IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

SPECIAL SECTION ON GEOMETRIC ALGEBRA IN SIGNAL PROCESSING

Received February 17, 2020, accepted March 2, 2020, date of publication March 6, 2020, date of current version March 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2978860

SVDTWDD Method for High Correct Recognition Rate Classifier With Appropriate Rejection Recognition Regions

GUOWEI YANG[1,2], **SHAOHUA QI**[1], **TENG YU**[1], **MINGHUA WAN**[2], **ZHANGJING YANG**[2], **TIANMING ZHAN**[2], **FANLONG ZHANG**[2], **AND ZHIHUI LAI**[3]

[1]School of Electronic Information, Qingdao University, Qingdao 266071, China
[2]Jiangsu Key Laboratory of Auditing Information Engineering, School of Information Engineering, Nanjing Audit University, Nanjing 211815, China
[3]School of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China

Corresponding author: Teng Yu (yutenghit@foxmail.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC0804002, in part by the National Science Foundation of China under Grant 61772277, Grant 61873071, Grant 61876213, Grant U1831127, Grant 61976117, Grant 61603192, and Grant 61976145, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20171494, in part by the Key Research and Development Program Science Foundation in Colleges and Universities of Jiangsu Province under Grant 18KJA520005.

ABSTRACT At present, regions of the same class determined by Support Vector Machines (SVM) classifier, Support Vector Domain Description (SVDD) classifier and Deep Learning (DL) classifier may occupy regions of other classes or unknown classes in feature space. There exists a risk that samples of other classes or unknown classes are wrongly classified as a known class. In this paper, the Support Vector Domain Tightly Wrapping Description Design (SVDTWDD) method with appropriate rejection regions and the corresponding incremental learning algorithm are proposed to overcome the above problem. The main work includes: (1) We develop a construction algorithm of the tightly wrapping set for the homogeneous feature set; (2) Based on the homogeneous feature set and tightly wrapping set, a novel algorithm is presented for obtaining the tightly wrapping surface of the homogeneous feature region; (3) The method for constructing all the public regions outside of the tightly wrapping surface and the intersections of wrapping regions in two different tightly wrapping surfaces, as the rejection region of the classifier; (4) An incremental algorithm is also presented based on the SVD-TWDD method. The experimental results with UCI data sets show that the correct recognition rate of our proposed method is nearly100% even if with a low rejection rate.

INDEX TERMS Classifier, geometric algebra, pattern recognition, support vector machine, support vector domain description, incremental learning, classification surface, wrapping learning.

I. INTRODUCTION

For easy understanding, we first introduce several concepts. For a classifier, in the public test data sets, the ratio of the number of rejected recognition samples to the total number of samples in the test data sets is called **the rejection recognition rate of a classifier**. In the test data sets without the rejected samples, the ratio of the number of correctly classified samples to the total number of samples is called **the correct recognition rate with a certain rejection recognition region (set)**. Assume $X \in R^N$ denote the feature value of samples, the area (set) C_ω formed by the features of all samples of a class ω is called the feature area (set) of homogeneous samples or the feature area (set) of actual homogeneous samples. If the classifier maps a region (set) ${}_cC_\omega$ to class ω, and points outside the ${}_cC_\omega$ are not mapped to class ω, at all, then ${}_cC_\omega$ is called the homogeneous feature region (set) of the class ω. The set ${}_{\omega T}C$ of the features of all training samples is called the homogeneous training sample feature set of the class ω.The region with ${}_{\omega T}C$ and all the points from R^N with distance to ${}_{\omega T}C$ that are less than δ is called ${}_{\omega T}C$-determined δ region ${}_{\omega T}C_\delta$, simplified as homogeneous class training feature region. As a result, there are 4 cases for a classifier: (1) $C_\omega = {}_cC_\omega$, i.e., the feature area (set) of homogeneous samples determined by the classifier equals with the feature area (set) of actual homogeneous samples.

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

In theory, the correct recognition rate of the classifier is 100%; (2) $C_\omega \supset {}_cC_\omega$, the feature area (set) of the homogeneous samples determined by the classifier loses part of the feature area (set) of the actual homogeneous samples. When the parts outside ${}_cC_\omega$ are set as rejection recognition areas (sets). Theoretically, the correct recognition rate of the classifier with a certain rejection recognition area (sets) is 100%; (3) $C_\omega \subset {}_cC_\omega$, i.e., the feature area (set) of the homogeneous sample determined by the classifier invades the feature area (set) of known samples of the homogeneous class or the feature area (set) of unknown samples of the homogeneous class. The classifier has the risk of misidentifying a sample of a known or unknown class as a known class. In theory, the correct recognition rate of the classifier cannot reach 100%; (4) $C_\omega - {}_cC_\omega \neq \Phi$, and ${}_cC_\omega - C_\omega \neq \Phi$, i.e., the feature regions (sets) of homogeneous samples determined by the classifier not only occupy the feature regions (sets) of known homogeneous samples of other types or the feature regions (sets) of homogeneous samples of unknown classes, but also lose the feature regions of some actual homogeneous samples (set). The risk of classifier misclassification is relatively large. In theory, the correct recognition rate of the classifier cannot reach 100%.

Over the past decades, many prominent results [1]–[29] in signal processing and in the design of classifiers have been achieved by geometric and algebra. However, at present, feature regions of the same class determined by Support Vector Machines (SVM) classifier, Support Vector Domain Description (SVDD) classifier and Deep Learning (DL) classifier may occupy feature regions of other classes or unknown classes. There is a risk that samples of other classes or unknown classes are wrongly classified as a known class. Hence, some SVM classifiers with high correct recognition rate, SVDD classifiers and DL classifiers still have about 2% error recognition rate [6]–[29]. These classifiers cannot be directly applied to the serious authentication and recognition applications, such as major disease detection, human identity authentication, and identification of banknote, bill, or terrorist. In these applications, it is often necessary to introduce a suitable rejection mechanism [6], so that classifiers can either reject a sample or classify it correctly. That is to say: (1) the rejection rate is very low; (2) the correct recognition rate is 100% or nearly 100%. If the rejection rate is high, then the applicable areas of the classifier are limited. If the correct recognition rate is not close to 100%, users dare not to use these classifiers to authenticate some particularly important things or events directly. Obviously, low rejection rate and high correct recognition rate are two contradictory events.

In order to achieve a low rejection rate and high correct recognition rate, it is necessary to design a classifier that making the feature region ${}_cC_\omega$ of the same ω class of samples (any point in the region is regarded as class ω, while the point outside the region is regarded as another class point or rejection recognition point), which contains almost all the actual feature region C_ω formed by the sample points of ω class (almost without losing its own domain), and almost does not invade the feature regions of other known classes and unknown possible classes. A good recognition system should also have the incremental learning function [7]. The incremental learning function enables the recognition system to inherit most of the existing knowledge of the system, optimize and upgrade the system continuously when the number of training samples or classes varies.

Aiming at solving the above problems, in this paper, we present a support vector domain tightly wrapping description design (SVDTWDD) method for classifier designing. In our method, a tightly wrapping surfaces of homogeneous samples for classification is constructed, in this way, we can obtain a classifier with high correct recognition rate as well as maintain a low rejection rate. Three contributions are made in our work.

1. We prove the existence of tightly wrapping Set of Homogeneous Training Feature Set.
2. The construction algorithm of tightly wrapping Set of homogeneous Training Feature set is given in our paper.
3. Based on the tightly wrapped set and sets from other classes, we develop an algorithm for tightly wrapping surface of homogeneous training feature region.

Based on the SVDTWDD, a novel method for incremental learning is developed.

II. RELATED WORKS

In this section, previous works of popular classifiers are reviewed.

