Knowledge-Based Systems 196 (2020) 105754

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

One-class support vector classifiers: A survey

Shamshe Alam, Sanjay Kumar Sonbhadra *, Sonali Agarwal, P. Nagabhushan

Indian Institute of Information Technology Allahabad, Prayagraj, 211015, India

A R T I C L E I N F O

Article history:
Received 20 January 2019
Received in revised form 6 March 2020
Accepted 7 March 2020
Available online 18 March 2020

Keywords:
One-class classification (OCC)
One-class support vector classifiers (OCSVCs)
Parameter estimation
Feature selection
Sample reduction
Distributed environment

A B S T R A C T

Over the past two decades, one-class classification (OCC) becomes very popular due to its diversified applicability in data mining and pattern recognition problems. Concerning to OCC, one-class support vector classifiers (OCSVCs) have been extensively studied and improved for the technology-driven applications; still, there is no comprehensive literature available to guide researchers for future exploration. This survey paper presents an up to date, structured and well-organized review on one-class support vector classifiers. This survey comprises available algorithms, parameter estimation techniques, feature selection strategies, sample reduction methodologies, workability in distributed environment and application domains related to OCSVCs. In this way, this paper offers a detailed overview to researchers looking for the state-of-the-art in this area.

1. Introduction

The task of classification is an important and admired topic of research in the area of image processing and pattern recognition. Conventionally, a classifier (binary or multi-class) needs at least two well-defined classes to be distinguished and may misclassify testing samples when a dataset suffers from data irregularity problems like class distribution skew, class imbalance, absent features, small disjunct, etc. Specifically, when a class is ill-defined (under-sampled or absent), then the classification model does not work as anticipated. Initially, Minter [1] identified this issue and named it “single-class classification”. Later, Koch et al. [2] called this phenomenon “one-class classification” (OCC). Afterwards, many researchers termed this phenomenon differently based on the application domains to which one-class classification was performed like “novelty detection” by Bishop [3], “outlier detection” by Ritter and Gallegos [4] and “concept learning” by Japkowicz [5]. In OCC task, there are enough target class samples and very fewer outliers, i.e., either negative class samples (class of no interest) are partially available or absent. This property of the dataset makes decision boundary detection a complex and challenging task. It is also witnessed that like conventional classification problems such as measuring the estimation of classification error, the complexity of a solution, the generalization of classification methods and the curse of dimensionality also appear in one-class classification. For one-class classification, several machine learning models have been proposed like one-class nearest neighbour, one-class deep neural network and autoencoder, one-class random forest, one-class support vector classifiers, one-class support higher-order tensor machine, one-class ensemble model, etc. [6–11].

Based on an extensive literature analysis, one-class support vector classifiers (OCSVCs) are found suitable for anomaly and novelty detection in numerous applications such as document classification [12], disease diagnosis [13,14], fraud detection [15,16], intrusion detection [17,18] and novelty detection [19]. These varied applications of OCSVC make this classifier interesting and important in the field of data mining and pattern recognition. Though several research articles have been published concerning OCSVCs during last two decades, comprehensive literature is still not available that covers all important issues to help research community for further developments. This review paper summarizes and shelters all important issues concerning to OCSVCs. Feasible algorithms, feature selection, training sample reduction, parameter estimation, workability over distributed/streaming data and related application areas are identified as key issues. Fig. 1 represents the clustered representation of relevant research works, where the notations from ‘a’ to ‘f’ represent the following:
(a) OCSVC algorithms,
(b) Parameter estimation techniques,
(c) Feature selection methods,
(d) Training sample reduction methods,
(e) Distributed and online OCSVCs,
(f) Applications of OCSVCs.

Rest of the paper is organized as follows: Section 2 contains OCSVC algorithms and in Section 3 kernel parameter estimation techniques have been discussed. Section 4 describes a detailed review of feature selection methods, whereas Section 5 gives a

* Corresponding author.
E-mail address: rsi2017502@iiita.ac.in (S.K. Sonbhadra).

https://doi.org/10.1016/j.knosys.2020.105754

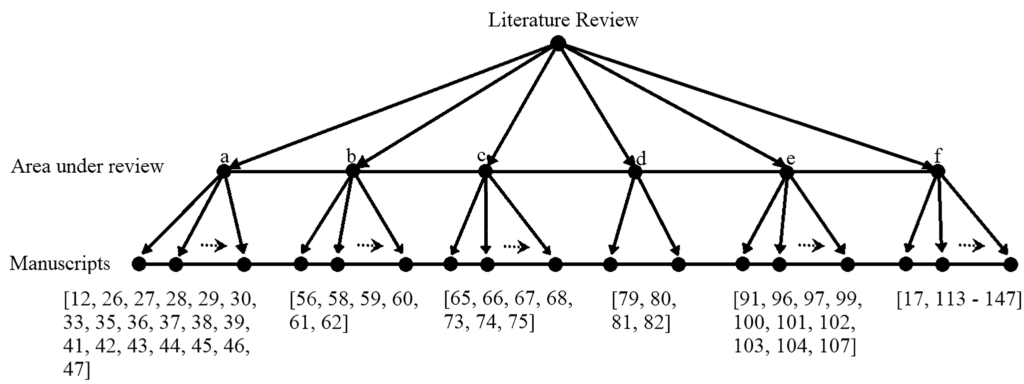


Fig. 1. Clustering of the reviewed literature according to their relevance.

review of sample reduction techniques. Distributed OCSVCs have been covered in Section 6, and Section 7 discusses the applications of OCSVCs. The last section contains concluding remarks and future scope.

2. OCSVC algorithms

The support vector machine (SVM) was introduced by Vapnik [20] and mainly used for tackling binary classification problems. Later, several extensions of SVM were proposed by researchers like least-squares SVM, linear programming SVM, sparse SVM, twin SVM, Universum SVM, twin spheres SVM etc. [21–25]. The one-class classification problem was solved by Tax et al. [26–30] via isolating the target class samples from outliers in sample space. In these research articles, it is witnessed that one-class classification suffers from complications akin to other conventional classification methods. Hence, one-class classification problems are established as standard classification tasks like the solution's complexity, estimation of classification error, etc. As a solution to these difficulties, a method called support vector data description (SVDD) has been proposed, where a hypersphere encloses all target class (class of interest) samples, and the boundary points (the data points nearby the decision boundary) of hypersphere are support vectors. The SVDD rejects a test sample as an outlier if it falls outside of the hypersphere; otherwise, accepts it as a target class sample, as shown in Fig. 2. The SVDD is defined as follow:

$$L(R, a, \alpha_i, \gamma_i, \xi_i) = R^2 + C\sum_i \xi_i - \sum_i \alpha_i\{R^2 + \xi_i - (\parallel x_i \parallel^2 - 2a.x_i + \parallel a \parallel^2)\} - \sum_i \gamma_i \xi_i \tag{1}$$

subject to : $\parallel x_i - a \parallel^2 \leq R^2 + \xi_i,\ where\ \ \xi_i \geq 0\ \ \forall$

where R represents the radius of the hypersphere, x_i is a negative class sample (outlier), boundary dots are support vectors, and slack variable ξ penalizes the outlier. The objective is to minimize the volume of hypersphere (i.e. minimize R) so that it covers all target class samples with the penalty of slack variables for the negative class data points so that there can be increase in classification accuracy. In these research articles, it is also observed that description radius R is inversely proportional to the number of support vectors which makes decision boundary loose or tight.

By introducing kernel functions, the hypersphere of SVDD can become more flexible. After extensive experiments performed on different datasets, Tax [26] observed that generally, the Gaussian kernel outperforms over the polynomial kernel and resulted in tighter descriptions, but needs more data for a stretchy boundary. In this research, experiments have been performed with different

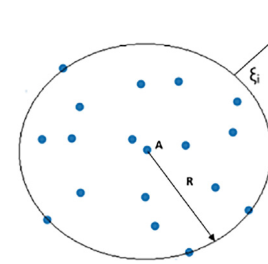


Fig. 2. The hypersphere covering the target data samples.

values of the Gaussian width parameter s and concluded that larger the kernel width, fewer support vectors are selected that leads to the more spherical description as shown in Fig. 3. In Figs. 3a, 3b and 3c, big dots represent the support vectors for different kernel widths. Tax demonstrated the feasibility of the proposed method on machine fault diagnostic and handwritten digit datasets. This method is inefficient for the high dimensional datasets and also starts refusing the low-density target samples as outliers when there are huge variations in density among the positive class samples. To make this technique adaptive in high dimensional feature space, Tax and Duin [31] proposed an approach to generate artificial outliers evenly in a hypersphere by transforming objects generated from a Gaussian distribution. This research suggested that the method to generate outliers in a hypersphere artificially is suitable for up to thirty dimensions.

To solve OCC problems, Schlökopf et al. [32,33] proposed an alternative one-class classification approach different from SVDD, where a hyperplane is constructed instead of hypersphere (Fig. 4). This hyperplane separates the target class data points with the maximal margin from the origin, where all the outliers are assumed to fall on the plane through the origin. This model is known as one-class support vector machine (OCSVM). This algorithm returns a function g(x) that takes the value +1 for the target class region and −1 elsewhere. It maps the data into the feature space corresponding to the kernel to separate them from the origin with maximum margin. For a new data point x, the value of the function g(x) is determined by evaluating the side of the hyperplane it falls in feature space. To separate the target class samples from the origin, following quadratic equation must be solved:

$$\max_{w,\xi,\rho} \frac{1}{2} \parallel w \parallel^2 + \frac{1}{\upsilon N}\sum_i^n \xi_i - \rho \tag{2}$$

subject to: $w.\phi(x_i) \geq \rho - \xi_i$ and $\xi_i \geq 0$ for all $i = 1, 2, 3 \ldots n$

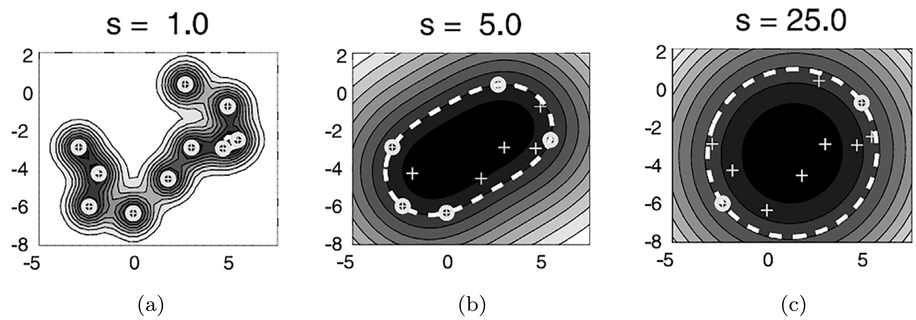


Fig. 3. Surface and support vectors [27].

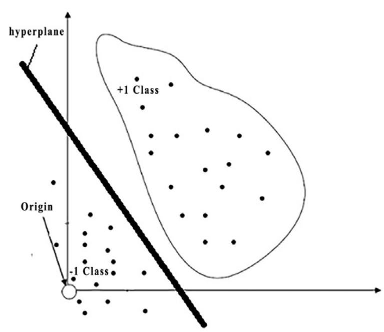


Fig. 4. Hyperplane in one-class support vector machine.

where ϕ is the function to represent a point x_i in feature space, and slack variable ξ_i penalizes the outlier. The objective is to find a hyperplane characterized by ω and ρ which has maximal distance from the origin in feature space and separates all target class samples from the origin. $\upsilon \in (0,1]$ is the parameter that sets an upper bound on the fraction of outliers and lower bound on the number of support vectors. It is evident in this article that the Gaussian kernel performs better than all other kernels.

Both the revolutionary methods for one-class classification (SVDD and OCSVM) performed equally with a Gaussian kernel, where origin plays a decisive role. In OCSVM, all the negative class data points are pre-assumed to lie on the origin. This phenomenon is a major drawback, observed by Campbell et al. [34]. Unlike Schlökopf et al. [32], Manevitz et al. [12] treated outliers as representative of the second class and assumed that along with origin all nearby samples to origin are outliers (Fig. 5a). Geometrically, the data vectors of all sub-spaces are to be treated as outliers. To evaluate the results, Reuters dataset has been used along with linear, sigmoid, polynomial and radial basis kernels. The experimental results were disappointing over OCSVM proposed by Schlökopf et al. [32], but performs better when number of text categories increases. Later, Li et al. [35] presented an improved version of OCSVM proposed by Schlökopf et al. [33] for intrusion detection with better accuracy. This method considered all data points closer to the origin as outliers along with origin (Fig. 5b).

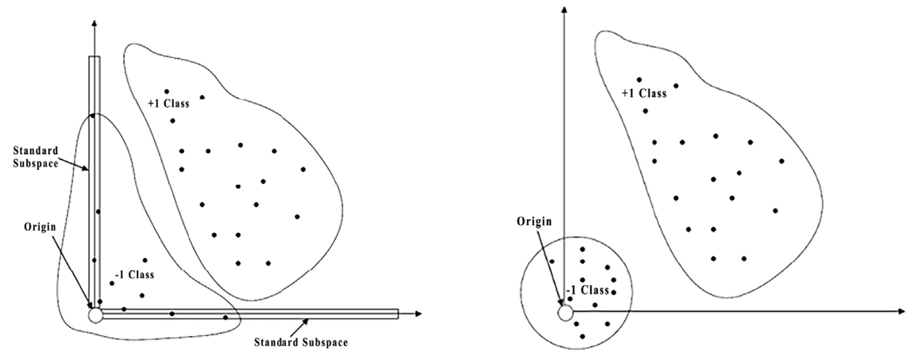
The work of Tax et al. [31] was extended by Luo et al. [36] who proposed cost-sensitive OCSVC algorithms named frequency-based SVDD (F-SVDD) and write-related SVDD (W-SVDD) for intrusion detection problem. Unlike SVDD, F-SVDD and WS-SVDD provide dissimilar costs to processes than to system users. This approach of dissimilar cost allotment results in better intrusion detection performance while compared to conventional SVDD.

In the race of enhancement of OCSVCs, Hao et al. [37] extended the work of Schlökopf et al. [33] with the fuzzy integration. In this approach, for each training data, fuzzy membership has been applied to reformulate OCSVM. In the real world, training samples come from several sources with dissimilar feature sets; hence, the fuzziness factor is essential for all the training data points to represent the maximum fuzziness degree. Experiments performed on the handwritten digit dataset exhibited good classification accuracy, specially for pretended similar classes (for example, digit 2 and 7 due to various writing styles). Later, the least squares one-class SVM (LS-OCSVM) proposed by Choi [38] reduces the time complexity from $\mathcal{O}(N^3)$ to $\mathcal{O}(N^2)$ by replacing the inequality constraints with equality constraints in QP. Despite LS-OCSVM has fast training speed, its testing speed is slow because it gets several support vectors when solving QP.

