

Review of Domain-Based Novelty Detection Using One Class Support Vector Machine

Hayder A. Hussein^{1,2}, Said Amirul Anwar^{2*}



¹ Computer Networks Engineering, College of Information Engineering, Al-Nahrain University, Baghdad 64074, Iraq

² Faculty of Intelligent Computing, Universiti Malaysia Perlis (UniMAP), Arau 02600, Malaysia

Corresponding Author Email: said@unimap.edu.my

Copyright: ©2025 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.301211>

Received: 12 October 2025

Revised: 21 December 2025

Accepted: 26 December 2025

Available online: 31 December 2025

Keywords:

novelty detection, anomaly detection, One-Class Support Vector Machines, domain-description methods, one-class classification, SVDD

ABSTRACT

Novelty detection is a critical task in machine learning and data mining, aiming to identify previously unseen or abnormal patterns that are not represented during model training. Among the available novelty detection paradigms, domain-description approaches are particularly attractive because they learn an explicit boundary of normal data without strong distributional assumptions. This paper provides a focused review of domain-based novelty detection methods with emphasis on Support Vector Machine (SVM) formulations, particularly Support Vector Data Description (SVDD) and One-Class Support Vector Machines (OCSVM). We summarize the theoretical foundations of one-class classification and review recent research that enhances SVDD and OCSVM through robust boundary learning, improved feature representations, and computational efficiency. Based on the analyzed literature, the main technical directions for improving OCSVM and SVDD can be grouped into three trends: (i) robustification via modified loss functions and outlier-resistant formulations, (ii) integration with feature learning frameworks such as deep models and hybrid architectures, and (iii) acceleration strategies for large-scale and high-dimensional settings. Despite consistent performance improvements across applications, parameter sensitivity, optimization complexity, and limited adaptability under evolving data distributions remain persistent challenges. Finally, we outline concrete research opportunities toward scalable, adaptive, and self-tuning domain-description models for reliable deployment in real-world novelty detection scenarios.

1. INTRODUCTION

One of the primary problems with machine learning and data mining is novelty detection (ND), sometimes referred to as anomaly detection, outlier detection, or one-class classification. Differentiating recognized items (regular patterns) from samples that are abnormal (outliers) is ND's primary duty [1]. The capability to identify novel classes could possess a substantial and favorable outcome, wherein the test data (or unlabelled data) might comprise details regarding items not known at the time of the training process [2]. ND encompasses learning a model that can recognize any departure from normalcy by absorbing the characteristics of the training dataset's normal data samples. Real-world applications include, among other things, currency validation, machine failure detection, medical diagnostics, and user verification in computer systems. Normal data samples are usually abundant in ND-related applications, but abnormal data samples are sometimes scarce or, in certain situations, nonexistent. As a result, the majority of ND algorithms prioritize the typical data and primarily use it to create a data description [3].

The predictive performance in ND is determined by measuring the model's ability to classify samples belonging to the normal category, i.e., elements that have properties alike the elements used for training, and distinguish those samples,

referred to as novel, outliers, or abnormal [4]. The novelty detection technique is employed to devise a model that isolates new patterns from a specified dataset. The model must be formulated in a manner that inputs having features different from those used during training should have the inputs classified as novel patterns, while those matching the training inputs should be classified accordingly [5]. In recent years, numerous fields have focused on the function of one-class categorization in pattern recognition. The approach of simulating a widely dispersed class to categorize an unknown testing class is known as one-class classification, and it is frequently employed for ND. When there is just one known class available to train the model (i.e., classifier) rather than two or more, it is regarded as a specific case of a classification problem. The ability of the constructed model to differentiate samples of the normal class that is, samples with traits comparable to those used in training from the other examples, which are referred to be abnormal or novel, or outliers, is the basis for measuring prediction performance in ND [6].

ND can also be viewed as much like novel class detection [1, 7, 8]. Two circumstances should be substantiated to affirm the advent of a novel class: the threshold condition and the cohesion-separation condition. Cluster-specific assumptions are made for the first condition, where elements of the novel class are assumed to be like the other elements of the class; however, those elements are different from the elements

belonging to other classes. The second condition indicates that the novel class element count must exceed a specified threshold. The threshold is employed for differentiating between the novel class and outlier's cases; when the quantity of applicants is fewer than the stated level, we presume that they represent anomalies in the current classes. Experts determine the requisite threshold value, and it is specific to the application [9]. In the wake of the specified criteria, a learner must evaluate sets of instances to ascertain the creation of an additional class. Thus, current techniques are either based on chunks [10-12] or according to timing limitations [13, 14].

Although terms such as anomaly detection, outlier detection, novelty detection, and one-class classification have nuanced differences in the literature, this review treats them interchangeably, focusing on their shared objective of identifying deviations from learned normal data using domain-based methods.