The motivation of SVM is that SVM maps all feature vectors into a high-dimensional space, in which a maximally spaced hyperplane is established, and the corresponding primitive space surface of the hyperplane is regarded as the classification decision surface. Two parallel hyperplanes are constructed on both sides of a hyperplane that separates two types of eigenvectors (data). Separating hyperplanes maximizes the distance between two parallel hyperplanes. Obviously, the same kind of feature region ${}_cC_\omega$ determined by SVM are usually unbounded regions, while the actual homogeneous feature regions C_ω are bounded. Therefore, the homogeneous feature regions ${}_cC_\omega$ are determined by SVM encroach on the actual feature regions of other classes or the feature regions of unknown classes. Moreover, the embezzlement is serious and there is a greater risk of misclassification of samples. Therefore, it is not suitable for the applications, such as identification of major diseases, identification of people through biological characteristics, identification of banknotes, identification of bills and identification of terrorists. Moreover, when new training samples or new class are added, the work of solving SVM needs to be re-carried out. In the design of multi-classification SVM classifier [8], [9], when changing a training sample or adding a new category, the corresponding classifier learning and training process need to start over, and the classifier cannot inherit any results from the previous training process, so the multi-classifier of SVM has no incremental learning function. Among those improved SVM methods [10]–[13] (taking into account the imbalance of different feature regions and other characteristics), there exist two drawbacks: (1) There is no suitable rejection recognition mechanism and it is inconvenient to determine the rejection recognition region. Because the current methods may not improve the correct recognition rate by determining the rejection recognition region; (2) The same type of feature region ${}_{c}C_{\omega}$ determined by SVM may encroach on the feature region of the unknown classes, and there is a risk of misjudging the other classes or a unknown-class sample as a known-class one. Generally, the SVM classifier has about 2% error recognition rate, and the correct recognition rate cannot approach 100%. The known feature regions determined by SVM tend to occupy the unknown feature space as much as possible; (3) The SVM classifier has no incremental learning function: when the number of classes are added or subtracted, the learning and training work need to be completely restarted. Similarly, if any training data varies, the learning and training work need to start over again.

The idea of SVDD or hypersphere SVM classification algorithm [14]–[19] is to map all feature vectors into a high-dimensional space, in which a minimum radius hypersphere satisfying some constraints is established, the hypersphere contains almost all the homogeneous sample points. The original space surface corresponding to the hypersphere or concentric hypersphere is regarded as the classification decision surface. In the second section, the experiments show that the bread-wrapping region ${}_{c}C_{\omega}$ of SVDD or hypersphere SVM classification decision-making may encroach on the feature region $C_{\omega'}$ of unknown classes, that is, the actual feature region C_{ω} of homogeneous samples is not tightly wrapped in the classification decision-making region ${}_{c}C_{\omega}$. Therefore, there are similar problems in the hypersphere SVM classifier: (1) No suitable rejection recognition mechanism is introduced. Because it is not convenient to determine the appropriate rejection recognition region, or to determine the rejection recognition region reluctantly, but it may not improve the correct recognition rate; (2) classification decision-making bread-wrapped region ${}_{c}C_{\omega}$ may encroach on the feature region $C_{\omega'}$ of the unknown class, and there is a risk of misclassifying the samples from other classes or unknown class as known classes. In general, the classifier still has about 2% error recognition rate, the correct recognition rate cannot approach 100%; (3) the classifier has no incremental learning function. If the number of training data or classes varies, the learning and training work need to start over again.

In the era of big data, various deep learning classifiers are constantly developed to improve the correct recognition rate, and some object recognition tasks can achieve 99% correct recognition rate. In the past, it was almost impossible to recognize street scenery characters. Nowadays, the correct recognition rate of street scenery characters is also very high [20]–[26]. Unfortunately, these classifiers still have an error rate of about 2%, and they share similar problems as SVM and SVDD. Therefore, these classifiers are not suitable for the serious authentication and identification applications.

III. AN EXAMPLE OF THE SVDD CLASSIFICATION DECISION SURFACE THAT FAILS TO TIGHTLY WRAP THE HOMOGENEOUS SAMPLES IN FEATURE SPACE

Example 1: As shown in Figure 1, in a two-dimensional space, there exists an actual feature field surrounded by a regular triangle. 1000 points are sampled by using uniform sampling, and the decision surface is obtained by using a Gauss kernel function and SVDD training with penalty coefficient $C = 0.5$. The surface is approximately a circle. The inner circle is the feature region ${}_{c}C_{\omega}$ of the classifier. Obviously, the feature region ${}_{c}C_{\omega}$ occupies the actual other known or unknown class feature region $C_{\omega'}$ (i.e. part of the region outside the regular triangle) and loses part of its own feature region ${}_{c}C_{\omega}$ (i.e. two small regions of the regular triangle).

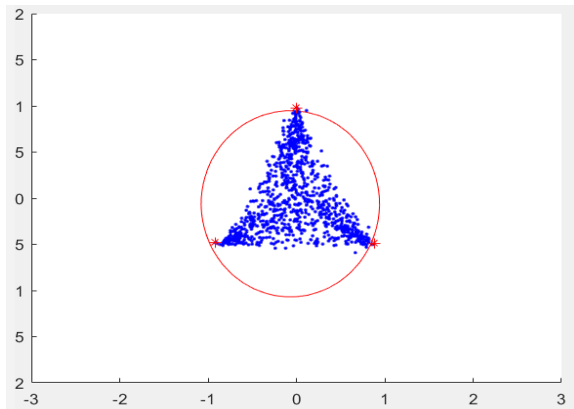

FIGURE 1. For SVDD, its classification decision surface cannot tightly wrap the homogeneous feature space.

IV. THE EXISTENCE THEOREM OF TIGHTLY WRAPPING SET OF HOMOGENEOUS TRAINING FEATURE SET

Given a point set ${}_{\omega T}C \subseteq R^N$ and the constants $\varepsilon > 0$, $r > 1$, we call the set ${}_{\omega T}C$ is a $\varepsilon/\sqrt{rN}$ compact connected set if for any two points X, Y of ${}_{\omega T}C$, there exist different points $X_1, \cdots, X_h$ such that all the distances $\delta(X, X_1), \delta(X_1, X_2), \cdots, \delta(X_h, Y)$ are less than $\varepsilon/\sqrt{rN}$. The parameters ε and r mainly been used to describe "compact connectivity" of discrete point sets. If $\varepsilon = 0.01, \mathrm{N} = 100, r = 9$, then $\varepsilon/\sqrt{rN} = 0.0003$. $\varepsilon/\sqrt{rN} = 0.0003$ show that ${}_{\omega T}C$ very tightly connected. At the same time, it shows that there must be other points in the sphere domain of any point with a radius of 0.0003 in the set ${}_{\omega T}C$.

Define the set $C_{\varepsilon/\sqrt{rN}} = {}_{\omega T}C \cup \{X : X \in \mathrm{R}^N, \delta(X, {}_{\omega T}C) \leq \varepsilon/\sqrt{rN}\}$, we call the set ${}_{\omega T}C$ is a $\varepsilon/\sqrt{rN}$ compactly simple connected set if $C_{\varepsilon/\sqrt{rN}}$ is mathematically simple connected. $\delta(X, {}_{\omega T}C) = \min\{\delta(X, Y) | Y \in {}_{\omega T}C\}$ is the minimum distance from X to this finite number of points. Intuitively, there exists no hole in a compactly simple connected set. We call the set ${}_{\omega T}C$ is a $\varepsilon/\sqrt{rN}$ compactly convex set if there exists any straight line segments, whose start

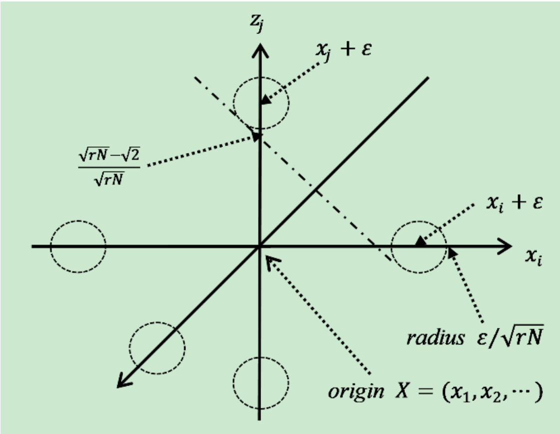

FIGURE 2. The new coordinate system with origin $X = (x_1, \cdots, x_i, \cdots, x_N) \in {}_{\omega T}C$.

point and end point are all in $C_{\varepsilon/\sqrt{rN}}$. Obviously, a $\varepsilon/\sqrt{rN}$ compactly simple connected set must be a $\varepsilon/\sqrt{rN}$ compact connected set. And a $\varepsilon/\sqrt{rN}$ compactly convex set must be a $\varepsilon/\sqrt{rN}$ compactly simple connected set.