Seeking for enhanced OCSVC, Yang et al. [39] proposed a neighbourhood-based OCSVM method for functional magnetic resonance imaging (fMRI) data to identify schizophrenia. This approach integrates the neighbourhood consistency hypothesis [40] along with OCSVM to compute primal values that denote the distance between data points and hyperplane in kernel space. Based on primal values and their neighbours, a new decision value is computed for each voxel of a fMRI image. A voxel is said to be activated if the decision value is higher than the given threshold else it is non-activated. Experimental results show that it gives more stable results than k-Means and fuzzy k-Means clustering algorithms.

Lui [41] extended the models proposed by Schlökopf et al. [33] and Hao et al. [37], and proposed a robust solution to fuzzy one-class SVM. In this research, by introducing robustness and fuzziness parameters, the reformulation of one-class SVM has been done. An association between fuzzy membership and boundary perturbation has been analysed and proved that for a given perturbation bound, the fuzzy membership has a lower bound regardless of the value of the regularization parameter of OCSVM. The input data in the real world is originated from various distributed sources, so training samples must be treated differently by assigning different fuzzy factors, and this method provides effective discrimination for the input data to improve the performance of OCSVM significantly.

In conventional OCSVM, every training object is treated equally and due to this fact, it is very sensitive to noise. Observing this, Zhu et al. [42] extended the work of Tax et al. [27] and Schlökopf et al. [33], and proposed a novel model, i.e., weighted OCSVM (W-OCSVM). In W-OCSVM, a weighting strategy has been applied to the training samples, where the boundary points are assigned higher weights, and core data points of distribution are assigned lesser weights. This weighting mechanism helps to reduce noise sensitivity. In this approach, a neighbourhood sphere has been used as shown in Fig. 6, and weight is assigned to a point according to the neighbourhood distribution. For a data point, if a substantial number of neighbours are in one half of the sphere, then it is concluded as a nearby decision boundary point; therefore, more weight should be given to the data point; otherwise it is a non-boundary point and must be assigned with



(a) Origin and subspace samples are outliers [12].

(b) Origin and nearby samples are outliers [35].

Fig. 5. Different assumptions for outliers.

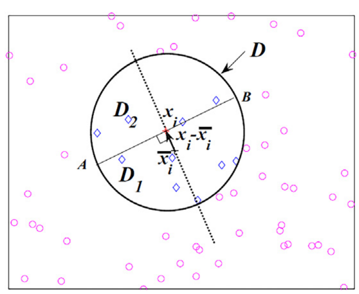


Fig. 6. The D_1 and D_2 are two equal parts of neighbourhood sphere D, the neighbours of x_i are denoted as diamonds, x_i is denoted as asterisk, AB is normal to $x_i - \bar{x}_i$. [42].

lesser weight. In this approach, only the neighbour's distribution knowledge is sufficient to determine the weight of a point and the results obtained were significantly better than standard OCSVM. Later, Ji et al. [43] extended the standard SVDD given by Tax et al. [27] and proposed adaptive weighted one-class SVM (A-WOCSVM) which is stouter than W-OCSVM proposed by Zhu et al. [42] due to enriched anti-outlier ability. This approach uses samples' local density and with their distances from the origin to describe the behaviour of SVDD.

The support vector data description minimizes the volume of hypersphere around target data. It gives a satisfactory performance but leads to a loose boundary for multivariate datasets [48]. To make one-class classifier more robust, Wang et al. [44] proposed ellipsoidal data description (ELPDD), where minimum volume enclosing ellipsoid (MVEE) [49,50] was used with consideration of dimensional variance and covariance of the dataset. MVEE was constructed around target class in the kernel principal component analysis (k-PCA) [51] subspace, and the resulting model can be learned via SVM-like objective function with a log-determinant penalty instead of the L2-norm. Experimental results show that the ELPDD gives good accuracy but time complexity is significantly high.

It is observed that for anomaly detection, OCSVCs are the best suitable one-class classifiers but very sensitive to the outliers or noise. To overcome this problem, Tian et al. [45] introduced ramp loss function [52] to the conventional OCSVM and proposed Ramp-OCSVM. This model is more robust than conventional OCSVM due to the non-convexity properties of the ramp loss function. Extensive experiments have been performed to show the effectiveness of the proposed model and results show that the Ramp-OCSVM outperforms over the conventional OCSVM. To solve the problem of sensitivity to the outliers, Xing et al. [46] proposed more robust OCSVM based on the rescaled hinge loss function [53]. The hinge loss function is unbounded; therefore, the larger loss caused by outliers cannot be prevented, which may reduce the anti-outlier ability and generalization performance of OCSVM. Whereas, the rescaled hinge loss function is bounded after rescaling. This model shows better performance over conventional OCSVM.

Most recently, Xing et al. [47] proposed a robust Adaboost based OCSVM inspired by Adaboost based variations of SVMs [54,55]. Adaboost is an iterative machine learning algorithm to construct a classifier ensemble. It uses only one classification algorithm to construct diverse weak base classifiers, which are trained on datasets selectively sampled from an initial training dataset. To make Adaboost fit for combining OCSVMs and achieve robustness, a weighted combination based loss function is designed to substitute the exponential loss function in the conventional Adaboost. Experiments were performed on twenty UCI datasets, and results show that the proposed robust Adaboost based OCSVM outperforms over state of the art.

From the preceding literature on various OCSVC algorithms, it is observed that many authors have extended the work proposed by Tax et al. [26–28] and Schlökopf et al. [33], and worked on the training data before feeding to model. Manevitz et al. [12] treated outliers as representative of the second class and assumed that along with origin all nearby data points (in lower-dimensional sub-space) to origin are considered as outliers. Later, Li et al. [35] considered all data points closer to the origin as outliers, whereas Luo et al. [36] used F-SVDD and WS-SVDD to assign dissimilar costs to processes than to system users.

Hao et al. [37] proposed a fuzzy membership factor to each training sample so that each data point will be of different importance for better data description. Later, Yang et al. [39] proposed a neighbourhood-based OCSVM method for fMRI data to identify schizophrenia. This approach integrates the neighbourhood consistency hypothesis [40] along with OCSVM to compute primal values which denote the distance between data points and hyperplane in kernel space. Whereas, Yong Lui [41] reformulated the fuzzy OCSVM and concluded that there is always a lower bound of fuzzy membership value for a training data regardless the value of regularization parameter that helps to improve the classification accuracy. Zhu et al. [42] proposed a new approach in which the training data points outside of boundary should be given less weight, whereas points inside the boundary should be given higher weight to make data description tight and more

Table 1
Summary of OCSVC algorithms.

Citation	Methodology/Specification	Achievements
Tax et al. [26–30]	Support vector data description (SVDD) around a class of interest or target class.	The data description methodology shows better results for sparse and complex data.
Schlökopf et al. [33]	In place of the hypersphere, a hyperplane separates target class samples with the maximal margin from the origin.	Kernel trick allows softness in decision rule and experimental results are comparable to [27].
Manevitz et al. [12]	With origin, all nearby data points (in lower dimension subspace) to origin are considered as outliers.	Gives better results compared to [33] when the number of categories increases.
Li et al. [35]	Considered all data points closer to the origin as outliers.	Performance is better for anomaly detection in the intrusion detection system compared to [33].
Luo et al. [36]	Proposed cost-sensitive algorithms named F-SVDD and WS-SVDD to provide dissimilar costs to processes than system users.	Better intrusion detection performance compared to [31].
Hao et al. [37]	Fuzzy membership to data points.	Good classification accuracy for handwritten character recognition compared to [33].
Choi [38]	Least Square OCSVM.	Reduces time complexity from $\mathcal{O}(N^3)$ to $\mathcal{O}(N^2)$.
Yang et al. [39]	Integrates the neighbourhood consistency hypothesis [40] along with OCSVM to compute primal values which denote the distance between data points and hyperplane in kernel space.	Achieved good classification accuracy for diagnosis of schizophrenia with fMRI.
Lui [41]	By introducing robustness and fuzziness parameter, reformulation of one-class SVM has been done.	Provides effective discrimination for the input data and achieves good classification accuracy.
Zhu et al. [42]	Data points nearer to the decision boundary assigned with higher weights.	Classification accuracy is better than SVDD [27].
Ji et al. [43]	Samples' local density and with their distances from origin to describe the behaviour of SVDD.	Enhanced anti-outlier ability.
Wang et al. [44]	Ellipsoidal data description (ELPDD).	Gives better accuracy compared to SVDD.
Tian et al. [45]	Introduced ramp loss function to the conventional OCSVM and proposed Ramp-OCSVM.	This model is more robust compared to the conventional OCSVM due to the non-convexity properties of the Ramp loss function.
Xing et al. [46]	Proposed more robust OCSVM based on the rescaled hinge loss function.	This model outperforms over conventional OCSVM.
Xing et al. [47]	Adaboost based OCSVM.	It outperforms over state of the art.

accurate. Afterwards, Ji et al. [43] proposed a new strategy for giving weight to training points and stated that along with distance, density should also be considered as a second key factor to calculate a resultant weight to a data point. Later, Wang et al. [44] proposed ellipsoidal data description (ELPDD) as more robust OCSVC compared to conventional SVDD. Meanwhile, Tian et al. [45] introduced Ramp-OCSVM which is better than the conventional OCSVM. After that, two enhanced OCSVMs were proposed based on rescaled hinge loss function and Adaboost by Xing et al. [46,47]. A relative study of the developments of the above literature is shown in Table 1.

3. Kernel parameter estimation techniques for OCSVCs

Kernel techniques provide more flexibility to one-class support vector classifiers. Tax et al. [27] and Schlökopf et al. [33] showed that Gaussian kernel outperforms than any other kernel due to single tuning parameter, i.e., Gaussian width, which normalizes the number of support vectors of the separation boundary. In contrast, the increase in width parameter increases the volume of the enclosed region and decreases the number of support vectors. The denial rate of OCSVC is measured by the upper bound on the fraction of training points outside the enclosed region and the lower bound of the number of support vectors, which is generally user-specific. The selection of the values for the kernel width and the regularization parameters has a key impact on OCSVM or SVDD [56,57]. Tax et al. [56] proposed a heuristic approach to estimate these parameters for SVDD in a fully unsupervised manner. Their proposed heuristic optimizes the estimated false-positive and false-negative rates by solving the following minimization problem:

$$\Lambda(s, \upsilon) = \frac{\#SV}{N} + \lambda\ \#SV_b \left(\frac{s}{s_{max}}\right)^d \tag{3}$$

where s is kernel width parameter and υ is the regularization parameter. SV and s_{max} represent the set of support vectors and the maximum distance in the training set respectively. SV_b is the set of support vectors (i.e., those with $0 < \alpha_i < C$), and λ is a regularizer. The first term ($\frac{\#SV}{N}$) is an estimate of the error on the target class, and the second term controls the error on the outlier class. Since the RBF kernel has been used for this heuristic, it is also possible to use the same parameter setting to train an OCSVM.

For a kernel-based one-class classifier, Zhuang et al. [58] proposed a three-tier framework for parameter optimization. The first stage is known as the learning stage, where the classifier was trained with only target class samples only, with a randomly chosen parameter, whereas in the second stage (evaluation stage), overall performance was measured based on both target and outlier data points considering all training samples. In the last stage (optimization stage), a uniform grid was defined in the parameter space, and each grid point has been evaluated for the global optimum. The experimental results proved that the proposed framework performed better compared to standard OCSVM and SVDD; but, computationally this framework is not efficient due to the exhaustive nature of grid search.

Wang et al. [59] proposed a heuristic approach to tune the Gaussian kernel parameter via boundary tightness detection to detect whether the decision boundary is over-fit or loose, and iterates until an optimal boundary is obtained or the difference between S_u and S_l is less than a given threshold. Following equations have been proposed for the calculation of S_u, D_u, S_l, and D_l after the number of experiments.

$$S_u = 0.4103 D_u^2, D_u = \max_i \max_j \parallel X_i - X_j \parallel \tag{4}$$

$$S_l = 0.4103 D_l^2, D_u = \max_i \max_{j \neq i} \parallel X_i - X_j \parallel \quad (5)$$

where x_i and x_j denote any two training samples. At each iteration $S = (S_u + S_l)/2$ is calculated and if the boundary is found tight then assign $S_l = S$ otherwise $S_u = S$ and stop the iteration when the difference $(S_u - S_l)$ is less than the given threshold. The final value of S is the optimal Gaussian parameter value. The proposed algorithm is efficient and ensures the proper tightness of the decision boundary, but unable to deal with a condition when the boundary is satisfying both the criteria (tight as well as loose). As a solution, an algorithm named detecting tight or loose (DTL) was proposed by Xiao et al. [60] in which S is set to S_l when both loose and tight condition found. In this paper, the authors proposed two approaches to the Gaussian kernel parameter estimation for OCSVM. The first one is the distance farthest nearest (DFN), which uses the information of the farthest and the nearest neighbours of each sample. The objective function for DFN is given by:

$$f_o(s) = \frac{1}{n} \sum_{i=1}^{n} \max_j \parallel \phi(x_i) - \phi(x_j) \parallel^2 - \frac{1}{n} \sum_{i=1}^{n} \max_{j \neq i} \parallel \phi(x_i) - \phi(x_j) \parallel^2 \quad (6)$$

This Eq. (6) is to be maximized to get optimal Gaussian parameter s, where ϕ is the function that represents a point x_i in feature space. The first term is the measure of separation of the class of interest and outlier, whereas the second term measures inner class separation. Xiao et al. [60] also compared DFN with other indirect methods like variance–mean (VM) and max-distance (MD). VM measures the dispersion of the kernel matrix and the Gaussian parameter s. This dispersion is obtained by maximizing the formula $VAR^2/(MEAN + \epsilon)$, where VAR and $MEAN$ are variance and mean of non-diagonal entries of kernel matrix respectively and ϵ is small positive constant. VM method usually tends to over-fit the decision boundary as shown in Fig. 7b. The max-distance method (MD) calculates the Gaussian parameter by the maximum distance of positive samples, which usually tends to lose the decision boundary and under-fits the decision boundary, as shown in Fig. 7a.

However, DFN ignores a portion of the mid-region, so it misdeals with the multi-mode dataset and attains unsatisfactory results. Since the DFN method is unable to make complete use of the sample distribution information, Xiao et al. [60] proposed another method named as DTL, which optimizes parameter s using the feedback of OCSVM until it converges. DTL detects the tightness of the decision boundary through a set of organized rules, and tunes iteratively. If inside the decision boundary, one or several holes are detected whose interior contains no training sample, then the boundaries are referred to be loose, whereas if the boundaries near two neighbouring samples are concave or separated, then they are judged to be tight. The schematic representation is shown in Fig. 8 before and after applying the DTL algorithm.