With a focus on domain-based methods for novelty detection in the machine learning literature, this survey attempts to present an up-to-date and organized summary of recent research and approaches to novelty detection using a semi-supervised learning approach using One-Class Support Vector Machines (OCSVM). The rest of the paper is organized in this manner: Section 2 discusses the novelty detection methodologies, including various categories of approaches from the literature. Meanwhile, Section 3 delves into Support Vector Machines (SVM), where this section details a support vector data description review (SVDD) and OCSVM are presented. This is followed by a detailed technical review of the fundamental algorithmic improvement made to OCSVM from the literature, and a review of real-world applications of novelty detection using OCSVM. At the end, the review is summarized in Section 4 and followed by the concluding remarks.

2. NOVELTY DETECTION METHODOLOGIES

Novelty detection (ND), also referred to as anomaly detection or one-class classification, aims to identify patterns that deviate from the characteristics of normal data observed during training. Depending on how normal behavior is modeled, existing novelty detection techniques can be broadly categorized into several methodological families. This section briefly reviews these categories with the specific purpose of positioning domain-description methods, particularly those based on Support Vector Machines (SVMs), relative to alternative approaches [15].

In probabilistic approaches, normal behavior is characterized by calculating the data's probability density function, and samples with low likelihood under the learned distribution are considered novel. While effective for low-dimensional and well-behaved data, these methods often suffer in high-dimensional settings due to density estimation complexity and sensitivity to distributional assumptions. In contrast, distance-based methods rely on similarity measures, such as nearest-neighbor distances or clustering structures, under the assumption that normal data form dense regions in feature space; however, their effectiveness degrades with increasing dimensionality and data volume [16].

Reconstruction-based methods, including autoencoder-based models, detect novelty by measuring reconstruction error when projecting input data through a learned representation. These approaches are capable of modeling

complex nonlinear data structures but typically require large amounts of training data and careful model tuning. Moreover, their decision boundaries are implicit, which can limit interpretability and reliability in industrial or safety-critical applications [17].

In contrast to the above paradigms, domain-description methods explicitly aim to identify a border enclosing typical data without assuming a specific underlying probability distribution. Rather than modeling density or distances, these approaches characterize the *support* of the normal data distribution, making them particularly suitable for novelty detection scenarios involving high-dimensional data, limited training samples, or unknown anomaly characteristics [15].

Among domain-description methods, Support Vector Machine-based approaches, notably OCSVM and Support Vector Data Description (SVDD), are the most widely adopted. These methods formulate novelty detection as an optimization problem that constructs a tight boundary around normal data by maximizing margin or minimizing hypersphere volume in a high-dimensional feature space. By relying on boundary-defining samples (support vectors), SVM-based domain-description methods achieve strong generalization performance and robustness [10].

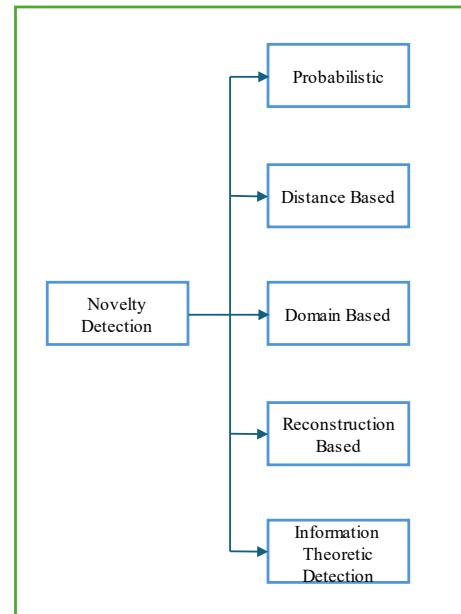


Figure 1. Categories of novelty detection methods and position of domain-description approaches (adapted from the study [18])

Figure 1 summarizes the main categories of novelty detection techniques and highlights the position of domain-description methods within the broader landscape. As illustrated, SVM-based techniques constitute a central and well-established branch of domain-description approaches, motivating their extensive adoption across diverse application domains [18].

Given these characteristics, domain-description methods based on OCSVM and SVDD have become a dominant framework for novelty detection. The following section therefore focuses on the theoretical foundations of SVM-based domain-description methods, followed by a detailed review of recent algorithmic improvements and real-world applications [15].