Furthermore, we call the point X of ${}_{\omega T}C$ being $\varepsilon/\sqrt{rN}$ compact boundary point if $\delta\left(X, \mathrm{bd}\left(C_{\varepsilon/\sqrt{rN}}\right)\right) = \varepsilon/\sqrt{rN}$, where $\mathrm{bd}\left(C_{\varepsilon/\sqrt{rN}}\right)$ denotes the boundary of the set $C_{\varepsilon/\sqrt{rN}}$. At last, define the set $C_\varepsilon = {}_{\omega T}C \cup \{X : X \in \mathrm{R}^N, \delta(X, {}_{\omega T}C) \leq \varepsilon\}\}$.

Theorem 1: Let ${}_{\omega T}C \subseteq \mathrm{R}^N$ be a $\varepsilon/\sqrt{rN}$ convex compactly connected bounded set and Γ be the quantity of $\varepsilon/\sqrt{2N}$ compact boundary points. For any point $X = (x_1, \cdots, x_i, \cdots, x_N) \in {}_{\omega T}C$, one can obtain $2N$ points compactly connected:

$$(x_1 \pm \varepsilon, \cdots, x_i, \cdots, x_N), \cdots\cdots, (x_1, \cdots, x_i \pm \varepsilon, \cdots, x_N), \cdots\cdots, (x_1, \cdots, x_i, \cdots, x_N \pm \varepsilon). \quad (1)$$

Let $\varepsilon/\sqrt{rN} - \varepsilon$ tightly wrapping set:

$$I({}_{\omega T}C) = \left(C_\varepsilon - C_{\varepsilon/\sqrt{rN}}\right) \cap \{(x_1, \cdots, x_i \pm \varepsilon, \cdots, x_N) \mid (x_1, \cdots, x_i, \cdots, x_N) \in {}_{\omega T}C\}. \quad (2)$$

Then, when $r > \frac{(\sqrt{N}+\sqrt{2})^2}{N}$, there exist Γ points in $I({}_{\omega T}C)$ at least. And for each boundary point in ${}_{\omega T}C$, at least one of above mentioned $2N$ points is in $I({}_{\omega T}C)$. Furthermore, for any $X = (x_1, \cdots, x_i, \cdots, x_N) \in I({}_{\omega T}C)$,

$$\delta\left(X, \mathrm{hy}\left(C_{\varepsilon/\sqrt{rN}}\right)\right) > \varepsilon/\sqrt{rN}, \quad (3)$$

where $\mathrm{hy}\left(C_{\varepsilon/\sqrt{rN}}\right)$ denotes a hyper plane in $C_{\varepsilon/\sqrt{rN}}$.

Proof: (i) Let $X = (x_1, \cdots, x_i, \cdots, x_N) \in {}_{\omega T}C$ be a $\varepsilon/\sqrt{rN}$ compact boundary point. Thus one has:

$$\delta\left(X, \mathrm{bd}\left(C_{\varepsilon/\sqrt{rN}}\right)\right) = \varepsilon/\sqrt{rN}. \quad (4)$$

As illustrated in Figure 2, one can construct a new coordinate system with X being the origin, and denote the coordinate axis passing the points X and $(x_1, \cdots, x_i \pm \varepsilon, \cdots, x_N)$ by XX_i, $i = 1, \cdots, N$. Then, one can obtain a sphere $B(x_1, \cdots x_{i-1}, x_i \pm \varepsilon, x_{i+1} \cdots, x_N)$ with sphere center $(x_1, \cdots x_{i-1}, x_i \pm \varepsilon, x_{i+1} \cdots, x_N)$ and its radius $\varepsilon/\sqrt{rN}$. As a result, the points $(x_1, \cdots, x_{i-1}, x_i \pm \frac{\sqrt{rN}-\sqrt{2}}{\sqrt{rN}}\varepsilon, x_{i+1}, \cdots, x_N)$ are the intersection between coordinate XX_i and sphere B.

Let us denote $\Pi_{\mathrm{ij}\ldots}$, $\mathrm{i}, \mathrm{j} = 1 \cdots, \mathrm{N}$ as the tangent plane of spheres $B(x_1, \cdots x_{i-1}, x_i + \varepsilon, x_{i+1} \cdots, x_N)$ and $B(x_1, \cdots x_{j-1}, x_j + \varepsilon, x_{j+1} \cdots, x_N)$ which are vertical to coordinate planes $X_iXX_j \cdots, i, j = 1, \cdots, N$. Then, there exist two intersection points between $\Pi_{\mathrm{ij}\ldots}$ and coordinate system, which are

$$(x_1, \cdots x_{i-1}, x_i + \frac{\sqrt{rN}-\sqrt{2}}{\sqrt{rN}}\varepsilon, x_{i+1} \cdots, x_N) \quad \text{and}$$
$$(x_1, \cdots x_{j-1}, x_j + \frac{\sqrt{rN}-\sqrt{2}}{\sqrt{rN}}\varepsilon, x_{j+1} \cdots, x_N). \quad (5)$$

According the definition of hyper plane, given N different points, if they are not in any $N-2$ dimension hyper plane, there exists an $N-1$ dimension hyper plane such that these N points are all in it. Given $2N$ points:

$$(x_1 \pm \frac{\sqrt{rN}-\sqrt{2}}{\sqrt{rN}}\varepsilon, x_2 \cdots, x_N), \cdots, (x_1, \cdots, x_{i-1}, x_i \pm \frac{\sqrt{rN}-\sqrt{2}}{\sqrt{rN}}\varepsilon, x_{i+1}, \cdots, x_N), \cdots, (x_1, \cdots, x_{N-1}, x_N \pm \frac{\sqrt{rN}-\sqrt{2}}{\sqrt{rN}}\varepsilon). \quad (6)$$

There are C_{2N}^N different ways to choose N points, where for some ways the N points are all in some $N-1$ dimension hyper plane, but not in any $N-2$ dimension hyper plane. As a result, the number of $\mathrm{N}-1$ dimension hyper plane, which does not contain the origin point $X = (x_1, \cdots, x_i, \cdots, x_N) \in {}_{\omega T}C$ and $(x_1, \cdots, x_{i-1}, x_i \pm \frac{\sqrt{rN}-\sqrt{2}}{\sqrt{rN}}\varepsilon, x_{i+1}, \cdots, x_N)$, is

$$\frac{C_{2N}^1 C_{2N-2}^1 \cdots C_2^1}{N!} = 2^N. \quad (7)$$

For example, the result formula (7) is equal to 8 in three dimensional space.

And the polyhedron enclosed by these $\mathrm{N}-1$ dimension hyper planes are convex, denoted as Ω. Among all polyhedrons with 2 N points (6) as vertexes, the Ω's volume is the smallest. The point $X = (x_1, \cdots, x_i, \cdots, x_N) \in {}_{\omega T}C$ is the center points of Ω.