These two proposed approaches are useful to select suitable Gaussian kernel parameter, enabling the resulting OCSVM model to outperform the conventional one. The DFN exhibits low computation complexity and extensively used, but it is unable to use valuable information of the OCSVM model adequately, whereas DTL uses feedbacks from OCSVM models to tune parameter s. The authors compared the performance of these methods for the selection of the Gaussian kernel parameter with ten UCI datasets and concluded that DFN attains higher precisions than DTL and VM.

Table 2
Summary of kernel parameter estimation techniques.

Citation	Methodology/Specification	Achievements
Tax et al. [56]	A heuristic method was proposed to estimate parameters in unsupervised manner.	Applicable to both SVDD and OCSVM and gives better results.
Zhuang et al. [58]	A three-stage framework is used with grid search and cross-validation.	Proposed framework perform better compared to standard OCSVM.
Wang et al. [59]	A heuristic approach to choosing kernel width via tightness detecting algorithm.	Achieves optimal decision boundary.
Xiao et al. [60]	DFN and DTL are used to estimate the Gaussian parameter.	DFN and DTL outperform over MD and VM.
Theissler et al. [61]	k-fold cross-validation has been used for autonomous tuning of the SVDD parameter.	User intervention not required.
Ghafoori et al. [62]	Proposed two parameter estimation algorithms, namely quick model selection (QMS) and revised duplex max-margin model selection (RDMMS)	This models give better performance compared to the other pre-existing heuristic methodologies.

Inspired by the fact that kernel parameters are data-driven, Theissler et al. [61] proposed an autonomous method for parameter estimation in SVDD with RBF kernel from the training set. The regularization parameter influences the volume of description, and the kernel width determines the number of support vectors. These two parameters define the tightness of the boundary. The goal of the proposed methodology is to determine parameters exclusively based on the training set of target instances without user participation. The radius for the RBF kernel is given by:

$$R^2 = 1 - 2 \sum_{i=1}^{M} \alpha_i K(\vec{x}_b, \vec{x}_i) + \sum_{i,j=1}^{M} \alpha_i \alpha_j K(\vec{x}_i, \vec{x}_j) \quad (7)$$

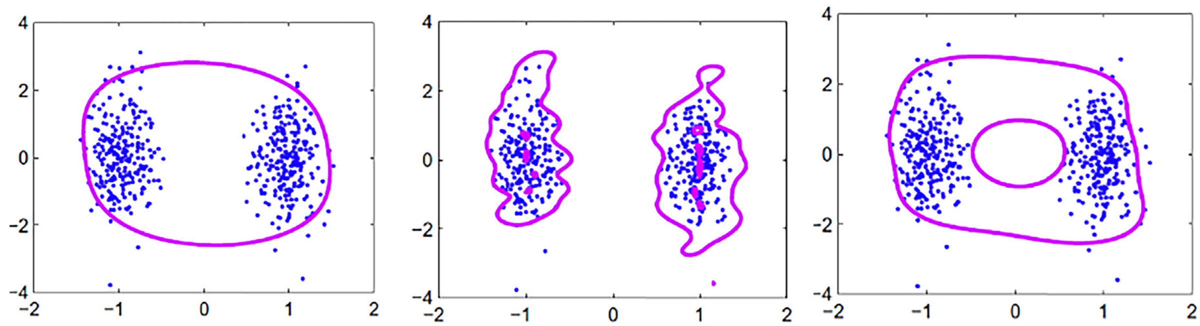
where x_b is the support vector at the boundary and x_i is a data point. Since the goal is to minimize the error on the target class, R is desired to be close to 1. Correspondingly, weighting error rate λ_i and radius give rise to find the pair $\{e_i, R_i\}$ closest to the origin by minimizing the following function:

$$\lambda_i = \sqrt{e^2_{\omega_{n_i}} + |1 - R_i|^2} \qquad \forall i \quad (8)$$

For each tuning step, k-fold is used to determine the error rate and radius R. The training set is randomly split into k folds: 1 validation set and $k-1$ training sets, and the error e_{ω_n} and R are averaged over the k runs.

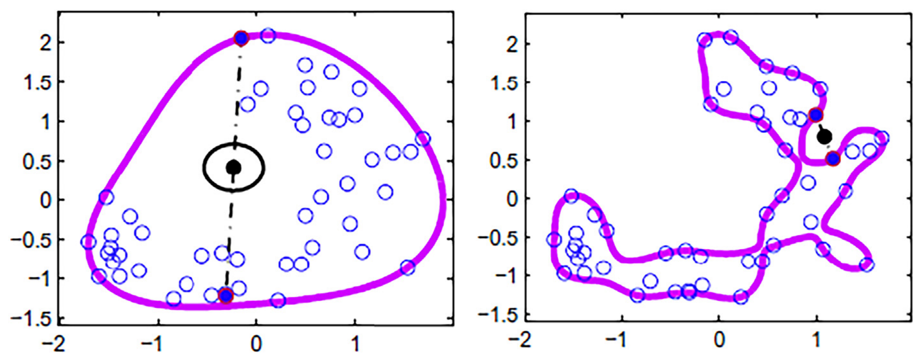
Recently, Ghafoori et al. [62] proposed two unsupervised methods for estimating the optimal parameter settings to train OCSVM and SVDD models, based on analysing the structure of the data. In this research, the authors proposed two parameter estimation algorithms, namely quick model selection (QMS) and revised duplex max-margin model selection (RDMMS), for the OCSVM and SVDD algorithms to estimate optimal parameter settings without any need for ground truth labels or exhaustive grid-search over the parameter space. It is observed that the proposed models give better performance compared to the other pre-existing heuristic methodologies.

From the preceding literature about the various methodologies for one-class SVCs parameter tuning, it has been observed that the Gaussian kernel gives the most promising result. Tax



(a) MD. (b) VM. (c) DFN.

Fig. 7. Comparison of DFN from VM and DFN [60].



(a) Before applying DTL. (b) After applying DTL.

Fig. 8. Schematic for DTL [60].

et al. [56] proposed a heuristic approach to estimate these parameters for SVDD in a fully unsupervised manner, whereas a three-tier parameter estimation framework was proposed by Zhuang et al. [58], which is better in terms of computational complexity. Afterwards, Wang et al. [59] proposed an iterative algorithm that is more robust than [58], and gives an optimized width parameter, that is assumed to be obtained after getting a decision boundary with optimized tightness. Xiao et al. [60] proposed two methods for determining kernel parameter: DFN and DTL, where the second one is the improvement in the algorithm proposed by Wang et al. [59]. The DFN method is unable to make complete use of the sample distribution information, whereas DTL optimizes parameter s using the feedback of OCSVM until it converges. All the above methods require the regularization parameter to be set by the user and require an initiating value. Whereas, Theissler et al. [61] proposed a methodology that is autonomous regarding parameter determination for SVDD with RBF kernel. Recently, Ghafoori et al. [62] proposed two unsupervised methods for estimating the optimal parameter settings to train OCSVM and SVDD models, based on analysing the structure of the data. These models give better performance compared to the other pre-existing heuristic methodologies. Table 2 shows a comparative outcome of preceding literature.

4. Feature selection and reduction techniques for OCSVCs

Just like the conventional binary and multi-class classifiers, data preprocessing is very important for OCSVCs. More specifically, it includes data cleaning, class balancing [55,63,64] (not necessary for OCSVCs if the target class is well defined), feature selection/transformation (dimensionality reduction) and sample reduction (also known as data reduction). There is a deep impact of feature selection/reduction mechanism on the performance of a classifier. In a dataset, not all features are discriminating and the presence of such features may reduce the performance of the classifier. The presence of massive features may also introduce noise, and the classifiers can easily over-fit due to the curse of dimensionality. As a solution to this, the feature selection/reduction is an important pre-processing task to fit a classifier with the minimum number of features. An effective feature selection mechanism reduces the dimension of samples and reduces the overall computation cost. Many dimensionality reduction algorithms have been proposed for conventional classifiers, whereas for one-class SVCs, limited number of papers have been published till date. Dimensionality reduction can be performed either with discriminative feature selection or feature transformation. Tax et al. [65], Lian [66], Liu et al. [67] and Feng et al. [68], demonstrated different variations of feature selection mechanism as dimensionality reduction approaches with principal component analysis (PCA) [69] for OCSVC. In general, PCA extracts the eigenvectors associated with top eigenvalues as projection direction and discards the lower eigenvalued eigenvectors, but Tax et al. and Lian et al. proved that eigenvectors associated with small eigenvalues might play a key role in classification.

Tax et al. [65] found that there is trade-off [70] in feature extraction and demonstrated that for one-class classification, the low-variance directions are most informative. In this context, researchers assumed a uniform distribution of outliers and Gaussian distribution for target class, and discussed the behaviour of SVDD on Gaussian target distribution followed by effect of applying PCA. For experiments, two datasets were used: face dataset [71] and Concordia dataset [72]. Experimental results show that for high sample sizes and large differences in the variances of the target distribution, low variance principal components exhibit smaller error. This is caused by the fact that the error on the outliers mainly depends on the eigenvalues of the estimated target covariance matrix S_n, and the error on the target class on the eigenvalues of $S_n \sum^{-1}$, where $\sum$ is the true target covariance matrix.

Later, Lian [66] demonstrated the use of lower eigenvalued projection direction of PCA using USPS database of handwritten digits for OCSVM [33]. Whereas, Liu et al. [68] proposed a novel method for anomaly detection in visual surveillance system using motion directional PCA and OCSVM. In this research, PCA is applied to each separate direction of the features (keeping 95% energy) instead of applying directly to high-dimensional features. These PCAs, in different directions, were independent. In motion directional PCA, every PCA packs the motion vector statistics features in the same direction into a smaller feature vector.

The indicator diagram plays an important role in health monitoring and fault diagnosis of reciprocating compressors. A novel approach for indicator recognition and novelty detection has been proposed by Feng et al. [68], where an indicator diagram is transformed into a high-dimensional feature space using discrete 2D-curvelet transform. Then, nonlinear PCA is implemented to map the high-dimensional features to three-dimensional feature space under the multi-class recognition situation, whereas PCA is applied for feature reduction under the one-class situation. Here, the multi-class SVM and one-class SVM are taken as the classifiers and novelty detector respectively. Experimental results show that the proposed approach performed better than the traditional wavelet-based approach.

Later, Jeong et al. [73] proposed two feature selection approaches for SVDD. The first one is the SVDD-radius-recursive feature elimination (SVDD-radius-RFE) method that trains the classifier and uses the square of SVDD radius as the criteria function for each feature. This method reduces the size of the boundary by iteratively reducing the feature set based on the consequence of eliminating an individual feature in criteria function, whereas the second method SVDD-dual-objective-RFE, maximizes the dual function of SVDD as criteria, i.e., the Lagrange's multipliers. Both of these methods produce similar results.

Afterwards, Nagabhushan et al. [74] proposed a three-tier target class supervised feature sub-setting for dimensionality reduction. In the first stage of the proposed approach, less important features are rejected, whereas in the second stage, redundancy is eliminated by rejecting redundant features and in the last stage, the optimum feature set is derived from the subsequent subset of features. These operations remove the unnecessary features to minimize the intra-class variance and ensure that the obtained features are optimal to represent a positive class.

Yousef et al. [75] experimented various feature selection strategies with OCSVC and found that classification accuracy can be increased if feature selection methodology applied to the training data. In this paper, eight feature selection methodologies have been used and found 30% average variance in the accuracy with different feature selection approaches. The datasets used in the experiments are available microRNA from selected plant species from miRBase [76]. Among eight feature selection algorithms, low information gain (LIG) selects the desired number of features with the lowest information iteratively, whereas random feature selection (RFS) randomly selects the features. Random feature from feature clusters (RFC) selects one feature per cluster randomly from feature clusters. Selecting features from clusters (SFC) selects up to three clusters and uses all features within. Selecting features with high information gain (HIG) chooses features with higher information per dataset and high information gain from feature cluster (HIC) selects one feature with the highest information per feature cluster from higher information features set. Zero-norm feature selection (ZNF) removes the features whose vector has zero values and defines the zero-norm to be the nonzero values for all positive examples. Using the number of nonzero values in a vector and threshold, it determines the relevance of a feature and removes the ones below a predefined threshold value. Pearson correlation-based feature selection (PCF) selects the features with low correlation. With experiments, it has been observed that among all feature selection methods, SFC and HIC outperform for all datasets. However, performance varies with different datasets and the variation is identical for all feature selection methods. The average accuracy of all feature selection methods can further help to define dataset quality. This research shows good variation and influence of feature selection methodologies for OCC.

Table 3
Summary of feature selection and reduction techniques for OCSVCs.

Citation	Methodology/Specification	Achievements
Tax et al. [65], Heng Lian [66], Liu et al. [67] and Feng et al. [68]	Variations of PCA. Low variance principal components are also important.	Better classification accuracy.
Jeong et al. [73]	SVDD-radius-RFE and SVDD-Dual-Optimization-RFE.	Both the methods give better accuracy and data description.
Nagabhshan et al. [74]	Feature Subsetting.	Minimizes intra-class variance and obtain optimal feature set that gives better classification accuracy over conventional OCSVM.
Yousef et al. [75]	Compared eight feature selection methodologies: SFC, HIC, RFS, RFC, HIG, ZNF, PCF and LIG.	Better classification accuracy if a good feature selection algorithm used.

The preceding literature justifies the importance of feature selection methodologies for one-class SVCs. Tax et al. [65], Lian [66], Liu et al. [67] and Feng et al. [68] showed that eigenvectors associated with lesser eigenvalues could be of immense importance, whereas eigenvectors associated with higher eigenvalues may be ignored. Afterwards, Jeong et al. [73] proposed two new methods for feature selection based on specific criteria function, whereas Nagabhshan et al. [74] proposed target class supervised feature sub-setting to obtain an optimal subset of features. Yousef et al. [75] experimented microRNAs of plant and found that if feature selection methodologies are used then up to 30% increased classification accuracy can be achieved. As a concluding remark, this section shows that there is no best-known feature selection method for OCSVCs. However, there are enough possibilities for improvements over existing methods. A comparative study of the outcome of the preceding lecture is shown in Table 3.