The novelty detection can be categorized into five broad classes, as shown in Figure 1 [18]. As can be seen from this figure, various novelty detection techniques have been developed, each leveraging distinct methodological frameworks to isolate anomalous instances. Probabilistic techniques calculate the density of the normal class and interpret low-density regions as indicative of abnormality, assuming novel data exhibits a low probability under the modeled distribution. Distance-based methods rely on nearest neighbor and clustering analysis, positing that normal data points are tightly grouped, while anomalies lie at greater distances from their nearest neighbors. Reconstruction-based approaches train models to reproduce input data and assess novelty through reconstruction error, where larger discrepancies signify deviations from the learned structure. On the other hand, domain-based techniques characterize the data distribution by forming boundaries around the normal class without focusing on density peaks, effectively capturing the support of the training distribution. Information-theoretic approaches evaluate the informational complexity of data using metrics such as entropy, detecting novelty based on significant changes in the information content introduced by new observations. These frameworks collectively address the diverse structural properties of data in unsupervised anomaly detection tasks [15].

Based on these, the distance-based novelty detection methods are the most used, which in this technique the statistical similarity metrics that classify features using clearly specified distance measurements. Statistical distances that quantify the (dis)similarity between two sets of features whether univariate or multivariate are the foundation of these techniques, pertinent to the damaged and undamaged circumstances, respectively. However, when dealing with enormous volumes of random high-dimensional characteristics, such methods do not appear to be efficient or successful. Because of their complicated High computing costs, substantial data storage requirements, and high-dimensional features like time series model residuals may also negatively affect the effectiveness of machine learning and novelty detection algorithms and result in significant limitations during the decision-making stage [17].

In certain instances, each sample was assigned a novelty score determined by a distance metric, such Euclidean. The greatest Euclidean distance between every sample of normal training data and the calculation of the centroid of all normal training data. The calculation of a test sample's distances from the normal data centroid defines its output. At the centroid, a spherical border of radius is defined by a certain threshold. If the distance exceeds a predetermined threshold, the test sample is deemed abnormal in relation to the existing normal training data. It could be possible to generalize to normal data that is absent from the training set by using a threshold value to define a region that is larger than the one inhabited by normal training data [19].

Domain-based methods for detecting novelty do not assume anything regarding the data's distribution; instead, they use just the data that are closest to the novelty boundary to establish its position. The SVM based approach is one of the domain-based novelty identification techniques are the most used in the literature. Among them, the One-Class SVM (OC-SVM) and Support Vector Data Description (SVDD) [20]. For domain-based techniques, a boundary that considers the training dataset's structure must be established. Since these techniques define the target class border, or the domain, rather

than the class density, they are usually insensitive to the sample and density of the target class. The placement of unknown data concerning the boundary then determines their class membership. Like two-class SVM, novelty detection SVM (sometimes referred to as OC-SVM in the literature) uses only the support vectors, or the data that are closest to the novelty border (in the converted space), to identify where the novelty boundary is. When determining the novelty, all other training set data those which are not support vectors, are ignored [15].

Among the various categories of novelty detection techniques, domain-description methods have attracted particular because of their capacity to model complicated data boundaries without strong distributional assumptions. SVM-based approaches, especially OCSVM and SVDD, represent the most prominent realizations of this paradigm. The following section therefore focuses on these methods, their theoretical foundations, and recent algorithmic advancements.

3. SUPPORT VECTOR MACHINES

SVM is a widely used classifier for generating decision boundaries that divide data into various classes. For binary pattern classification of linearly separable data, the original SVM network is a perfect fit. A hyperplane that maximizes the distance between two classes is used by the SVM. Support vectors are the training points that are located close to the border, establishing this dividing edge. Numerous adjustments and enhancements have been made to the original concept since it was first proposed. A technique called Robust Support Vector Machines (RSVM) deals with the over-fitting issue that noise in the training dataset causes. With this method, the standard SVM is combined with an averaging technique (class centre) to smooth the decision surface and regulate the degree of regularization [21].

3.1 Support Vector Data Description

One commonly used method of a one-class classifier is SVDD. This technique maps target data into a high-dimensional feature space, intending to identify a set of support vectors (SV) representing the spherical border of the target data. The procedure takes place in feature space; hence, the description boundary of SVDD is adjustable. To make up for shortcomings in earlier one-class classifier research, SVM served as the model for the development of SVDD. Numerous classification techniques before the usage of support vectors relied on estimating the target data set's probability distribution [15].

Recently, certain SVDD technique extensions have been put forth to enhance the hyper-spherically shaped novelty boundary's boundaries. The first extension is proposed in the previous study [22], where a new SVDD incorporates the idea of density weighting—that is, assigning each data point a weight based on its relative density, as determined by applying the k-nearest neighbor (k-NN) method to the target data's density distribution. This novel approach emphasizes data points in high-density regions by incorporating the additional weight into the SVDD search for an ideal description. Eventually, the optimal description moves to these places.

Recent studies further extend Deep SVDD by integrating representation learning with contrastive objectives, improving feature compactness for normal samples while enhancing

separability from abnormal patterns. For example, contrastive Deep SVDD frameworks combine contrastive learning and SVDD loss to strengthen anomaly discrimination in complex feature spaces [23].