(ii) On the contrary, one can assume that the $2N$ points: $(x_1 \pm \varepsilon, \cdots, x_i, \cdots, x_N), \cdots\cdots, (x_1, \cdots, x_i \pm \varepsilon, \cdots, x_N), \cdots\cdots, (x_1, \cdots, x_i, \cdots, x_N \pm \varepsilon)$ are all in $C_{\varepsilon/\sqrt{rN}}$. According the definition of $C_{\varepsilon/\sqrt{rN}}$, there are the $2N$ points:

$$(x_1' \pm \varepsilon, \cdots, x_i', \cdots, x_N'), \cdots\cdots, (x_1', \cdots, x_i' \pm \varepsilon, \cdots, x_N'), \cdots\cdots, (x_1', \cdots, x_i', \cdots, x_N' \pm \varepsilon) \quad (8)$$

in ${}_{\omega T}C$ is in the sphere $B\left(x_1', \cdots, x_i' \pm \varepsilon, \cdots, x_N'\right)$. Ω' is the smallest convex polyhedron with (8) as vertexes. Then Ω' must be in $C_{\varepsilon/\sqrt{rN}}$, since the ${}_{\omega T}C$ is a $\varepsilon/\sqrt{rN}$ convex compact set. Obviously, $X = (x_1, \cdots, x_i, \cdots, x_N) \in \Omega'$, and Ω' is also enclosed by 2^N hyper planes. The intersection points between coordinate XX_i and hyper planes with common vertex $\left(x_1', \cdots, x_i' + \varepsilon', \cdots, x_N'\right)$ are not inside of the line segment with endpoints of points $X = (x_1, \cdots, x_i, \cdots, x_N)$ and $(x_1, \cdots, x_{i-1}, x_i + \frac{\sqrt{rN}-\sqrt{2}}{\sqrt{rN}}\varepsilon, x_{i+1}, \cdots, x_N)$, but in the extension of above-mentioned line segment. Thus the vertex angle of Ω' corresponding the vertex point $\left(x_1', \cdots, x_i' + \varepsilon', \cdots, x_N'\right)$ covers the $(x_1, \cdots, x_{i-1}, x_i + \frac{\sqrt{rN}-\sqrt{2}}{\sqrt{rN}}\varepsilon, x_{i+1}, \cdots, x_N)$, so

$$(x_1, \cdots, x_{i-1}, x_i + \frac{\sqrt{rN}-\sqrt{2}}{\sqrt{rN}}\varepsilon, x_{i+1}, \cdots, x_N) \in \Omega'. \quad (9)$$

Homogeneously, one has $(x_1, \cdots, x_{i-1}, x_i - \frac{\sqrt{rN}-\sqrt{2}}{\sqrt{rN}}\varepsilon, x_{i+1}, \cdots, x_N) \in \Omega'$. According the definition of the smallest convex set, one has

$$X = (x_1, \cdots, x_i, \cdots, x_N) \in \Omega \subset \Omega' \subset C_{\varepsilon/\sqrt{rN}}. \quad (10)$$

Next, we prove that the distance $\delta\left(X, hy\left(C_{\varepsilon/\sqrt{rN}}\right)\right) > \varepsilon/\sqrt{rN}$ holds, when $r > \frac{\left(\sqrt{N}+\sqrt{2}\right)^2}{N}$. In fact one can verify that the hyper plane equation passing the points $(x_1 + \frac{\sqrt{rN}-\sqrt{2}}{\sqrt{rN}}\varepsilon, x_2 \cdots, x_N), \cdots, (x_1, \cdots, x_{i-1}, x_i + \frac{\sqrt{rN}-\sqrt{2}}{\sqrt{rN}}\varepsilon, x_{i+1}, \cdots, x_N), \cdots, (x_1, \cdots, x_{N-1}, x_N + \frac{\sqrt{rN}-\sqrt{2}}{\sqrt{rN}}\varepsilon)$ is:

$$(y_1 - x_1) + \cdots + (y_N - x_N) - \frac{\sqrt{rN}-\sqrt{2}}{\sqrt{rN}}\varepsilon = 0 \quad (11)$$

Note that this hyper plane is one of the boundary planes of Ω. According to the distance definition between point and surface, the distance between $X = (x_1, \cdots, x_i, \cdots, x_N)$ and surface of Ω is

$$\begin{aligned} \rho_{X\Omega} &= \rho_{Y=X} \\ &= \frac{|A_1(y_1 - x_1) + \cdots + A_n(y_N - x_N) + B|}{\sqrt{A_1^2 + A_2^2 \cdots + A_n^2}} \\ &= \frac{\left|(y_1 - x_1) + \cdots + (y_N - x_N) - \frac{\sqrt{rN}-\sqrt{2}}{\sqrt{rN}}\varepsilon\right|}{\sqrt{A_1^2 + A_2^2 \cdots + A_n^2}} \\ &= \frac{\sqrt{rN}-\sqrt{2}}{N\sqrt{r}}\varepsilon > \varepsilon/\sqrt{rN}, \end{aligned} \quad (12)$$

which is a contradiction with $\Omega \subset \Omega' \subset C_{\varepsilon/\sqrt{rN}}$ and $X = (x_1, \cdots, x_i, \cdots, x_N) \in {}_{\omega T}C$ *is* a $\varepsilon/\sqrt{rN}$ compact boundary point. Note that $(\sqrt{N}+\sqrt{2}) \wedge 2/N$ decreases as N increases, and its value is not bigger than 5.8289.

V. THE CONSTRUCTION ALGORITHM OF TIGHTLY WRAPPING SET OF HOMOGENEOUS TRAINING FEATURE SET

In this paper, the so-called $\varepsilon/\sqrt{rN}$—ε tightly wrapping set is $I({}_{\omega T}C) = ({}_{\omega T}C_\varepsilon - {}_{\omega T}C_{\varepsilon/\sqrt{rN}}) \cap \{(x_1, \cdots, x_i \pm \varepsilon, \cdots, x_N)\,|(x_1, \cdots, x_i, \cdots, x_N) \in {}_{\omega T}C\}$. It is a set related to a certain set ${}_{\omega T}C$, the points of which are restricted within a skin cavity (region) with thickness less than $\varepsilon - \varepsilon/\sqrt{rN}$, the skin cavity wraps the outer side of set ${}_{\omega T}C$.

A. OPTIMIZATION ALGORITHM OF COMPACTNESS PARAMETER

Let us define set $C \subseteq R^N$ as the homogeneous feature points set coming from the homogeneous feature region $T \subseteq R^N$, and assume ${}_{\omega T}C$ as a $\varepsilon/\sqrt{5.9N}$ convex compact bounded set. If the homogeneous feature region T is fixed, the number of points in ${}_{\omega T}C$ increases as ε decreases.

On one hand, given the convex set T and parameter ε, it's easy to construct a set ${}_{\omega T}C$ to be $\varepsilon/\sqrt{5.9N}$ convex compact. On the other hand, given a set ${}_{\omega T}C$ *as* $\varepsilon/\sqrt{5.9N}$ convex compact, it's difficult to compute the smallest ε. Next, we present an algorithm to find the smallest ε. This algorithm is based on the theory that a simplex determined by $N+1$ points, which are not in the same hyper plane, is a minimal volume convex polyhedron containing these $N+1$ points. The algorithm details are shown as follows.

Algorithm 1 Optimization Algorithm of Compactness Parameters

Input: M points in ${}_{\omega T}C$, denoting these points by $X_1, X_2 \cdots, X_M$.

Step 1: compute the 1^{st} neighbor X_{j_1} of point X_j:
$j_1 = argmin_{i \neq j}\left\|X_j - X_i\right\|$

Step 2: compute the 2^{nd} neighbor X_{j_2} of point X_j:
$j_2 = argmin_{i \neq j, i \neq j_1}\left\|X_j - X_i\right\|$

$\vdots$

Step N+1: compute the $(N+1)^{th}$ nearest neighbor $X_{j_{N+1}}$ of point
$X_j : j_{N+1} = argmin_{i \notin \{j, j_1, \ldots, j_N\}}\left\|X_j - X_i\right\|$

Step N+2: compute the max distance between X_j and its neighbors:
$g(X_j) = max_{i \in \{j_1, \ldots, j_N\}}\left\|X_j - X_i\right\|$

Step N+3: compute the sub-optimal ε:
$\varepsilon \geq \sqrt{5.9N}\ max_{1 \leq j \leq M}\, g(X_j)$
$= \sqrt{5.9N} max_{1 \leq j \leq M}\, g(X_j)$

Output: ε

One can prove under the ε obtained by the above algorithm, the ${}_{\omega T}C$ is $\varepsilon/\sqrt{5.9\text{N}}$ convex compact. Besides, the algorithm is convergent and the computation complexity is O(M^2) when M is finite.