5. Sample reduction techniques for OCSVCs

The existing training sample reduction techniques concerning OCSVCs are discussed in this section. In the present era of big and stream data, enormous data is produced from geologically distributed data sources. In the presence of massive training samples, the classifiers may suffer from uneven computational cost and unusual resource consumption during the training phase. Training sample reduction is a solution to enhance the training efficiency and reduce computation cost. Many training sample reduction methods have been offered for SVMs [77,78], but still, satisfactory work has not been done for OCSVCs.

Tax et al. [27] showed that support vectors are boundary points of the data distribution, and considering this phenomenon, Li [79] proposed an approach for training sample reduction that chooses the most promising exterior training samples among all samples. Fig. 9 shows that almost all nearest neighbours of any boundary point reside on one side of the tangent plane passing

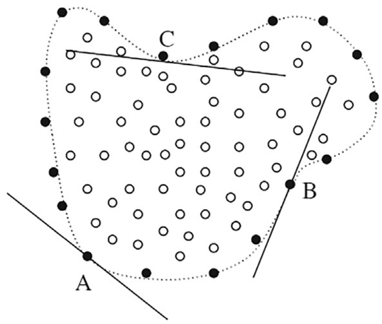


Fig. 9. Concave and convex surfaces [79].

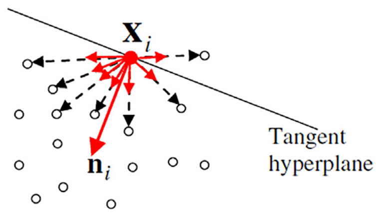


Fig. 10. Normal vector n_i of the tangent plane for a given point x_i [79].

through it, where black dots are extreme boundary points, and lines represent tangent planes passing through these boundary points. The dotted line indicates the space surface of data distribution. This method keeps only boundary points as training samples and discards other data points surrounded by these extreme points. Geometrically, the normal vector to the tangent plane towards a maximum number of neighbours generally makes an angle of less than 90 degrees with the neighbour points. For a potential extreme point x_i with k neighbours, relative vector v_{ij} can be found as:

$$v_{ij} = x_i - x_j \qquad where \qquad j = 1, 2, 3\ldots k. \tag{9}$$

Each v_{ij} is normalized to v_{ij}^u for the estimation of normal vector to the tangent plane. The normal vector n_i for any boundary point x_i is calculated as:

$$n_i = \sum_{j=1}^{k} v_{ij}^u \tag{10}$$

In Fig. 10, the angle θ_{ij} between normal vector n_i of tangent plane and data point vector v_{ij} takes values between 0 to $\pi/2$ on the positive side (dense distribution of the data) of the tangent plane. A sample is treated as an extreme boundary point if the value of the angle θ_{ij} is zero and the tangent plane on the convex or concave surface, as shown in Fig. 10. The number of neighbours with $\theta_{ij} \geq 0$ out of k-nearest neighbours has been used to determine the extreme point as follows:

$$l_i = 1/k \sum_{j=1}^{k} \theta_{ij} \geq 0 \tag{11}$$

If a sample is on the concave surface, most of its neighbours fall on one side of the tangent plane and very few on the other side. Therefore, the value of $l_i \geq 1-\alpha$ elects it as a boundary point, where α is non-negative but a small parameter and empirically chosen from a range of [0, 0.2]. A data point x_i on a convex surface is chosen as an extreme point if all its k-nearest neighbours are on one side of the tangent plane and value of $l_i = 1$. All samples selected by this process are treated as reduced training set to train OCSVC. This approach operates on local data distribution

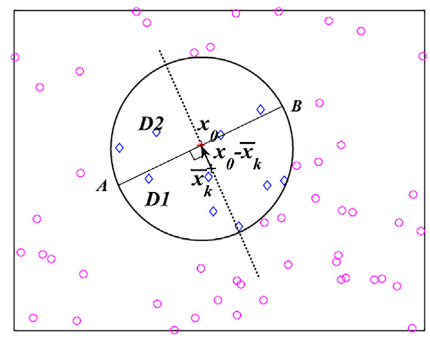


Fig. 11. The neighbourhood sphere [81].

and geometrical analysis; hence, preserves class information. It is evident that though this training sample reduction technique outputs fewer training samples and requires less memory, it suffers from computational overhead due to the nearest neighbours approximation. It is also observed that if the sample space is too large in high dimensional feature space, this approach may select samples with less information; thus unsuited for big data applications.

To reduce the training samples, Sun et al. [80] proposed a heuristic sample reduction (HSR) technique. In this approach, initially, the cluster centres of the training set are identified by k-means. Afterwards, all samples are sorted in descending order according to the distance from corresponding cluster centre. Later, the farthest sample is moved from the sorted set into the reduced training set. The training samples which have a shorter distance to the farthest sample are deleted from the training set. This process iterates until the sorted training set is empty. In this way, HSR reduces the time complexity of support vector data descriptor (SVDD) [27] from $\mathcal{O}(N^3)$ to $\mathcal{O}(\tilde{N}^3)$, where N is the number of original samples and $\tilde{N}$ is the number of the reduced training set.

Zhu [81] observed the problem associated with the method of Li [79] and offered a more effective and robust method for training sample reduction using the uneven neighbour's distribution property, suitable to high dimensional datasets. Let, for a sample x_0 in dataset D, x_i (i = 1, 2, $\cdots$, k) is one of the k-nearest neighbours and $\bar{x}_k$ is defined as mean of neighbours of x_0. The neighbourhood sphere is divided into two parts D_1 and D_2 (Fig. 11), where the divisor is the hyperplane AB passing through x_0 and perpendicular to $(x_0 - \bar{x}_k)$. D_1 is the hemisphere, where $\bar{x}_k$ is located and the other is defined as D_2. If there are more neighbours in D_1 than D_2, then the sample generally locates near the boundary, whereas if number of neighbours in both hemispheres are nearly equal, it generally locates within the data distribution. The sample-neighbour angle is defined as:

$$c_{0,i} = cos(\theta) = \frac{\langle x_0 - \bar{x}_k, x_0 - x_i\rangle}{\|x_0 - \bar{x}_k\|, \|x_0 - x_i\|} \tag{12}$$

where θ is the angle between $x_0 - \bar{x}_k$ and $x_0 - x_i$ and cosine sum is defined as:

$$c_0^{sum} = \sum_{j=1}^{k} c_{0,i} = \frac{\langle x_0 - \bar{x}_k, x_0 - x_i\rangle}{\|x_0 - \bar{x}_k\|, \|x_0 - x_i\|} \tag{13}$$

When the count of neighbours in D_1 is much greater than D_2 then c_0^{sum} is close to k, whereas if neighbours are nearly equal then c_0^{sum} is close to 0. Therefore, the proposed method selects training data samples having c_0^{sum} greater than a user-defined threshold value. This threshold value can be fixed by an

Table 4
Summary of sample reduction techniques for OCSVCs.

Citation	Methodology/Specification	Achievements
Li [79]	Extreme points selection algorithm	This approach is suitable for low dimensional datasets.
Sun et al. [80]	Heuristic sample reduction (HSR).	HSR reduces the time complexity of SVDD from $\mathcal{O}(N^3)$ to $\mathcal{O}(\tilde{N}^3)$, where $\tilde{N}$ is the size of reduced training set.
Zhu et al. [81]	Neighbour's distribution properties.	Reduces training samples by 85% and suitable for the high dimensional dataset.
Alam et al. [82]	The mean vector of samples and angular directional distances of neighbours of a sample play the key role.	The time complexity of the method is less compared to pre-existing methods and suitable for high dimensional datasets.

iterative way of calculating the error. To speed up the k-nearest neighbour problem, the parallel computation has been imposed for subsets of the dataset. This approach reduces the computation of k-nearest neighbour from $\mathcal{O}(n^2)$ to $\mathcal{O}(mn)$, where n is the size of the whole dataset and m is the number of samples per subset. The above method generates approximately 15% of the total dataset as the training set and ensures the preservense of class information, which has been verified on different UCI benchmark datasets. Meanwhile, better neighbour discovery methods and autonomous threshold estimation of k may give improved results.

Recently, Alam et al. [82] proposed a novel sample reduction algorithm, where the mean vector and angular directional distances of samples play the key role. This approach eliminates unnecessary training samples, reduces training time of SVDD, and preserves the most promising boundary points. The time complexity of the method is less compared to pre-existing methods. It is observed from experiments that the proposed method outperforms over pre-existing algorithms [79,81] and suitable for high dimensional datasets.

This section gives detailed description of existing sample reduction methodologies for OCSVCs. Firstly, Li [79] proposed a training sample reduction algorithm in which the extreme points are selected using the tightness criteria of the decision boundary, and samples nearer to the decision boundary are well-kept for the training. This technique reduces the training time but not suitable for high dimensional datasets. Later, Sun et al. [80] proposed heuristic sample reduction (HSR) method for sample reduction based on k-means clustering, whereas based on neighbour distribution property, Zhu et al. [81] proposed another training sample reduction technique that reduces the training samples to approximately 15% of the original dataset, and suitable for high dimensional datasets too. Recently, Alam et al. [82] used mean vector and angular directional distances of neighbours of samples to extract the most promising boundary points as the reduced training set. This approach rejects unnecessary training samples and reduces the training time of SVDD. A comparative study of preceding literature is summarized in Table 4.

6. Distributed and online OCSVCs

Although, several OCSVC algorithms have been proposed for anomaly or novelty detection for batched data but not well explored for the distributed or online environment. Real-world applications like earth science, weather forecasting, satellite and aviation control, social networking, etc., continuously generate samples/features over the time-period. Information extraction from such complex streaming data is always a challenging task because of geographically distributed and heterogeneous data sources. Concerning distributed streaming data, centralized processing may lead to communication and computation overhead. Therefore, incremental classification techniques are needed to process these distributed data for anomaly or outlier detection. Several online SVM learning algorithms [83–90] have been published in recent years. The milestone incremental learning approach is the incremental SVM [91], which provides a framework for distributed or online one-class learning.

Tax et al. [91] proposed an incremental SVDD approach by optimization of following abstract notion of incremental SVM:

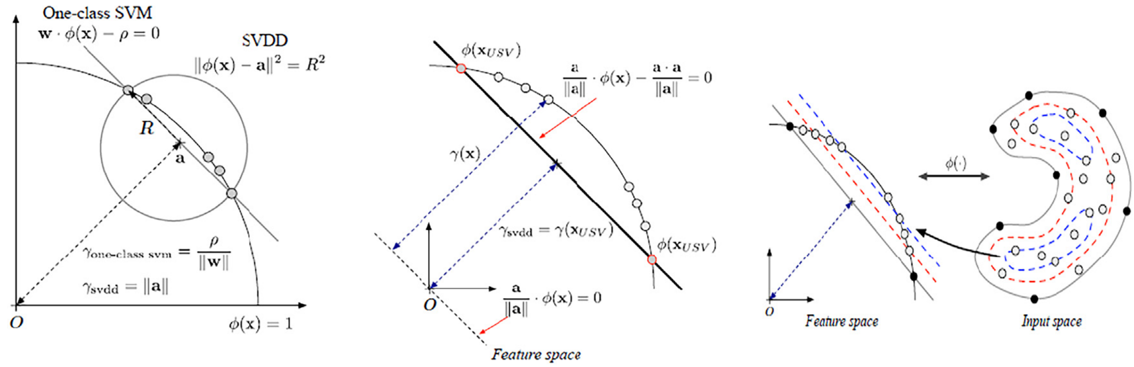
$$\max_{\mu} \max_{\substack{0\le x\le C\\ a^T x+b=0}} : w = -c^T x + \frac{1}{2}x^T Kx + \mu(a^T x + b) \tag{14}$$

where c and a are $n \times 1$ vectors, K is a $n \times n$ matrix, and b is a scalar. Abstract parameters a, b and c can be defined accordingly for different SVM algorithms. For the standard support vector classifiers [92], $c = 1$, $a =$ y, $b = 0$ and C = given regularization constant, whereas for υ-SVC [93], $c = 1$, $a =$ y, $b = 0$ and C = $1/(\upsilon$N) have been chosen. For the SVDD [30], the parameters are defined as $c = \mathrm{diag}(K)$, $a =$ y and $b = -1$. When a new point k is added, its weight x_k is initialized to 0 and the weights of other data points and μ should be updated in order to obtain the optimal solution for new grown dataset. Likewise, when a point k is to be removed, its weight x_k is enforced to 0, while updating the weights of the remaining points and μ so that the solution obtained with $x_k = 0$ is optimal for the reduced dataset.

To deal with stream data, Cauwenberghs et al. [94] proposed an incremental/decremental online SVM approach named as C & P algorithm. Training of SVM on arrival of a new sample with historical support vectors might give approximate results [85]. This method satisfies the Kuhn–Tucker (KKT) conditions to all historical data and adiabatically adds a new data point to the solution for incremental learning, whereas leave-one-out (LOO [92]) is used in decremental learning as it predicts the generalization power of a trained classifier. The adiabatic process ensures that the value of margin vector coefficients change during each incremental step to keep all elements in equilibrium, i.e., keep their KKT conditions satisfied. This approach was not for OCSVCs, but further, the same concept inspired Laskov et al. [95] to give a new solution for incremental SVDD which was the extension of the algorithm proposed by Tax et al. [91].

Later, Sillito et al. [96] proposed an incremental one-class learning algorithm for outlier detection. In this approach, outliers were identified with the help of a probability distribution function, where initially, the training data distribution is obtained with the help of kernel density estimation. When the number of training samples reaches to the maximum feasible limit for kernel density estimation, the kernel density estimate is treated as a maximally complex Gaussian mixture model and keep the model complexity constant by merging a pair of components for each new kernel added. This method outperforms over conventional incremental OCSVM [91].

Tavakkoli et al. [97] proposed an incremental SVDD training mechanism that runs in constant time irrespective of training sample space size since its retraining is done only with support vectors. This incremental training algorithm is based on the theorem proposed by Osuna et al. [98], which concludes that a large QP problem can be broken into a series of smaller sub-problems. The optimization of these sub-problems converges when new samples are added as long as at least one sample violates the KKT conditions. A new sample will be discarded if it satisfies the KKT condition because its inclusion does not minimize the objective function. Otherwise, this training sample should be included in the new working set. Since the algorithm runs only on support vectors, it ensures an optimized result.



(a) The margins of hyperplane γ_{OCSVM} and γ_{SVDD} respectively.

(b) sample margin $\gamma(x)$ for data x in SVDD.

(c) Data distribution in feature space and data description boundary in input space.

Fig. 12. Geometric representation of SVDD and OCSVM [99].