B. THE CONSTRUCTION ALGORITHM OF TIGHTLY WRAPPING SET

This part illustrates how to construct a $\varepsilon/\sqrt{5.9N}-\varepsilon$ tightly wrapping set $I({}_{\omega T}C)$:

$$\begin{aligned} I\left({}_{\omega T}C\right) = &\left(C_\varepsilon - C_{\varepsilon/\sqrt{5.9N}}\right) \\ &\cap \left\{\left(x_{j1},\cdots,x_{ji}\pm\varepsilon,\cdots,x_{jN}\right) \mid \right. \\ &\left.\times\left(x_{j1},\cdots,x_{ji},\cdots,x_{jN}\right)\in {}_{\omega T}C\right\} \end{aligned} \quad (13)$$

Firstly, we can construct a $\varepsilon/\sqrt{5.9N}$ hyper-spherical neighborhood discriminant function

$$f_j\left(X\right) = \varepsilon/\sqrt{5.9N} - \left\|X - X_j\right\|, \quad 1 \le j \le M. \quad (14)$$

If $f_j\left(X\right) \ge 0$, then X lies in the hyper-spherical neighborhood $\prod(X_j)$, whose spherical center is X_j and radius is $\varepsilon/\sqrt{5.9N}$.

Secondly, for each point $X_j = \left(x_{j1},\cdots,x_{ji},\cdots,x_{jN}\right), 1 \le j \le M$, one can construct $2N$ points:

$$\left(x_{j1},\cdots,x_{ji}\pm\varepsilon,\cdots,x_{jN}\right), \quad 1 \le i \le N \quad (15)$$

Thirdly, check if each constructed point is in $\prod(X_j)$, $1 \le j \le M$. Then the set of the all points, which are not in any $\prod(X_j)$, $1 \le j \le M$, is a $\varepsilon/\sqrt{5.9N}-\varepsilon$ tightly wrapping set, i.e., $I\left({}_{\omega T}C\right)$. Based on Theorem 1, the number of points in $I\left({}_{\omega T}C\right)$ is not less than the number of boundary points of ${}_{\omega T}C$.

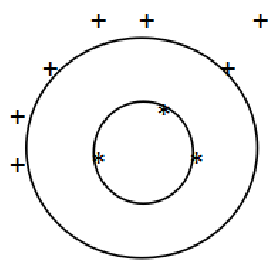

FIGURE 3. Illustration of tightly wrapping surface.

VI. THE ALGORITHM FOR TIGHTLY WRAPPING SURFACE OF HOMOGENEOUS TRAINING FEATURE REGION BASED ON THE TIGHTLY WRAPPED SET AND THE OTHER CLASS SAMPLE SET

A feature transformation $\phi : R^N \to H$ can maps the feature space to a high dimensional space, denoting the corresponding kernel function as $\mathrm{k} : \mathrm{R}^\mathrm{N}\times\mathrm{R}^\mathrm{N} \to \mathrm{R}$. As illustrated in Figure 3, we assume the radius of small hyper sphere and big one are r and $\sqrt{r^2+\rho^2}$ respectively. The '*' points in the small sphere are mapped points of the corresponding ones in ${}_{\omega T}C$. The '+' points outside the big sphere are mapped points of the corresponding ones in $\mathrm{I}\left({}_{\omega T}C\right)$ and the other class sample set $\bigcup\limits_{j\neq i} C_j$. We aim to pursue a proper transformation $\phi : \mathrm{R}^\mathrm{N} \to \mathrm{H}$ such that the small sphere almost contains all '$*$' points with smallest r and largest $\sqrt{\mathrm{r}^2+\rho^2}$. As a result, the primal space surface corresponding to the small hyper sphere of high dimension is the tightly wrapped surface of ${}_{\omega T}C$.

For easy calculation, we assume there are m_1 points in ${}_{\omega T}C$, and $m_2 = n - m_1$ points in $I(C)\cup\bigcup\limits_{j\neq i} C_j$, the point c is the center of high dimensional sphere. We construct the tightly wrapped surface of the homogeneous area by computing the optima of the following optimization function.

$$\begin{aligned} &\min_{r,c,\rho,\xi}(r^2 - \upsilon\rho^2 + \tfrac{1}{\upsilon_1 m_1}\sum_{i=1}^{m_1}\xi_i + \tfrac{1}{\upsilon_2 m_2}\sum_{j=m_1+1}^{n}\xi_j), \\ &s.t.\ \|\phi(X_i)-c\|^2 \le r^2+\xi_i, \quad 1\le i\le m_1, \\ &\quad \left\|\phi(X_j)-c\right\|^2 \ge r^2+\rho^2-\xi_j, \quad m_1 \le j \le n, \\ &\quad 0 \le \xi_k, \quad 1 \le k \le n, \end{aligned} \quad (16)$$

where ξ_i, ξ_j are slacking variables, $\frac{1}{\upsilon_1 m_1}, \frac{1}{\upsilon_2 m_2}$ denote the punishment terms. To solve this optimization problem, we utilize the Lagrange function as:

$$\begin{aligned} L(r,c,\xi,\alpha,\beta) = &\ r^2 - \nu\rho^2 + \tfrac{1}{\upsilon_1 m_1}\sum_{i=1}^{m_1}\xi_i + \tfrac{1}{\upsilon_2 m_2}\sum_{j=1}^{n}\xi_j \\ &+\sum_{i=1}^{m_1}\alpha_i(\|\phi(X_i)-c\|^2 - r^2 - \xi_i) \\ &-\sum_{j=m_1+1}^{n}\alpha_j(\|\phi(X_i)-c\|^2 - r^2 - \rho^2 + \xi_j) \\ &-\sum_{k=1}^{n}\beta_k\xi_k \end{aligned} \quad (17)$$

Then the optima should satisfy the following conditions:

$$\begin{cases} \dfrac{\partial L}{\partial r} = 2r(1-\sum\limits_{i=1}^{n}\alpha_i y_i) = 0 \\ \dfrac{\partial L}{\partial \rho} = 2\rho(-\upsilon+\sum\limits_{j=m_1+1}^{n}\alpha_i y_i) = 0 \\ \dfrac{\partial L}{\partial \xi_i} = \dfrac{1}{\upsilon_1 m_1} - \alpha_i - \beta_i = 0, \quad 1 \le i \le m_1 \\ \dfrac{\partial L}{\partial \xi_j} = \dfrac{1}{\upsilon_2 m_2} - \alpha_j - \beta_j = 0, \quad m_1+1 \le j \le n \\ \dfrac{\partial L}{\partial c} = 2c\sum\limits_{i=1}^{n}\alpha_i y_i - 2\sum\limits_{i=1}^{n}\alpha_i y_i \phi(X_i) = 0 \end{cases} \quad (18)$$

As a result, we can obtain:

$$c = \frac{\sum_{i=1}^{n}\alpha_i y_i \phi(X_i)}{\sum_{i=1}^{n}\alpha_i y_i} = \sum_{i=1}^{n}\alpha_i y_i \phi(X_i) \quad (19)$$

So, the dual problem is as follows:

$$\begin{aligned} &\max_\alpha \sum_{i=1}^{n}\alpha_i y_i \phi(X_i)\cdot\phi(X_i) - \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \phi(X_i)\cdot\phi(X_j) \\ &s.t.\ 0 \le \alpha_i \le \frac{1}{\upsilon_1 m_1}, \quad 1 \le i \le m_1 \\ &\quad 0 \le \alpha_j \le \frac{1}{\upsilon_2 m_2}, \quad m_1+1 \le j \le n \\ &\quad \sum_{i=1}^{n}\alpha_i y_i = 1, \sum_{i=1}^{n}\alpha_i = 2v+1 \end{aligned} \quad (20)$$

where $\phi(X_i)\cdot\phi(X_j)$ are instead by $K(X_i,X_j)$, the above optimization can be rewritten as:

$$\begin{aligned}&\max_{\alpha}\sum_{i=1}^{n}\alpha_i y_i K(X_i,X_i)-\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(X_i,X_j)\\ &s.t.\ 0\le\alpha_i\le\frac{1}{\upsilon_1 m_1},\quad 1\le i\le m_1\\ &\qquad 0\le\alpha_j\le\frac{1}{\upsilon_2 m_2},\quad m_1+1\le j\le n\\ &\qquad \sum_{i=1}^{n}\alpha_i y_i=1,\ \sum_{i=1}^{n}\alpha_i=2\upsilon+1\end{aligned}\tag{21}$$

This is a quadratic optimization problem and there exist many algorithms for solving it, such as the sequential minimum optimization algorithm and the complexity of this algorithm is $o(n^2)$.

Furthermore, to obtain r, ρ^2 and $r^2+\rho^2$, we consider following two sets:

$$\begin{aligned}S_1&=\{x_i\,|0<\alpha_i<\tfrac{1}{\upsilon_1 m_1},\quad 1\le i\le m_1\},\\ S_2&=\{x_j\,|0<\alpha_j<\tfrac{1}{\upsilon_2 m_2},\quad m_1+1\le j\le n\}\end{aligned}\tag{22}$$

Let $n_1=|S_1|$, $n_2=|S_2|$, by KTT conditions, we have:

$$r^2=\frac{1}{n_1}P_1,\quad \rho^2=\frac{1}{n_2}P_2-\frac{1}{n_1}P_1.\tag{23}$$

where

$$\begin{aligned}P_1&=\sum_{x_i\in S_1}\|\phi(X_i)-c\|^2=\sum_{x_i\in S_1}(k(X_i,X_i)\\ &\quad -2\sum_{k=1}^{n}\alpha_k y_k k(X_i,X_k)+\langle c,c\rangle)\end{aligned}\tag{24}$$

$$\begin{aligned}P_2&=\sum_{x_j\in S_2}\|\phi(X_i)-c\|^2=\sum_{x_j\in S_2}(k(X_j,X_j)\\ &\quad -2\sum_{k=1}^{n}\alpha_k y_k k(X_j,X_k)+\langle c,c\rangle)\end{aligned}\tag{25}$$

$$\begin{aligned}\langle c,c\rangle&=\langle\sum_{i=1}^{n}\alpha_i y_i\phi(X_i),\sum_{j=1}^{n}\alpha_j y_j\phi(X_j)\rangle\\ &=\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j k(X_i,X_j)\end{aligned}\tag{26}$$

Thus, the class decision function is given by:

$$\begin{aligned}f(x)&=\mathrm{sgn}(r^2-\|\phi(X)-c\|^2)\\ &=\mathrm{sgn}(r^2+2\sum_{k=1}^{n}\alpha_k y_k k(X,X_k)-k(X,X)-\langle c,c\rangle)\end{aligned}\tag{27}$$

The class decision surface W is given by

$$\begin{aligned}&r^2-\|\phi(X)-c\|^2\\ &=r^2+2\sum_{k=1}^{n}\alpha_k y_k k(X,X_k)-k(X,X)-\langle c,c\rangle=0\end{aligned}\tag{28}$$

The construction procedures of tightly wrapped set and classification decision surface show that the constructed surface can compactly wrapping the region of homogeneous features, i.e., the volume difference between the wrapped region and the homogeneous feature region is less than a certain bounded number multiplying ε. Furthermore, the distances, also called projection magnitudes, between the boundary of feature region and boundary of class decision surface are all less than ε.

VII. TIGHTLY WRAPPING LEARNING THEOREM

Theorem 2 (Tightly Wrapping Learning Theorem): Let $\delta>0$ and ${}_{\omega T}C$ is a $\varepsilon/\sqrt{rN}$ compact compactly connected set of points with bounded M in a N-dimensional feature space R^N and ${}_{\omega T}C$ is a $\varepsilon/\sqrt{rN}$ compact convex set with $\varepsilon<\frac{\delta}{2^N N(M+1)^{N-1}}$, $r>5.9$. Then the difference between the volume of $C_{\varepsilon/\sqrt{rN}}$ and the volume of the region enclosed by the surface W of (28) is less than δ. That is, when δ is sufficient small, the surface W tightly wraps the feature region of the same kind.

Proof: Based on the theory of the hyper-sphere SVM or SVDD method, as long as the surface W of (28) (the approximate solution can be obtained by numerical calculation method for points on the surface) is found, the distance from the boundary of the feature region to the classification decision boundary in each coordinate direction (the projection length of the coordinate direction) is greater than $\varepsilon/\sqrt{rN}$ and less than ε, according to the spatial position of the surface of (28). The difference between the volume of $C_{\varepsilon/\sqrt{rN}}$ and volume of the region enclosed by the curved surface is less than the difference $(2M+2\varepsilon)^N-(2M)^N$ between the volume of the cube with the edge length $2M$ and the cube with the edge length $(2M+2\varepsilon)$. So the difference between the volume of $C_{\varepsilon/\sqrt{rN}}$ and volume of the region enclosed by the curved surface is less than $2^N\varepsilon N(M+\varepsilon)^{N-1}$.

Hence the difference between the volume of $C_{\varepsilon/\sqrt{rN}}$ and the volume of the region enclosed by the surface W of (28) is less than

$$2^N\varepsilon N(M+\varepsilon)^{N-1}<\delta\tag{29}$$

That is, as long as δ is small enough, the surface W is able to tightly wraps the feature region of the homogeneous class.

The above "The construction algorithm of tightly wrapping set of homogeneous feature set" and "The algorithm for tightly wrapping surface of homogeneous feature region based on wrapping set and the other class sample set " are collectively referred as the SVDTWDD method for support vector domain tightly wrapped description design.

Example 2: Figure 4 is the set of tightly wrapping points and tightly wrapping surfaces of feature region of an object (identifying the same feature region of an object) enclosed by an equilateral triangle in example 1. Obviously, the wrapping surface obtained by this method is much tighter than that obtained by SVDD (as in example 1).

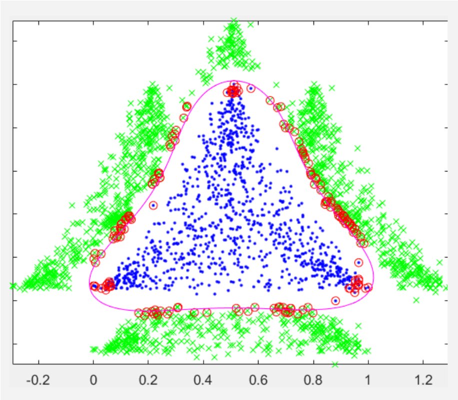

FIGURE 4. The set of the tight wrapping points and tight wrapping surfaces of feature domains. The blue dots and green crosses represent samples of two different classes respectively in feature space, the red circles are the support vectors of the decision surface represented by the magenta curve.

VIII. APPROPRIATE REJECTION RECOGNITION REGIONS FOR MULTI-CLASS CLASSIFIERS

For the multi-class classification problem, we assume η classes as $\omega_1, \cdots, \omega_\eta$. By using the above algorithm, we can find a small parameter ε. The discriminant functions of each class are given as follows:

$$\begin{aligned} {}_jf(X) &= \mathrm{sgn}({}_jr^2 - \left\|\phi(X) - {}_jc\right\|^2) \\ &= \mathrm{sgn}({}_jr^2 + 2\sum_{k=1}^{n} {}_j\alpha_k y_k k(X, X_k) - k(X, X) - \langle {}_jc, {}_jc\rangle) \\ &\quad 1 \le j \le \eta, \end{aligned} \tag{30}$$

And the decision surfaces are:

If there is only one j that ${}_jf(X) \ge 0$, then judge X belongs to class ω_j.