For large scale datasets, Kim et al. [99] proposed fast incremental OCSVM based on sample margin, that reduces computational overhead (time and space) with better classification accuracy. In this approach, non-support vectors are rejected with the concept of sample margin. Geometrically, Fig. 12a shows that the norm of the centre of SVDD [31] is equal to the margin of hyperplane of one-class SVM [33], and in unit norm feature space SVDD can be reformulated by a hyperplane equation as follows:

$$\| \phi(x) - a \|^2 \leq R^2 \Leftrightarrow w_{SVDD}.\phi(x) - \rho_{SVDD} \geq 0 \tag{15}$$

where a is the hypersphere centre of SVDD. The normal vector w_{svdd} and the bias ρ_{svdd} of SVDD hyperplane are as below:

$$w_{SVDD} = \frac{a}{\| a \|}, \qquad \rho_{SVDD} = \| a \| \tag{16}$$

Sample margin is defined by the distance from the image of data x to the virtual hyperplane which passes through the origin of feature space and parallel to the optimal hyperplane as shown in Fig. 12b. In SVDD, sample margin is defined as below:

$$\gamma_{SVDD}(x) = \frac{a.\phi(x)}{\| a \|} \tag{17}$$

where a is the centre of SVDD's hypersphere and y(x) is the image of data x in feature space. In one-class SVM, sample margin is defined as follows:

$$\gamma_{OCSVM}(x) = \frac{w.\phi(x)}{\| w \|} \tag{18}$$

Because data examples exist on the surface of a unit hypersphere, the sample margin has 0 as the minimum value and 1 as the maximum, i.e.

$$0 \leq \gamma(x) \leq 1 \tag{19}$$

Also, sample margin of unbounded support vectors x_{USV} $(0 < \alpha_{x_{USV}} < \frac{1}{\upsilon N})$ are the same as the margin of hyperplane. Hence :

$$\gamma(x_{USV}) = \gamma_{SVDD} = \| a \| \tag{20}$$

$$\gamma(x_{USV}) = \gamma_{OCSVM} = \frac{\rho}{\| w \|} \tag{21}$$

The sample margin represents the distribution of samples in the feature space. Fig. 12c shows the distribution of the sample margin of training data and the hyperplane of OCSVM in feature space. The Kuhn–Tucker conditions have to be preserved for all trained data before and after a new data x_c is added for training. That is, the change of Lagrange multipliers ($\Delta\alpha$) is determined to hold the Kuhn–Tucker conditions. In general, the Lagrange multipliers of error support vectors (BSVs) and in-class support vectors (NSVs) do not vary during each update step. Only the Lagrange multipliers of marginal support vectors (USVs) and newly added data x_c change their values to satisfy the Kuhn–Tucker conditions during the update process.

Afterwards, Bounsiar et al. [100] proposed a distributed OCSVM to deal with distributed data. In real-world classification problems, a class may have multiple subclasses which form different clusters in representation space, with different densities and extent. OCSVM or SVDD uses the same kernel width on all support vectors, regardless if they are in a dense region of the data distribution or sparse one. To overcome this weakness, multiple cluster OCSVM (MCOS) has been proposed, where a clustering algorithm is used to divide the training set into coherent clusters, and an OCSVM was built with separate kernels on each cluster. By combining all local representations, the final representation of data is obtained. One of the advantages of using MCOS is that it never goes worse than OCSVM and gives better performance. There is a scope for improvement with the clustering algorithm if the optimum number of clusters can be drawn guided by training samples for a given classification problem.

Das et al. [101] analysed the problem of one-class classification over the distributed environment at NASA Ames Research Center and proposed an algorithm for anomaly detection using OCSVM for distributed vertically partitioned data. The goal was to minimize the communication cost among distributed data sources and data transfer ratio to identify anomalies or outliers. The results obtained in this research work were appreciable, where the amount of communication required is less than 1% of that is required for central processing. The proposed algorithm consists of two steps. The first step uses a local anomaly detection algorithm, where OCSVM is executed at each node, which identifies the outliers based on local features present on that node. Further, these local outliers are collected to one central node (master node) and tested against the global model to finalize the global outliers. The working of the first step of the proposed algorithm is shown in Fig. 13, where $\{P_1, P_2, \ldots, P_p\}$ are the sites at different geographical regions; $\{x_1, x_2, \ldots, x_m\}$ are the local subset of features from the global feature set and $\{O_1, O_2, O_3, \ldots, O_p\}$ are the local identifiers from each site. P_0 is the master site, and O_d is the final global outlier that has been identified by the algorithm.

The objectives of the second step of the proposed algorithm are to minimize the communication cost and attain high accuracy. To fulfil these objectives; the pruning rule for communication

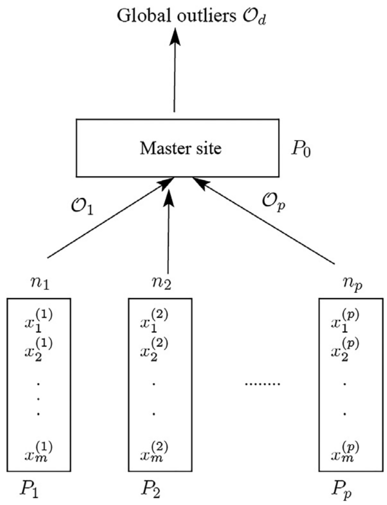


Fig. 13. Distributed architecture [101].

between the sites has been proposed. According to the pruning rule, if a point is not an outlier for any local set, then it cannot be a global outlier. This statement says that a sample data point, $\vec{x} \epsilon$ D (where D is overall data) may be a global outlier(O_d), if it is an outlier with respect to at least one feature or a subset of features i.e. $\vec{x}_i^{(j)}\ \epsilon\ O_j$, where $j = \{1, 2, \ldots, p\}$. The assumption that most of the global outliers will be a local outliers for at least one of the distributed sites, which may not be true in general, but it has been verified theoretically that with an increased number of features over distributed sites, the percentage of correct detection increases exponentially. The proposed model shows that the false-positive rate of the algorithm is 0 and the resulting hyperplane is same as in the case, where the entire dataset at one location. The message complexity reduction can be verified using Eqs. (22) and (23). When data samples are centralized, the message complexity is:

$$m \times n_1 + m \times n_2 + m \times n_3 + \cdots + m \times n_p = m \times \sum_{i=1}^{p} n_i \tag{22}$$

and for distributed algorithm the local outlier are firstly centralized from all distributed sites to a sample of size T_s for building a global model, the message complexity in this method is:

$$\begin{aligned} |O_1| \times \sum_{i=1}^{p} n_i + |O_2| \times \sum_{i=1}^{p} n_i + \cdots + |O_p| \times \sum_{i=1}^{p} n_i + T_s \\ \times n_1 + T_s \times n_2 + T_s \times n_3 \\ + \cdots + T_s \times n_p = \sum_{i=1}^{p} O_i \times \sum_{i=1}^{p} n_i + T_s \times \sum_{i=1}^{p} n_i \\ = \sum_{i=1}^{p} n_i \times \left(T_s + \sum_{i=1}^{p} O_i \right) \end{aligned} \tag{23}$$

The proposed distributed algorithm is more efficient than its centralized counterpart. The experiments were conducted on NASA MODIS satellite image and the CMAPSS datasets. The results of experiments show that the classification accuracy of this algorithm is approximately 99%.

Afterwards, Castillo et al. [102] proposed a novel distributed OCSVM for big data environment. The objective is to find a global model using several OCSVMs. These parallel OCSVMs work on local data partition and able to exchange data during learning to improve their individual specialization. The mathematical formulation in the proposed model makes the optimization problem separable and avoids some data points considered as outliers in the final solution. This method is unaffected by the position of the outliers and able to fit the data more precisely. Extensive experiments have been performed on several datasets and comparison of the proposed model is done with other one-class classifiers. Experimental results show that the proposed model outperforms over others.

Later, Hansi et al. [103] proposed a fast incremental SVDD (FISVDD) algorithm for online data. This method works on the observation that all potential support vectors have the same distance to the centre of the hypersphere. Each iteration involves only the existing support vectors and the new data point. However, this approach depends only on matrix manipulations and the complexity in each iteration is only $\mathcal{O}(k^2)$, where k is the number of support vectors. Experimental results indicate that FISVDD demonstrates significant gain in efficiency with almost no loss in either outlier detection accuracy or objective function value. For anomaly detection in wireless sensor networks, two novel distributed one-class support vector machines were proposed by Miao et al. [104]. In this research, to achieve decentralized implementation without transmitting the original data, the kernel function is replaced with a random approximate function [105]. Furthermore, to find an appropriate dimension, a sparse constraint [106] is added into the decentralized cost function to get another one. These two cost functions are minimized by stochastic gradient descent to derive two distributed algorithms.

Recently, Kefi et al. [107] proposed an incremental one-class covariance-guided SVM (iCOSVM). The iCOSVM improves over other incremental OCSVM by the integration of the covariance matrix into the optimization problem. If new samples satisfy KKT conditions, then their Lagrange multipliers are zero and they will not change the previous support vector set. However, if samples in the newly added training set violate the KKT conditions, some of them will become new support vectors surely, and a number of non-support vectors in the previous training set will likely be converted to new support vectors. Extensive experiments have been performed over artificial and real-time datasets to validate the performance of the proposed model.

From the preceding literature, it is observed that limited research work has been done for distributed OCSVCs. Tax et al. [91] proposed an incremental SVDD with the replacement of old sample with the newly arrived sample that outperforms over conventional batch SVDD model. Later, Tavakkoli et al. [97] proposed an incremental SVDD training mechanism that runs in constant time irrespective of training sample space size, since retraining is done only with support vectors. For large scale datasets, Kim et al. [99] proposed fast incremental OCSVM based on sample margin, that reduces computational overhead (time and space) with better classification accuracy. In continuation of earlier research works, Bounsiar et al. [100] proposed a multi-cluster one-class SVM (MCOS) algorithm. Das et al. [101] proposed the distributed OCSVM with 99% accuracy on geographically distributed data with less than 1% communication overhead. Afterwards, Castillo et al. [102] proposed a novel distributed OCSVM for big data environment, which is unaffected by the position of the outliers and able to fit the data more precisely. Later, Hansi et al. [103] proposed a fast incremental SVDD (FISVDD) algorithm for online data. This method works on the observation that all potential support vectors have the same distance to the centre of the hypersphere. For anomaly detection in wireless sensor network, two novel distributed one-class support vector machines were proposed by Miao et al. [104]. Later, Kefi et al. [107] proposed an incremental one-class covariance-guided SVM (iCOSVM). A comparative study of the preceding literature is shown in Table 5.

Table 5
Summary of distributed OCSVCs.

Citation	Methodology/Specification	Achievements
Tax et al. [91]	A new sample is added while some existing sample is removed.	Performance is better than batch model of SVDD.
Sillito et al. [96]	Probability distribution function.	Outperforms over the method of Tax et al. [91] for outlier detection.
Tavakkoli et al. [97]	If a sample does not satisfy KKT condition, then it must be considered in the working set.	Runs in constant time and takes computational resources.
Kim et al. [99]	Operates on the principle of sample margin.	Reduces computational overhead (time and space) with better classification accuracy.
Bounsiar et al. [100]	Multi-cluster OCSVM, k-means clustering algorithm.	The complexity is never worse than OCSVM.
Das et al. [101]	Distributed outlier detection algorithm.	Reduces the computing complexity.
Castillo et al. [102]	Objective is to find a global model using several OCSVMs	Outperforms over other one-class classifiers. Unaffected by the position of the outliers and able to fit the data more precisely.
Hansi et al. [103]	All potential support vectors have the same distance to the centre of the hypersphere.	Time complexity in each iteration is only $\mathcal{O}(k^2)$, where k is the number of support vectors.
Miao et al. [104]	Random approximate function [105] is used and sparse constraint [106] is added to approximate the appropriate dimension.	Outperformed over other incremental OCSVCs.
Kefi et al. [107]	Integration of the covariance matrix into the optimization problem	Outperformed over conventional incremental OCSVCs.

7. Applications of OCSVCs

In this section, applications of OCSVCs are discussed. It is observed that the OCSVCs are most importantly used for anomaly and novelty detection in every application domain. An anomaly is referred to as an outlier in data mining and pattern recognition tasks. For one-class classification or anomaly detection, apart from OCSVCs other machine learning models [6–11] also have been proposed like one-class nearest neighbour, one-class random forest, one-class deep neural network and autoencoder, one-class support higher-order tensor machine, one-class ensemble model, etc. One-class nearest neighbour (OCNN) is the improved form of the classical nearest neighbour model, and works fine for lower dimensional datasets, but becomes unstable for large dimensional datasets (beyond 15–16 features). In this model, the distance of an input x to the nearest neighbour NN(x) and the distance of the nearest neighbour to its nearest neighbour (NN(NN(x))) are calculated [108]. If the first distance is larger than the second, then the input sample is considered as an outlier. The main disadvantage of the nearest neighbour approach is its high computational cost for large-size datasets as the whole training set has to be stored and evaluated.

One-class random forest (OCRF) [109] is a decent choice for anomaly/novelty detection because it is flexible and does not suffer from the problem of overfitting. But the main disadvantage is the prerequisite of artificial outlier generation. It also requires more computational resources and is less intuitive. At present, even though the broad developments made by deep learning approaches in the different application domains [110–112], there is a relative insufficiency of deep learning methods for anomaly detection. Due to the diverse applicability of OCSVCs, several research articles have been published in different application domains continuously since the birth of OCSVCs. Based on an extensive literature survey, OCSVCs are found as the most popular approaches for anomaly and novelty detection. It carries potential research directions to enhance the suitability and applicability of anomaly and novelty detection tasks like document classification, intrusion detection, novelty detection, machine fault detection, fraud detection, etc. This section gives a comprehensive overview of different application domains, where OCSVCs have been applied successfully. The literature survey has been done on the following application domains:

- Text/Document applications
- Audio applications
- Video applications
- Industrial applications
- Hyperspectral and Remote sensing data application

7.1. Text/document applications

Traditional text and document classifiers need at-least two well-defined classes, and this idea does not work for one-class document classification. For document analysis, many research articles have been reported concerning different classifiers during the last two decades [7], but very less number of researches has been done with respect to one-class classification. Treating the document classification as the one-class problem, Moshe et al. [113] proposed a robust solution to the authorship verification problem, where the classifier was trained with only one writer's sentences and performed verification using one-class support vector machine with RBF kernel. The experimental results were better than the conventional multi-class classification approach.

In the field of document classification, the handwritten signature verification is very important classification problem for both online and offline acquisition modes. An offline signature verification system can be modelled as writer-dependent (WD) and writer-independent (WI) [114,115]. Ferrer et al. [116] proposed WD handwritten signature verification systems using several classifiers, whereas Guerbai et al. [117] proposed the application of one-class SVM and showed its effectiveness over binary SVM for handwritten signature verification. The centre of excellence for document analysis and recognition (CEDAR) signature dataset [118] was used for comparative study and it is observed that OCSVM outperforms over binary-SVM.