If there are more than two j that ${}_jf(X) \ge 0$, then judge X belongs to the intersection of wrapping regions by two different tightly wrapping surfaces and rejection recognition regions. **The intersections of wrapping regions by two different tightly wrapping surfaces are the areas of Maximum decision risk for classifier**. Adding the intersections to the rejection recognition regions will ensure that the correct recognition rate of the classifier with certain rejection recognition regions is 100% or close to 100%.

If ${}_jf(X) < 0$ for all j, then judge X belongs to the rejection recognition regions.

As a result, the rejection recognition regions can be defined as the public regions outside of all the tightly wrapped surfaces and the intersections of wrapping regions by two different tightly wrapping surfaces.

IX. ALGORITHM FOR INCREMENTAL LEARNING

Let's say that the samples of new classes the intersections of wrapping regions by two different tightly wrapping surfaces. When new classes are added, all the results of the previous work need to be retrained. For the sample set of new classes, one only needs to pursue the wrapped set, classification decision functions and the classification decision tightly wrapped surfaces. At the same time, the rejection recognition region should be adjusted accordingly.

When new training samples are added and the previous classifier can classify correctly, the classifier does not need to make any adjustment. Otherwise, only the wrapped point set corresponding to the misclassification and rejection recognition samples (Note: New boundary points can be judged) need to be recalculated, and the classification decision function and classification decision-making surface based on the new class sample set and wrapping set can be recalculated. At the same time, the rejection recognition region should be adjusted accordingly.

When subtracting the misclassification samples from the training data, only the set of wrapping points corresponding to the misclassification samples (note: boundary points can be judged) need to be retrieved from the set of wrapping points (note: most of the previous calculation results can be used) and the classification decision function and classification decision-making tightly wrapped surface based on the new class sample set and the wrapped set can be obtained. At the same time, the rejection recognition region should be adjusted accordingly.

TABLE 1. Description of the data sets.

Datasets	pos	neg	m1	m2	d
Iris	50	100	50	100	4
Wine	59	119	59	119	13
Blood	570	178	487	15	4
Cancer	444	239	398	12	9

TABLE 2. The parameter values of different datasets in our method.

Dataset	v	v1	v2
Iris	10	0.1	0.1
Wine	10	0.1	0.1
Blood	50	0.1	0.1
Cancer	30	0.1	0.01

X. EXPERIMENTS

The effectiveness of the proposed algorithm is evaluated on UCI data set, in which our proposed method is compared to the classic SVM, support vector data description SVDD, small sphere large margin SVM (SSLM-SVM). Some description of data set is displayed in Table 1, where 'pos' and 'neg' denote the samples number of '+' class and '−' class. The symbols m1, m2 denote the number of selected samples in the experiments, d denotes the number of the feature dimensions.

The parameters in experiments are decided by using the grid searching and cross validating methods. The Gaussian kernel function is used:

$$K(u, v) = \exp(-\frac{1}{\delta}\left\|u - v\right\|^2) \tag{31}$$

TABLE 3. Experiment results.

Dataset	Classical SVM	SVDD	SSLM-SVM	SVDTWDD (No rejection region, training without optimization)	SVDTWDD (Rejection region is set)	
					Recognition rate	Rejection rate
Iris	98.3%	98,5%	100%	100%	100%	0%
Wine	97.75%	85.34%	100%	100%	100%	0%
Blood	68.67%	71.12%	71.33%	72.67%	99.32%	2.3%
Cancer	94.75	93.78	95.25%	97.45%	99.15%	1.8%

The parameter δ is selected in $\{\frac{\sigma_0^2}{16}, \frac{\sigma_0^2}{8}, \frac{\sigma_0^2}{4}, \frac{\sigma_0^2}{2}, \sigma_0^2, 2\sigma_0^2, 4\sigma_0^2, 8\sigma_0^2, 16\sigma_0^2\}$ where σ_0^2 is the average norm of samples.

For the SVDD-SVM, the parameters C1 is selected in {0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50, 100, 500} and C2 in $\{\frac{1}{4} \times \frac{m_1}{m_2}, \frac{1}{2} \times \frac{m_1}{m_2}, 2 \times \frac{m_1}{m_2}, 4 \times \frac{m_1 m_2}{\}}$.

For the SSLM-SVM and our proposed method, the parameters v is selected as {10,30,50,70,90}, v1 and v2 as {0.001, 0.01, 0.1}. The parameter values of different datasets in the SVDTWDD method are displayed in Table 2.

According to [14], the positive class samples and negative class samples are selected from the data set and trained by using SVM, SVDD, SSLM, deep sensing network and wrapping learning algorithm respectively. The remaining samples are used for testing. For the problem of two classifications using the wrapping learning algorithm, the experimental method is: first, the wrapping learning algorithm is implemented for the classes (positive classes) with more training samples, and during the implementation of the wrapping learning algorithm, the other class of samples are added to the wrapping sample set for training. Secondly, the initial discriminant function is used to test and optimize the classification method. In order to satisfy the test results, it is necessary to scale the discriminant function or to add the wrong samples to the training set for retraining. If the correct recognition rate of the test results is 100% or fairly satisfactory, then there is no need to implement the wrapping learning algorithm for the other class (negative class) samples. No rejection recognition region is set. If the result of the test is not satisfactory, the wrapping learning algorithm is applied to the other class of (negative) samples to repeat the above process, and the discriminant region of this class (negative) is set as the wrapping region of this class to remove the other (positive) discriminant region, i.e., the positive class discriminant priority. Otherwise we set the positive discriminant region as the positive discriminant region and remove the other (negative) discriminant region, i.e., the negative discriminant priority. The problem whether positive or negative classes are to be taken as priority, is determined based on the test results. Then, the rejection recognition region is set outside the positive discrimination region and the negative discrimination region. Finally, the rejection recognition rate is calculated. Generally, if the rejection recognition region is relatively large, even very large, the rejection recognition rate is relatively small. Table 3 below show the experimental results. It can be seen that the wrapping learning algorithm is effective and can be used to design a classifier with high correct recognition rate and suitable rejection recognition region.

XI. CONCLUSION AND DISCUSSION

In this paper, we have presented a novel method for designing a classifier with high recognition rate. In the biometric areas, such as identity identification, banknote identification, and ticket authentication, many existing classifiers, such as SVM, SVDD, hyper-sphere SVM and deep learning, still have some drawbacks. One is the lack of appropriate rejection recognition mechanism and the other one is the correct recognition rate is not 100%. To overcome these drawbacks, this paper demonstrates that the homogeneous features are not tightly wrapped in the decision-hyperplane of the hypersphere SVM using a concrete example. Then we give the theorem proof of the existence of tightly wrapping set of homogeneous features.

As a conclusion, we have proposed the novel SVDTWDD method: the optimization algorithm of compactness parameters, construction algorithm of tightly wrapping set of homogeneous features, the algorithm of the tightly wrapping surface of homogeneous feature region based on homogeneous feature set and tightly wrapping set, and the method of setting the rejection region of the multi-class classifier. Furthermore, we have discussed the incremental learning algorithms for cases of adding new categories, increasing or decreasing training samples. The experiments on UCI data show that the proposed classifier can achieve nearly 100% recognition rate with a low rejection rate.

REFERENCES

[1] W. Cao, W. Feng, Q. Lin, G. Cao, and Z. He, "A review of hashing methods for multimodal retrieval," *IEEE Access*, vol. 8, pp. 15377–15391, 2020.

[2] M. Ashraphijuo, V. Aggarwal, and X. Wang, "On deterministic sampling patterns for robust low-rank matrix completion," *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 343–347, Mar. 2018.