Later, Guerbai et al. [119] proposed writer independent signature verification using curvelet transform features [120] along with OCSVM. In this approach kernel of OCSVM was modified carefully with different distance metrics (Euclidian, Cityblock, Chebeychev, Correlation, Spearman, etc.). Most recently, Bouamra et al. [121] proposed an offline automatic signature verification system using black and white run-length distributions as features and proved that using run-length features, OCSVM gives high accuracies in different scenarios.

7.2. Audio applications

Speech recognition is always a challenging task and solved by many machine learning approaches like artificial neural networks, deep networks, support vector machines, OCSVCs, etc. Though, hidden Markov models (HMMs) are very popular for Automatic Speech Recognition (ASR) [122], effective properties of SVM (maximum marginal separation of two classes and convergence to the minimum cost) make it an important research area for signal processing. Meanwhile, it is observed that concerning OCSVCs, very limited works have been done so far in the field of speech recognition. Considering speech recognition as one-class classification problem, Rabaoui et al. [123] proposed a modified OCSVM framework [124] for sound detection and classification. This approach detects events in a continuous audio stream. In this research, an unsupervised sound detection method was also proposed, which does not require any pre-trained models. It uses exponential family model and OCSVMs to approximate the generalized likelihood ratio. Later, Belkacem et al. [125] proposed a two-step (speaker turn detection and speaker clustering) solution to speaker diarization. Experimental results show the effectiveness of OCSVM over conventional approaches. In the same field of research, Salsabil and Zied [126] proposed a stressed speech recognition system based on cepstral features [127,128] and one-class SVM. Mel frequency cepstral coefficients (MFCC) and Gammatone frequency cepstral coefficients (GFCC) have been used as cepstral features. Experiments were performed on speech under simulated and actual stress (SUSAS) database, and results show that the MFCC and GFCC features give the best performance for OCSVM with Gaussian kernel. It is also evident that OCSVM performed better over the conventional multi-class SVM.

7.3. Video applications

Keren et al. [129] used OCSVM and proposed one-class support vector tracker (OCST) with three major benefits over pre-existing tracking systems: more accurate positioning estimation, relieve the drifting problem and lower computational cost. The OCST takes the advantages of the dense HOG feature [130] and 2bitBP feature [131]. Experimental results ensure the effectiveness of the proposed approach. Afterwards, Vasileios et al. [132] proposed a method for video summarization based on human activity description using a new variant of OCSVM, known as subclass-OCSVM. This model exploits subclass information in the OCSVM optimization problem to minimize the data dispersion within each subclass and determine the optimal decision function. In this research, the performance of the proposed OCSVM model is tested on three Hollywood movies belonging to three different categories: action, adventure and drama, whereas training is done with eighteen Hollywood movie trailers from categories: action, comedy, thriller and drama. Experimental results show that modified OCSVM outperforms over traditional OCSVM.

An anomaly in crowd scenes was successfully detected using online adaptive OCSVM, proposed by Hanhe et al. [133]. Online OCSVM, along with sliding buffer updates the system incrementally with lesser computation cost and discards obsolete patterns. The proposed model in this research is capable of detecting local and global anomalies by keeping the Karush–Kuhn–Tucker (KKT) conditions satisfied with the growing dataset.

7.4. Industrial applications

For the past two decades, one-class support vector classifiers are the most important area of research for fault detection. In OCSVCs, training data may contain some outliers that will degrade the performance of the classifier. To overcome this problem, Prayoonpitak et al. [134] proposed a robust OCSVM based on Gaussian-based penalization to eliminate outliers from training set, and successfully tested on MFPT bearing dataset for fault detection. Experimental results show that the modified OCSVM outperforms over the conventional OCSVM. In this model, to control the rate of penalization weighted Gaussian kernel is used. Later, Yingchao et al. [135] proposed a more robust OCSVM to the above-discussed problem, which eliminated the suspected outliers, and implemented the same for fault detection on Tennessee Eastman process (TEP) benchmark dataset. TEP is a dataset of a real chemical engineering process proposed by Downs and Vogel [136] with 21 fault types. For performance comparison, other variations of OCSVCs were used, and results show that the proposed model outperforms over others. This model works on the concept of automatic adjustment of the tuning parameter υ iteratively using "Bisection" algorithm. First, the classifier identifies some suspected outliers by automatically selects a proper parameter υ then these outliers are excluded from the training set and OCSVM is retrained to obtain the decision boundary enclosing the target class. This boundary is used to identify outliers more accurately.

The OCSVM has been successfully implemented for fault detection in a chiller system by Beghi et al. [137]. Malfunctioning of any chiller system may lead to the user's discomfort, increased production and maintenance cost. Hence, prior investigation of such a system is a challenging task in industrial applications. To approximate the fraction of training errors and support vectors, the parameter υ is fixed to a small value i.e., 0.05 to ensure a small misclassification rate. Whereas, the Gaussian kernel width parameter is chosen via tightness detection. A heuristic approach has been proposed to evaluate the tightness of the decision boundaries. If there exists at least one large hole inside the boundaries, i.e., a region without samples, the tightness is considered as loose, and if the boundaries nearby two neighbouring samples are concave, it is considered as tight. Experiments were performed on the chiller dataset [138], and results prove that this model gives better fault detection. To separate suspected outliers from training samples for better performance of OCSVM, Shen et al. [139] proposed another model. This model modifies the penalty factor estimation which is controlled using the distance between a data point and centre of the dataset.

7.5. Hyperspectral and remote sensing data applications

Geostationary satellites capture the hyperspectral geographic images for different observations like flooding, droughts, forest fires, etc., whereas wireless sensor networks are widely used to collect the data concerning climate, rain and soil analysis. Classification is the most challenging task for these type of high dimensional datasets because mostly the investigation is target-specific. Munoz et al. [140] proposed RBF kernel-based SVDD approach for remote sensing images. In this research, the performance of SVDD has been compared with the Gaussian domain description classifier (a mixture of Gaussian domain description classifier and *k*-NN) and found that SVDD outperformed over others.

Zhang et al. [141] demonstrated hyperspectral classification using spectral angle mapper (SAM) and extended OCSVM, and proved that joint use of both spectral magnitude and gradient outperforms over the spectral data alone. SAM is a supervised classification algorithm, which utilizes spectral angular information for the classification of hyperspectral image data [142]. Experimental results show that proposed OCSVM outperformed to SAM. Afterwards, Xu et al. [143] proposed an approach to estimate the rice planting area using OCSVM that optimizes rice production prediction. In this research, the rice planting area in Jiangsu, China with landsat optical land imager (OLI) imagery has been used. Results show that the performance of OCSVM is extraordinary. In this research, for hyperspectral image classification the discriminant band selection mechanism has been applied to reduce the computation and storage cost.

Seeking for the same objective, Tang et al. [144] proposed a supervised band selection method that computes a discriminative weight for every band. Discriminative bands contribute more positive scores to the target class during the training stage, and this phenomenon enables the model to learn discriminative band weight vector for each class, then bands with higher discriminative weights are selected. In this research, the one-class SVM model has been chosen to learn discriminative weights for hyperspectral bands. The efficiency of the proposed model has been justified with the experiments on two widely used datasets: Pavia Centre, and Indian Pines. Later, Roodposhti et al. [145] proposed an approach to estimate draught sensitivity map for vegetation using one-class SVM with the use of a combination of both 30 years statistical data (1978 to 2008) of synoptic stations and 10 years MODIS imagery archive (2001 to 2010) of Kermanshah Province, Iran. The standardized precipitation index (SPI) and the enhanced vegetation index (EVI) were extracted from these datasets and have been used as indicators of soil moisture.

7.6. Intrusion detection based applications

Intrusion detection is one of the most important and critical issue in this era of distributed and big data environment. Wang et al. [146] applied OCSVM for intrusion detection and found that this classifier outperforms over another state of art classifiers. Looking for the same objective of intrusion detection in the electric power generation system, Maglaras et al. [147] proposed an integrated OCSVM associated with supervisory control and data acquisition (SCADA) network. Electric power generation, transmission and distribution tasks are managed by SCADA systems. SCADA systems are very complex and geographically distributed; hence, an effective intrusion detection system is needed to secure the system against cyber-attacks. This defence mechanism must be distributed, cheap and above all accurate, since false positive alarms or mistakes regarding the origin of the intrusion mean severe costs for the system. The proposed model provides details about the origin and the time of an intrusion. This model uses network traces to create a cluster of split OCSVMs. Outcomes of these split OCSVMs are combined with those of the central OCSVM. These split OCSVMs are weighted using social metrics, aggregated and finally classified using *k*-means clustering. Experimental results show that the proposed method can handle all attacks efficiently.

Khreich et al. [17] proposed runtime system anomaly detection at the host level using OCSVM. To reduce the false alarm rate, this technique combines the frequency with the temporal information from system call traces and OCSVM. In this model, the feature extraction approach starts by segmenting the system call traces into multiple n-grams of variable length and further, mapped to fixed-size sparse feature vectors, which are then used to train OCSVM detectors. The results achieved on a real-world system call dataset show that in the proposed model, feature vectors with up to 6-grams gives better performance.

8. Concluding observation

From the different sections of the above literature, it has been observed that limited availability of outliers or ill-defined non-target class data is always challenging and interesting part of OCC problems. In this review paper, all important issues along with their solutions proposed by many researchers have been discussed. The review has been grouped into six important areas: OCSVC algorithms, kernel parameter estimation techniques, feature selection and reduction, training sample reduction, distributed OCSVCs and application domains.

For OCC, Tax et al. [27] proposed support vector data description (SVDD) and Schlökopf et al. [33] proposed an alternative approach to SVDD with hyperplane, wherein both the methods origin plays an important role. Both methods give comparatively equivalent results when the Gaussian kernel is used. Whereas, Manevitz et al. [12] treated outliers as representative of the second class and assumed that along with origin, all nearby lower-dimensional data points to the origin must be considered as outliers. Later, Li et al. [35] presented an improved version of the approach of Schlökopf et al. [33] for intrusion detection with better accuracy. This method considered all data points closer to the origin as outliers along with the origin. Luo et al. [36] extended the work of Tax et al. [27,30] and proposed cost-sensitive OCSVM algorithms named frequency-based SVDD (F-SVDD) and write-related SVDD (WS-SVDD) for intrusion detection problem. In the race of enhancement of OCSVM, Hao et al. [37] extended the work of Schlökopf et al. [33] with the fuzzy integration. In this approach, for each training data, fuzzy membership has been applied to reformulate OCSVM. The result obtained by using fuzzy-OCSVM was not very satisfactory and generalized. Later, Yang et al. [39] proposed a neighbourhood-based OCSVM method for fMRI data to identify schizophrenia. This approach integrates the neighbourhood consistency hypothesis [40] along with OCSVM to compute primal values that denote the distance between data points and hyperplane in kernel space. Based on primal values and their neighbours, a new decision value is computed for each voxel of a fMRI image. Whereas, Lui [41] extended the work of Schlökopf et al. [33] and Hao et al. [37] and proposed robust-fuzzy-OCSVM. In this paper, by introducing robustness and fuzziness parameters, the reformulation of one-class SVM has been done. Zhu et al. [42] observed that all the training data points were getting equal attention for SVDD and proposed a weighted strategy based on neighbour's distribution in which points nearer to decision boundary should be given more weight than the inner points. Ji et al. [43] proposed a new weighting strategy in which along with distance factor, density factor has been used to get a better weight and proposed a better WOCSVM known as Adaptive-WOCSVM (A-WOCSVM). It has been observed that the perturbation factor is inversely proportional to the fuzziness factor, and for a given perturbation, there is always a lower bound on the fuzziness factor. Later, Wang et al. [44] proposed ellipsoidal data description (ELPDD) model for one-class classification and proved its effectiveness over conventional SVDD model. It is observed that for anomaly detection, OCSVCs are the best suitable one-class classifiers but very sensitive to the outliers or noise. To overcome this problem, Tian et al. [45] introduced ramp loss function [52] to the conventional OCSVM and proposed Ramp-OCSVM. This model is more robust compared to the conventional OCSVM due to the non-convexity properties of the ramp loss function. After that, two enhanced OCSVMs were proposed based on rescaled hinge loss function and Adaboost by Xing et al. [46,47]. Though several researchers have proposed different OCSVC algorithms; still, there is enough scope for a more flexible classification model to minimize the false-positive and false-negative rate.

From the literature covered, it is observed that to make a smooth data description, kernels are used, and the Gaussian kernel shows more promising results for one-class classification. Parameter estimation is a very important issue in kernel-based OCSVCs. Under Section 2, different kernel estimation techniques [56,58–62] were analysed and found that some approaches [58–60] are supervised in nature, whereas approaches [56,61,62] are unsupervised to overcome computational complexity up-to some extent. It is also identified that data-driven parameter estimation and selection of suitable kernel functions are potentially strong research areas.

Concerning to feature selection and dimensionality reduction for OCSVMs, it is observed that general methods of other multi-class classifiers cannot be used because of the data irregularity problems. Tax et al. [65], Lian [66], Liu et al. [67] and Feng et al. [68] demonstrated different variations of feature selection mechanisms as dimensionality reduction approaches with principal component analysis (PCA) [69] for one-class SVCs. Jeong et al. [73] proposed two new methods for feature selection based on specific criteria function, whereas Nagabhshan et al. [74] proposed target class supervised feature sub-setting to obtain an optimal subset of features. Later, Yousef et al. [75] experimented microRNAs of plant and found that if proper feature selection methodologies are used, then up to 30% of classification accuracy can be increased, which is a noteworthy result. The literature shows that there is no best-known feature selection method for OCSVCs. However, there are enough possibilities for improvements over existing methods.

In this era of big data, it is common to have a massive dataset that makes training of classifiers impractical. To overcome this challenge, Li [79] proposed an extreme data point selection algorithm that selects a data point as a potential support vector based on geometrical observation. This approach underperforms for high dimensional datasets. Whereas, Sun et al. [80] proposed heuristic sample reduction (HSR) method based on k-means clustering. As an improvement to the model of Li [79], Zhu et al. [81] proposed a new boundary detection algorithm capable of working with high dimensional massive datasets. The extreme point selection algorithm reduces the training dataset size by 20%, whereas the method proposed by Zhu et al. reduces it by 80%–85%. Recently, Alam et al. [82] proposed a more robust sample reduction algorithm, where the mean vector of samples and angular directional distances of neighbours of a sample are used as key parameters. This approach reduces the training time of SVDD and suitable for high dimensional datasets too. The time complexity of this method is less compared to other pre-existing methods. OCSVCs are boundary-based classification algorithms, and there is enough scope to devise a more efficient algorithm to identify the most promising boundary points to reduce the computation overhead.