[3] T. Zhan, M. Wan, L. Sun, Y. Xu, G. Yang, Z. Lu, and Z. Wu, "Robust regression-based Markov random field for hyperspectral image classification," *IEEE Access*, vol. 7, pp. 11868–11881, 2019.

[4] P. Yang, F. Zhang, and G. Yang, "Fusing DTCWT and LBP based features for rotation, illumination and scale invariant texture classification," *IEEE Access*, vol. 6, pp. 13336–13349, 2018.

[5] S. Franchini, A. Gentile, F. Sorbello, G. Vassallo, and S. Vitabile, "ConformalALU: A conformal geometric algebra coprocessor for medical image processing," *IEEE Trans. Comput.*, vol. 6, no. 4, pp. 955–970, Apr. 2015.

[6] W. Shoujue, "Bionic pattern recognition (topology pattern recognition) theory and application of a new pattern recognition model," *J. Chin. Electron.*, vol. 30, no. 10, pp. 1417–1420, 2002.

[7] Z. Zhihua, *Machine Learning*. Beijing, China: Tsinghua Univ. Press, 2016.

[8] V. Vapnik, *Statistical Learning Theory*. New York, NY, USA: Wiley, 1998.

[9] F. Tang, L. Adam, and B. Si, "Group feature selection with multiclass support vector machine," *Neurocomputing*, vol. 317, no. 23, pp. 42–49, 2018.

[10] Y. Tian, Z. Qi, X. Ju, Y. Shi, and X. Liu, "Nonparallel support vector machines for pattern classification," *IEEE Trans. Cybern.*, vol. 44, no. 7, pp. 1067–1079, Jul. 2014.

[11] L. Ding and S. Liao, "An approximate approach to automatic kernel selection," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 554–565, Mar. 2017.

[12] A. Cano, A. Zafra, and S. Ventura, "Weighted data gravitation classification for standard and imbalanced data," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1672–1687, Dec. 2013.

[13] R Khemchandani and S Chandra, "Twin support vector machines for pattern classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 905–910, Mar. 2007.

[14] D. M. J. Tax and R. P. W. Duin, "Support vector domain description," *Pattern Recognit. Lett.*, vol. 20, nos. 11–13, pp. 1191–1199, Nov. 1999.

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

[16] M. Wu and J. Ye, "A small sphere and large margin approach for novelty detection using training data with outliers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 2088–2092, Nov. 2009.

[17] Y. Xu and C. Liu, "A rough margin-based one class support vector machine," *Neural Comput. Appl.*, vol. 22, no. 6, pp. 1077–1084, May 2013.

[18] J. Chorowski, J. Wang, and J. M. Zurada, "Review and performance comparison of SVM-and ELM-based classifiers," *Neurocomputing*, vol. 128, pp. 507–516, Mar. 2014.

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

[20] Y. Xu, "Maximum margin of twin spheres support vector machine for imbalanced data classification," *IEEE Trans. Cybern.*, vol. 47, no. 6, pp. 1540–1550, Jun. 2017.

[21] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–443, 2015.

[22] C. Shi, G. Panoutsos, B. Luo, H. Liu, B. Li, and X. Lin, "Using multiple-feature-spaces-based deep learning for tool condition monitoring in ultra-precision manufacturing," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3794–3803, May 2019.

[23] A. Tsygvintsev, "On the overfly algorithm in deep learning of neural networks," *Appl. Math. Comput.*, vol. 349, pp. 348–358, May 2019.

[24] A. S. M. Alharbi and E. de Doncker, "Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information," *Cognit. Syst. Res.*, vol. 54, pp. 50–61, May 2019.

[25] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, Jul. 2018.

[26] J. Wang, P. Neskovic, and L. N. Cooper, "Pattern classification via single spheres," in *Proc. Int. Conf. Discovery Science*, 2005, pp. 241–252.

[27] G. Gao, J. Yang, X.-Y. Jing, F. Shen, W. Yang, and D. Yue, "Learning robust and discriminative low-rank representations for face recognition with occlusion," *Pattern Recognit.*, vol. 66, pp. 129–143, Jun. 2017.

[28] Y. Tai, J. Yang, Y. Zhang, L. Luo, J. Qian, and Y. Chen, "Face recognition with pose variations and misalignment via orthogonal procrustes regression," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2673–2683, Jun. 2016.

[29] Y. Yu, S. Tang, K. Aizawa, and A. Aizawa, "Category-based deep CCA for fine-grained venue discovery from multimodal data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 4, pp. 1250–1258, Apr. 2019.

GUOWEI YANG received the B.S. and M.S. degrees in mathematics from Jiangxi Normal University, Nanchang, China, in 1985 and 1988, respectively, and the Ph.D. degree from the University of Science and Technology Beijing, China, in 2004. He was a Professor with Qingdao University, China, in 1999. He is currently a Professor with Qingdao University and Nanjing Audit University. He is also the Director of the Artificial Intelligence Society and the Vice Director of extension engineering of Specialized Committee, China Society of Artificial Intelligence. The main research regions are pattern recognition, big data analysis, intelligent information processing, and intelligent audit.

SHAOHUA QI received the B.S. degree in electronic information engineering from the Ocean University of China, Qingdao, China, in 2016. He is currently pursuing the M.S. degree in signal and information processing with Qingdao University, Qingdao, China. From 2016 to 2017, he was a Technical Support Assistant Engineer with ZTE Corporation, Nanjing, China. His research interests include image processing and object detection using neural networks.

TENG YU received the B.S. degree in communication engineering from the Harbin Institute of Technology, China, in 2006, and the Ph.D. degree from Hanyang University, South Korea, in 2015. Since 2015, he has been an Assistant Professor with the School of Electronic and Information Engineering, Qingdao University, China. His current research interests include artificial intelligence, computer vision, and image processing.

MINGHUA WAN received the B.S. degree from the Automated Institute, Nanchang University of Aviation, in 2003, and the M.S. and Ph.D. degrees in pattern recognition and intelligent systems from the Nanjing University of Science and Technology (NUST), in 2007 and 2011, respectively. He is currently an Associate Professor with Nanjing Audit University. He is the author of more than 20 scientific articles in pattern recognition and computer vision. His current research interests include face recognition and detection and image processing.

ZHANGJING YANG received the B.Sc. degree in computer science and education from Nanjing Normal University, in 2004, and the M.Sc. and Ph.D. degrees in computer applications from the Nanjing University of Science and Technology, China, in 2010 and 2014, respectively. He is currently working with the School of Technology, Nanjing Audit University. His current interests are in the areas of pattern recognition, computer vision, and machine learning.

TIANMING ZHAN received the B.S. and M.S. degrees from the School of Math and Statistics, Nanjing University of Information Science and Technology, Nanjing, China, in 2006 and 2009, respectively, and the Ph.D. degree from the School of Computer Science and Engineering, Nanjing University of Science and Technology, in 2013. He is currently an Associate Professor with the School of Information Engineering, Nanjing Audit University. His research interests include hyperspectral image processing, machine learning, and data analysis.

FANLONG ZHANG received the B.S. and M.S. degrees from Liaocheng University, in 2007 and 2010, respectively, and the Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology (NUST), in 2015. He is currently an Associate Professor with the School of Information Engineering, Nanjing Audit University. His research interests include pattern recognition, computer vision, and face recognition in particular.

ZHIHUI LAI received the B.S. degree in mathematics from South China Normal University, in 2002, the M.S. degree from Jinan University, in 2007, and the Ph.D. degree in pattern recognition and intelligence system from the Nanjing University of Science and Technology (NUST), China, in 2011. He was a Postdoctoral Fellow with the Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, China. He has also been a Research Associate, Postdoctoral Fellow, and Research Fellow at The Hong Kong Polytechnic University, since 2010. He has published more than 50 academic articles, including more than 20 articles published on the IEEE Transactions. His research interests include face recognition, image processing and content-based image retrieval, pattern recognition, compressive sense, human vision modelization, and applications in the fields of intelligent robot research.