With the increasing volume of data and distributed processing in demand, Tax et al. [91] proposed an incremental SVDD approach to deal with the temporal arrival of samples. Later, Sillito et al. [96] proposed another incremental one-class learning algorithm for outlier detection. In this approach, outliers were identified with the help of probability distribution function, where initially, the training data distribution is obtained with the help of kernel density estimation. Afterwards, Tavakkoli et al. [97] proposed an incremental SVDD training mechanism that runs in constant time since its re-training is done only with support vectors. This incremental training algorithm is based on the theorem proposed by Osuna et al. [98], which concludes that a large QP problem can be broken into a series of smaller sub-problems. For large scale datasets, Kim et al. [99] proposed an incremental model that operates on the concept of sample margin. To work in distributed environment, Das et al. [101] proposed a distributed approach that comprises reduced computational cost and storage requirement. Afterwards, Castillo et al. [102] proposed another distributed OCSVM for big data environment which is unaffected by the position of the outliers and able to fit the data more precisely. Later, Hansi et al. [103] proposed a fast incremental SVDD (FISVDD) algorithm for online data. This method works on the observation that all potential support vectors have the same distance to the centre of the hypersphere. For anomaly detection in wireless sensor networks, two novel distributed one-class support vector machines were proposed by Miao et al. [104]. Later, Kefi et al. [107] proposed an incremental one-class covariance-guided support vector machine (iCOSVM). The iCOSVM improves over other incremental OCSVM by the integration of the covariance matrix into the optimization problem. In the present era of big and online data, more robust and adaptive distributed OCSVCs are required to handle the current requirements.

In Section 7, various important applications of OCSVCs have been discussed. This detailed survey proves that the scope of OCSVCs is diverse, and provides an up to date in-depth insight into OCSVCs to help in the formulation of novel future works in this field.

Declaration of competing interest

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

References

[1] T. Minter, Single-class classification, 1975.
[2] M.W. Koch, M.M. Moya, L.D. Hostetler, R.J. Fogler, Cueing, feature discovery, and one-class learning for synthetic aperture radar automatic target recognition, Neural Netw. 8 (7–8) (1995) 1081–1102.
[3] C.M. Bishop, Novelty detection and neural network validation, IEE Proc.-Vis. Image Signal Process. 141 (4) (1994) 217–222.
[4] G. Ritter, M.T. Gallegos, Outliers in statistical pattern recognition and an application to automatic chromosome classification, Pattern Recognit. Lett. 18 (6) (1997) 525–539.
[5] N. Japkowicz, Concept-Learning in the Absence of Counter-Examples: an Autoassociation-Based Approach to Classification, Rutgers University, 1999.
[6] O. Mazhelis, One-class classifiers: a review and analysis of suitability in the context of mobile-masquerader detection, S. Afr. Comput. J. 2006 (36) (2006) 29–48.
[7] S.S. Khan, M.G. Madden, One-class classification: taxonomy of study and review of techniques, Knowl. Eng. Rev. 29 (3) (2014) 345–374.
[8] M.A. Pimentel, D.A. Clifton, L. Clifton, L. Tarassenko, A review of novelty detection, Signal Process. 99 (2014) 215–249.
[9] R. Chalapathy, S. Chawla, Deep learning for anomaly detection: A survey, 2019, arXiv preprint arXiv:1901.03407.
[10] Y. Chen, L. Lu, P. Zhong, One-class support higher order tensor machine classifier, Appl. Intell. 47 (4) (2017) 1022–1030.
[11] A. Tajoddin, M. Abadi, Ramd: registry-based anomaly malware detection using one-class ensemble classifiers, Appl. Intell. 49 (7) (2019) 2641–2658.
[12] L.M. Manevitz, M. Yousef, One-class SVMs for document classification, J. Mach. Learn. Res. 2 (Dec) (2001) 139–154.
[13] G.S. Uttreshwar, A. Ghatol, Hepatitis B Diagnosis Using Logical Inference and Self-Organizing Map 1, Citeseer, 2008.
[14] G. Cohen, H. Sax, A. Geissbuhler, et al., Novelty detection using one-class parzen density estimator. an application to surveillance of nosocomial infections., Stud. Health Technol. Inf. 136 (2008) 21.
[15] M. Hejazi, Y.P. Singh, One-class support vector machines approach to anomaly detection, Appl. Artif. Intell. 27 (5) (2013) 351–366.
[16] L. Dong, L. Shulin, H. Zhang, A method of anomaly detection and fault diagnosis with online adaptive learning under small training samples, Pattern Recognit. 64 (2017) 374–385.
[17] W. Khreich, B. Khosravifar, A. Hamou-Lhadj, C. Talhi, An anomaly detection system based on variable n-gram features and one-class svm, Inf. Softw. Technol. 91 (2017) 186–197.
[18] L.A. Maglaras, J. Jiang, T.J. Cruz, Combining ensemble methods and social network metrics for improving accuracy of ocsvm on intrusion detection in scada systems, J. Inf. Secur. Appl. 30 (2016) 15–26.
[19] L. Yin, H. Wang, W. Fan, Active learning based support vector data description method for robust novelty detection, Knowl.-Based Syst. 153 (2018) 40–52.
[20] V. Vapnik, V. Vapnik, Statistical Learning Theory, wiley New York, 1998.
[21] S. Abe, Support Vector Machines for Pattern Classification, Vol. 2, Springer, 2005.
[22] D. Tomar, S. Agarwal, Twin support vector machine: a review from 2007 to 2014, Egypt. Inf. J. 16 (1) (2015) 55–69.

[23] J. Zhao, Y. Xu, H. Fujita, An improved non-parallel universum support vector machine and its safe sample screening rule, Knowl.-Based Syst. 170 (2019) 79–88.
[24] B. Richhariya, M. Tanveer, A reduced universum twin support vector machine for class imbalance learning, Pattern Recognit. (2020) 107150.
[25] S. Lu, H. Wang, Z. Zhou, All-in-one multicategory ramp loss maximum margin of twin spheres support vector machine, Appl. Intell. 49 (6) (2019) 2301–2314.
[26] D.M.J. Tax, One-Class Classification: Concept-Learning in the Absence of Counter-Examples (Ph. D. thesis), Delft University of Technology, 2001.
[27] D.M. Tax, R.P. Duin, Support vector domain description, Pattern Recognit. Lett. 20 (11–13) (1999) 1191–1199.
[28] D.M. Tax, R.P. Duin, Data domain description using support vectors., in: ESANN, Vol. 99, 1999, pp. 251–256.
[29] D.M. Tax, R.P. Duin, Combining one-class classifiers, in: International Workshop on Multiple Classifier Systems, Springer, 2001, pp. 299–308.
[30] D.M. Tax, R.P. Duin, Support vector data description, Mach. Learn. 54 (1) (2004) 45–66.
[31] D.M. Tax, R.P. Duin, Uniform object generation for optimizing one-class classifiers, J. Mach. Learn. Res. 2 (Dec) (2001) 155–173.
[32] B. SchölkopfÜ, R.C. Williamson, A. SmolaÜ, J. Shawe-TaylorÝ, Sv estimation of a distribution's support, 1999.
[33] B. Schölkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola, R.C. Williamson, Estimating the support of a high-dimensional distribution, Neural Comput. 13 (7) (2001) 1443–1471.
[34] C. Campbell, K.P. Bennett, A linear programming approach to novelty detection, in: Advances in Neural Information Processing Systems, 2001, pp. 395–401.
[35] K.-L. Li, H.-K. Huang, S.-F. Tian, W. Xu, Improving one-class SVM for anomaly detection, in: Machine Learning and Cybernetics, 2003 International Conference on, Vol. 5, IEEE, 2003, pp. 3077–3081.
[36] J. Luo, L. Ding, Z. Pan, G. Ni, G. Hu, Research on cost-sensitive learning in one-class anomaly detection algorithms, in: International Conference on Autonomic and Trusted Computing, Springer, 2007, pp. 259–268.
[37] P.-Y. Hao, et al., Fuzzy one-class support vector machines, Fuzzy Sets and Systems (159) (2008).
[38] Y.-S. Choi, Least squares one-class support vector machine, Pattern Recognit. Lett. 30 (13) (2009) 1236–1240.
[39] J. Yang, N. Zhong, P. Liang, J. Wang, Y. Yao, S. Lu, Brain activation detection by neighborhood one-class SVM, Cogn. Syst. Res. 11 (1) (2010) 16–24.
[40] L. Chen, P. Pu, Survey of Preference Elicitation Methods, Tech. rep., 2004.
[41] Y. Liu, B. Zhang, B. Chen, Y. Yang, Robust solutions to fuzzy one-class support vector machine, Pattern Recognit. Lett. 71 (2016) 73–77.
[42] F. Zhu, J. Yang, C. Gao, S. Xu, N. Ye, T. Yin, A weighted one-class support vector machine, Neurocomputing 189 (2016) 1–10.
[43] M. Ji, H.-J. Xing, Adaptive-weighted one-class support vector machine for outlier detection, in: Control and Decision Conference (CCDC), 2017 29th Chinese, IEEE, 2017, pp. 1766–1771.
[44] K. Wang, H. Xiao, Ellipsoidal data description, Neurocomputing 238 (2017) 328–339.
[45] Y. Tian, M. Mirzabagheri, S.M.H. Bamakan, H. Wang, Q. Qu, Ramp loss one-class support vector machine; a robust and effective approach to anomaly detection problems, Neurocomputing 310 (2018) 223–235.
[46] H.-J. Xing, M. Ji, Robust one-class support vector machine with rescaled hinge loss function, Pattern Recognit. 84 (2018) 152–164.
[47] H.-J. Xing, W.-T. Liu, Robust adaboost based ensemble of one-class support vector machines, Inf. Fusion 55 (2020) 45–58.
[48] D.M. Tax, P. Juszczak, Kernel whitening for one-class classification, Int. J. Pattern Recognit. Artif. Intell. 17 (03) (2003) 333–347.
[49] B. Silverman, D. Titterington, Minimum covering ellipses, SIAM J. Sci. Stat. Comput. 1 (4) (1980) 401–409.
[50] P. Kumar, E.A. Yildirim, Minimum-volume enclosing ellipsoids and core sets, J. Optim. Theory Appl. 126 (1) (2005) 1–21.
[51] B. Schölkopf, A. Smola, K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (5) (1998) 1299–1319.
[52] X. Huang, L. Shi, J.A. Suykens, Ramp loss linear programming support vector machine, J. Mach. Learn. Res. 15 (1) (2014) 2185–2211.
[53] G. Xu, Z. Cao, B.-G. Hu, J.C. Principe, Robust support vector machines based on the rescaled hinge loss function, Pattern Recognit. 63 (2017) 139–148.
[54] J. Sun, H. Fujita, P. Chen, H. Li, Dynamic financial distress prediction with concept drift based on time weighting combined with adaboost support vector machine ensemble, Knowl.-Based Syst. 120 (2017) 4–14.
[55] J. Sun, H. Li, H. Fujita, B. Fu, W. Ai, Class-imbalanced dynamic financial distress prediction based on adaboost-svm ensemble combined with smote and time weighting, Inf. Fusion 54 (2020) 128–144.
[56] D.M. Tax, R.P. Duin, Outliers and data descriptions, in: Proceedings of the 7th Annual Conference of the Advanced School for Computing and Imaging, 2001, pp. 234–241.
[57] G. Rätsch, S. Mika, B. Schölkopf, K.-R. Müller, Constructing boosting algorithms from svms: An application to one-class classification, IEEE Trans. Pattern Anal. Mach. Intell. (9) (2002) 1184–1199.
[58] L. Zhuang, H. Dai, Parameter estimation of one-class svm on imbalance text classification, in: Conference of the Canadian Society for Computational Studies of Intelligence, Springer, 2006, pp. 538–549.
[59] H. Wang, L. Zhang, Y. Xiao, W. Xu, An approach to choosing gaussian kernel parameter for one-class svms via tightness detecting, in: Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2012 4th International Conference on, Vol. 2, IEEE, 2012, pp. 318–323.
[60] Y. Xiao, H. Wang, L. Zhang, W. Xu, Two methods of selecting gaussian kernel parameters for one-class svm and their application to fault detection, Knowl.-Based Syst. 59 (2014) 75–84.
[61] A. Theissler, I. Dear, Autonomously determining the parameters for svdd with rbf kernel from a one-class training set, in: Proceedings of the WASET International Conference on Machine Intelligence, Vol. 7, pp. 1135–1143.
[62] Z. Ghafoori, S.M. Erfani, S. Rajasegarar, J.C. Bezdek, S. Karunasekera, C. Leckie, Efficient unsupervised parameter estimation for one-class support vector machines, IEEE Trans. Neural Netw. Learn. Syst. 29 (10) (2018) 5057–5070.
[63] J. Sun, J. Lang, H. Fujita, H. Li, Imbalanced enterprise credit evaluation with dte-sbd: decision tree ensemble based on smote and bagging with differentiated sampling rates, Inform. Sci. 425 (2018) 76–91.
[64] C. Zhang, J. Bi, S. Xu, E. Ramentol, G. Fan, B. Qiao, H. Fujita, Multi-imbalance: An open-source software for multi-class imbalance learning, Knowl.-Based Syst. 174 (2019) 137–143.
[65] D.M. Tax, K.-R. Müller, Feature extraction for one-class classification, in: Artificial Neural Networks and Neural Information Processing—ICANN/ICONIP 2003, Springer, 2003, pp. 342–349.
[66] H. Lian, On feature selection with principal component analysis for one-class svm, Pattern Recognit. Lett. 33 (9) (2012) 1027–1031.
[67] C. Liu, G. Wang, W. Ning, X. Lin, L. Li, Z. Liu, Anomaly detection in surveillance video using motion direction statistics, in: Image Processing (ICIP), 2010 17th IEEE International Conference on, IEEE, 2010, pp. 717–720.
[68] K. Feng, Z. Jiang, W. He, B. Ma, A recognition and novelty detection approach based on curvelet transform, nonlinear pca and svm with application to indicator diagram diagnosis, Expert Syst. Appl. 38 (10) (2011) 12721–12729.
[69] I. Jolliffe, Principal component analysis, in: International Encyclopedia of Statistical Science, Springer, 2011, pp. 1094–1096.
[70] S. Geman, E. Bienenstock, R. Doursat, Neural networks and the bias/variance dilemma, Neural Comput. 4 (1) (1992) 1–58.
[71] T. Heskes, Bias/variance decompositions for likelihood-based estimators, Neural Comput. 10 (6) (1998) 1425–1433.
[72] S.-B. Cho, Recognition of unconstrained handwritten numerals by doubly self-organizing neural network, in: Proceedings of 13th International Conference on Pattern Recognition, Vol. 4, IEEE, 1996, pp. 426–430.
[73] Y.-S. Jeong, I.-H. Kang, M.-K. Jeong, D. Kong, A new feature selection method for one-class classification problems, IEEE Trans. Syst. Man Cybern. C 42 (6) (2012) 1500–1509.
[74] P. Nagabhushan, H. Meenakshi, Target class supervised feature subsetting, Int. J. Comput. Appl. 91 (12) (2014) 11–23.
[75] M. Yousef, M.D. Saçar Demirci, W. Khalifa, J. Allmer, Feature selection has a large impact on one-class classification accuracy for micrornas in plants, Adv. Bioinf. 2016 (2016).
[76] S. Griffiths-Jones, H.K. Saini, S. van Dongen, A.J. Enright, Mirbase: tools for microrna genomics, Nucleic Acids Res. 36 (suppl_1) (2007) D154–D158.
[77] C. Liu, W. Wang, M. Wang, F. Lv, M. Konan, An efficient instance selection algorithm to reconstruct training set for support vector machine, Knowl.-Based Syst. 116 (2017) 58–73.
[78] J. Nalepa, M. Kawulok, Selecting training sets for support vector machines: a review, Artif. Intell. Rev. (2018) 1–44.
[79] Y. Li, Selecting training points for one-class support vector machines, Pattern Recognit. Lett. 32 (11) (2011) 1517–1522.
[80] W. Sun, J. Qu, Y. Chen, Y. Di, F. Gao, Heuristic sample reduction method for support vector data description, Turkish J. Electr. Eng. Comput. Sci. 24 (1) (2016) 298–312.
[81] F. Zhu, N. Ye, W. Yu, S. Xu, G. Li, Boundary detection and sample reduction for one-class support vector machines, Neurocomputing 123 (2014) 166–173.
[82] S. Alam, S.K. Sonbhadra, S. Agarwal, P. Nagabhushan, M. Tanveer, Sample reduction using farthest boundary point estimation (fbpe) for support vector data description (svdd), Pattern Recognit. Lett. (2020).
[83] J. Kivinen, A.J. Smola, R.C. Williamson, Online learning with kernels, IEEE Trans. Signal Process. 52 (8) (2004) 2165–2176.
[84] S. Ruping, Incremental learning with support vector machines, in: Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on, IEEE, 2001, pp. 641–642.

[85] N.A. Syed, S. Huan, L. Kah, K. Sung, Incremental Learning with Support Vector Machines, Citeseer, 1999.
[86] Z. Li, J. Zhang, S. Hu, Incremental support vector machine algorithm based on multi-kernel learning, J. Syst. Eng. Electron. 22 (4) (2011) 702–706.
[87] J. Zheng, F. Shen, H. Fan, J. Zhao, An online incremental learning support vector machine for large-scale data, Neural Comput. Appl. 22 (5) (2013) 1023–1035.
[88] Y. Qin, D. Li, A. Zhang, A new svm multiclass incremental learning algorithm, Math. Probl. Eng. 2015 (2015).
[89] J. Liu, T.-J. Pan, Z.-Y. Zhang, Incremental support vector machine combined with ultraviolet-visible spectroscopy for rapid discriminant analysis of red wine, J. Spectrosc. 2018 (2018).
[90] J. Xu, C. Xu, B. Zou, Y.Y. Tang, J. Peng, X. You, New incremental learning algorithm with support vector machines, IEEE Trans. Syst. Man Cybern.: Syst. (99) (2018) 1–12.
[91] D.M. Tax, P. Laskov, Online svm learning: from classification to data description and back, in: Neural Networks for Signal Processing, 2003. NNSP'03. 2003 IEEE 13th Workshop on, IEEE, 2003, pp. 499–508.
[92] B. Schölkopf, A.J. Smola, R.C. Williamson, P.L. Bartlett, New support vector algorithms, Neural Comput. 12 (5) (2000) 1207–1245.
[93] V.N. Vapnik, An overview of statistical learning theory, IEEE Trans. Neural Netw. 10 (5) (1999) 988–999.
[94] G. Cauwenberghs, T. Poggio, Incremental and decremental support vector machine learning, in: Advances in Neural Information Processing Systems, 2001, pp. 409–415.
[95] P. Laskov, C. Gehl, S. Krüger, K.-R. Müller, Incremental support vector learning: Analysis, implementation and applications, J. Mach. Learn. Res. 7 (Sep) (2006) 1909–1936.
[96] R.R. Sillito, R.B. Fisher, Incremental one-class learning with bounded computational complexity, in: International Conference on Artificial Neural Networks, Springer, 2007, pp. 58–67.
[97] A. Tavakkoli, M. Nicolescu, G. Bebis, et al., Incremental svdd training: Improving efficiency of background modeling in videos, in: International Conference on Signal and Image Processing, Kailua-Kona, HI, 2008, pp. 1–6.
[98] E. Osuna, R. Freund, F. Girosi, An improved training algorithm for support vector machines, in: Neural Networks for Signal Processing [1997] VII. Proceedings of the 1997 IEEE Workshop, IEEE, 1997, pp. 276–285.
[99] P.J. Kim, H.J. Chang, J.Y. Choi, Fast incremental learning for one-class support vector classifier using sample margin information, in: Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, IEEE, 2008, pp. 1–4.
[100] A. Bounsiar, M.G. Madden, Support vector one-class classification for multiple-distribution data, in: Signal Processing Conference, 2010 18th European, IEEE, 2010, pp. 1189–1193.
[101] K. Das, K. Bhaduri, P. Votava, Distributed anomaly detection using 1-class svm for vertically partitioned data, Stat. Anal. Data Min.: ASA Data Sci. J. 4 (4) (2011) 393–406.
[102] E. Castillo, D. Peteiro-Barral, B.G. Berdiñas, O. Fontenla-Romero, Distributed one-class support vector machine, Int. J. Neural Syst. 25 (07) (2015) 1550029.
[103] H. Jiang, H. Wang, W. Hu, D. Kakde, A. Chaudhuri, Fast incremental svdd learning algorithm with the gaussian kernel, 2017, arXiv preprint arXiv:1709.00139.
[104] X. Miao, Y. Liu, H. Zhao, C. Li, Distributed online one-class support vector machine for anomaly detection over networks, IEEE Trans. Cybern. 49 (4) (2018) 1475–1488.
[105] A. Rahimi, B. Recht, Random features for large-scale kernel machines, in: Advances in Neural Information Processing Systems, 2008, pp. 1177–1184.
[106] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1) (1996) 267–288.
[107] T. Kefi-Fatteh, R. Ksantini, M.-B. Kaâniche, A. Bouhoula, A novel incremental one-class support vector machine based on low variance direction, Pattern Recognit. 91 (2019) 308–321.
[108] M. Markou, S. Singh, Novelty detection: a review—part 1: statistical approaches, Signal Process. 83 (12) (2003) 2481–2497.
[109] C. Désir, S. Bernard, C. Petitjean, L. Heutte, One class random forests, Pattern Recognit. 46 (12) (2013) 3490–3506.
[110] A.O. Adewumi, A.A. Akinyelu, A survey of machine-learning and nature-inspired based credit card fraud detection techniques, Int. J. Syst. Assur. Eng. Manag. 8 (2) (2017) 937–953.
[111] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.
[112] B. Kiran, D. Thomas, R. Parakkal, An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos, J. Imaging 4 (2) (2018) 36.
[113] M. Koppel, J. Schler, Authorship verification as a one-class classification problem, in: Proceedings of the Twenty-First International Conference on Machine Learning, ACM, 2004, p. 62.
[114] D. Bertolini, L.S. Oliveira, E. Justino, R. Sabourin, Reducing forgeries in writer-independent off-line signature verification through ensemble of classifiers, Pattern Recognit. 43 (1) (2010) 387–396.
[115] R. Kumar, J. Sharma, B. Chanda, Writer-independent off-line signature verification using surroundedness feature, Pattern Recognit. Lett. 33 (3) (2012) 301–308.
[116] M.A. Ferrer, J. Vargas, A. Morales, A. Ordóñez, Robustness of offline signature verification based on gray level features, IEEE Trans. Inf. Forensics Secur. 7 (3) (2012) 966–977.
[117] Y. Guerbai, Y. Chibani, N. Abbas, One-class versus bi-class svm classifier for off-line signature verification, in: Multimedia Computing and Systems (ICMCS), 2012 International Conference on, IEEE, 2012, pp. 206–210.
[118] M.K. Kalera, S. Srihari, A. Xu, Offline signature verification and identification using distance statistics, Int. J. Pattern Recognit. Artif. Intell. 18 (07) (2004) 1339–1360.
[119] Y. Guerbai, Y. Chibani, B. Hadjadji, The effective use of the one-class svm classifier for handwritten signature verification based on writer-independent parameters, Pattern Recognit. 48 (1) (2015) 103–113.
[120] J. Ma, G. Plonka, The curvelet transform, IEEE Signal Process. Mag. 27 (2) (2010) 118–133.
[121] W. Bouamra, C. Djeddi, B. Nini, M. Diaz, I. Siddiqi, Towards the design of an offline signature verifier based on a small number of genuine samples for training, Expert Syst. Appl. 107 (2018) 182–195.
[122] R. Solera-Ureña, J. Padrell-Sendra, D. Martín-Iglesias, A. Gallardo-Antolín, C. Peláez-Moreno, F. Díaz-de María, Svms for automatic speech recognition: a survey, in: Progress in Nonlinear Speech Processing, Springer, 2007, pp. 190–216.
[123] A. Rabaoui, M. Davy, S. Rossignol, Z. Lachiri, N. Ellouze, Improved one-class SVM classifier for sounds classification, in: Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on, IEEE, 2007, pp. 117–122.
[124] A. Rabaoui, H. Kadri, Z. Lachiri, N. Ellouze, One-class svms challenges in audio detection and classification applications, EURASIP J. Adv. Signal Process. 2008 (1) (2008) 834973.
[125] B. Fergani, M. Davy, A. Houacine, Speaker diarization using one-class support vector machines, Speech Communication 50 (5) (2008) 355–365.
[126] S. Besbes, Z. Lachiri, Classification of speech under stress based on cepstral features and one-class svm, in: Control, Automation and Diagnosis (ICCAD), 2017 International Conference on, IEEE, 2017, pp. 213–218.
[127] M. Sahidullah, G. Saha, Design, analysis and experimental evaluation of block based transformation in mfcc computation for speaker recognition, Speech Communication 54 (4) (2012) 543–565.
[128] R. Fathima, P. Raseena, Gammatone cepstral coefficient for speaker identification, Int. J. Adv. Res. Electr. Electron. Instrum. Eng. 2 (2013).
[129] K. Fu, C. Gong, Y. Qiao, J. Yang, I. Guy, One-class svm assisted accurate tracking, in: Distributed Smart Cameras (ICDSC), 2012 Sixth International Conference on, IEEE, 2012, pp. 1–6.
[130] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 1, IEEE, 2005, pp. 886–893.
[131] Z. Kalal, J. Matas, K. Mikolajczyk, Online Learning of Robust Object Detectors During Unstable Tracking, IEEE, 2009.
[132] V. Mygdalis, A. Iosifidis, A. Tefas, I. Pitas, Exploiting subclass information in one-class support vector machine for video summarization, in: Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on, IEEE, 2015, pp. 2259–2263.
[133] H. Lin, J.D. Deng, B.J. Woodford, Anomaly detection in crowd scenes via online adaptive one-class support vector machines, in: Image Processing (ICIP), 2015 IEEE International Conference on, IEEE, 2015, pp. 2434–2438.
[134] T. Prayoonpitak, S. Wongsa, International Journal of Materials, Mechanics and Manufacturing 5 (3) (2017) 146–152,
[135] Y. Xiao, H. Wang, W. Xu, J. Zhou, Robust one-class svm for fault detection, Chemometr. Intell. Lab. Syst. 151 (2016) 15–25.
[136] J.J. Downs, E.F. Vogel, A plant-wide industrial process control problem, Comput. Chem. Eng. 17 (3) (1993) 245–255.
[137] A. Beghi, L. Cecchinato, C. Corazzol, M. Rampazzo, F. Simmini, G.A. Susto, A one-class svm based tool for machine learning novelty detection in hvac chiller systems, IFAC Proc. Vol. 47 (3) (2014) 1953–1958.
[138] M.C. Comstock, J.E. Braun, R. Bernhard, Development of Analysis Tools for the Evaluation of Fault Detection and Diagnostics in Chillers, Purdue University, 1999.
[139] S. Yin, X. Zhu, C. Jing, Fault detection based on a robust one class support vector machine, Neurocomputing 145 (2014) 263–268.
[140] J. Muñoz-Marí, L. Bruzzone, G. Camps-Valls, A support vector domain description approach to supervised classification of remote sensing images, IEEE Trans. Geosci. Remote Sens. 45 (8) (2007) 2683–2692.
[141] X. Zhang, H. Xu, P. Li, Hyperspectral classification using spectral magnitude and gradient, in: Hyperspectral Image and Signal Processing (WHISPERS), 2012 4th Workshop on, IEEE, 2012, pp. 1–4.

[142] F.A. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, A. Goetz, The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data, Remote Sens. Environ. 44 (2–3) (1993) 145–163.

[143] X. Xu, T. Cheng, X. Yao, Y. Tian, Y. Zhu, W. Cao, Mapping rice planting area from Landsat 8 imagery using one-class support vector machine, in: Proceedings of the Fifth International Conference on Agro-Geoinformatics, Tianjin, China, 2016, pp. 18–20.

[144] Y. Tang, E. Fan, C. Yan, X. Bai, J. Zhou, Discriminative weighted band selection via one-class svm for hyperspectral imagery, in: Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International, IEEE, 2016, pp. 2765–2768.

[145] M.S. Roodposhti, T. Safarrad, H. Shahabi, Drought sensitivity mapping using two one-class support vector machine algorithms, Atmos. Res. 193 (2017) 73–82.

[146] Y. Wang, J. Wong, A. Miner, Anomaly intrusion detection using one class svm, in: Information Assurance Workshop, 2004, pp. 358–364.

[147] L.A. Maglaras, J. Jiang, T. Cruz, Integrated ocsvm mechanism for intrusion detection in scada systems, Electron. Lett. 50 (25) (2014) 1935–1936.