On the other hand, Yin et al. [24] introduced an SVDD technique for robust novelty detection that is based on active learning. It may lessen the quantity of labelled data by utilizing a framework for active learning, generalize the data spread, and lessen the effect of noise by directing the selection process with the help of the local density.

A deep learning neural network model that adopts the SVDD was proposed by Kim et al. [25]. A variation of the SVM, the SVDD achieves a maximal margin in one-class classification tasks, resulting in great generalization performance. The goal of the suggested model is to achieve deep learning's representational power. With the SVDD, generalization performance is preserved. In the study by Wang and Cha [26], according to the multivariate statistics known as Stahel-Donoho (SD) outlying-ness in an arbitrary kernel space, the authors have suggested giving each observation a weight. Weighted penalties, whose weights depend smoothly on the outlying-ness criteria, were presented as a novel robust SVDD. By assigning a weight to each data item based on a few chosen weight function families. In order to find the minimum-sized hypersphere, SVDD down-weights observations containing a lot of outliers, which could be mistakes, recordings from unusual situations, or members of a different population. To create a better model for the regular data, it is crucial to be able to identify these observations during SVDD training.

A two-phase intelligent fault detection approach for rotating equipment that combines optimized SVDD and optimized SVM was proposed by Zhang et al. [27]. To be more precise, SVM is used for fault identification, and SVDD is used for fault detection. The grasshopper optimization algorithm (GOA) optimizes the parameters of SVM and SVDD. The input feature vector for SVDD and SVM is extracted using multiscale entropy (ME). The advantages and disadvantages of eight distinct entropy-based indicators for feature extraction are examined within the parameters of the suggested methodology. A technique for selecting features when building the SVDD model with only normal samples is shown.

Meanwhile, to efficiently monitor the nonlinear plantwide processes, a decentralized fault detection and diagnosis technique was suggested by Wang et al. [28]. It has two main themes: fault diagnosis based on SVDD and mutual information-Louvain-based process decomposition. First, the plantwide process is first mapped as an undirected graph that matches the process structure and mechanism knowledge. A Louvain algorithm with MI correlation is suggested to finely break down the process into manageable sub-blocks. Mutual Information (MI) is provided to illustrate the degree of correlation between various nodes (i.e., process variables). For each sub-block, a decentralized defect detection technique based on SVDD is then introduced, and the associated variable contribution rate is calculated. Decentralized SVDD (DSVDD), which identifies the abnormal status and pinpoints the variable causing this fault, is the basis for local fault diagnosis and detection. The DSVDD model is used to identify faults and determine their causes by calculating the corresponding contribution rate.

In the study by Rahimzadeh Arashloo [29], the author has introduced a dual norm into the goal function, demonstrating a way for managing the intrinsic sparsity or consistency of the issue to improve descriptive capacity. By generalizing the

model to ℓ_p -norm where $p \geq 1$, the proposed approach enables formulating a non-linear cost for slacks. Also, extend the proposed works of ℓ_p -norm concept from a strictly one-class environment to the training situation and demonstrate the benefits of the suggested extension by include marked negative objects.

DR-SVDD, or dynamic radius support vector data description, is a new fault diagnosis technique. was proposed by Lu et al. [30] to efficiently identify the fuel cell system with proton exchange membranes faults. Compared to the classic SVDD and enhanced SVDDs, this approach takes into account the distribution properties of the training set sample as well as the SVDD hypersphere radius information to provide a more sufficient and accurate description of the sample data. New slack variables are added to the model to improve its generalization performance, resilience, fault-tolerance, and computational complexity.

The Sparrow Search Algorithm (SSA) was introduced by Pan et al. [31] into the parameter optimization process of SVDD. This method reads the training data then determines and arranges the fitness values for every sparrow based on the characteristics that correspond to its location. The training samples are divided into two groups when SVDD is trained using a single group at random; While the other group conducts the test, one group trains the SVDD model determine how many samples are not in the hypersphere. This is known as the fitness value. The sparrow that has the lowest fitness was chosen as the elite, and its position data was recorded to ascertain whether the fitness was below or equal to the error rate threshold. The computed error rate is then used as the fitness value.

In the study by Wu et al. [32], the authors have proposed the Part Interval Stacked Auto-encoder and Support Vector Data Description (PISAE-SVDD), a new process monitoring and fault diagnosis algorithm. This algorithm improves the Stacked Auto-encoder's (SAE) loss function. The mean square error of the initial input and output data is the reconstruction error value in certain measurement data. In contrast, a specific acceptable range for the reconstruction error value is provided by uncertain measurement data. Within this range, the measurement data's error value is regarded as zero. In the same manner as for specific measurement data, the reconstruction error value of measurement data that exceeds the permitted range is computed. Meanwhile, SAE's powerful nonlinear characterization capability is used to retrieve the distinctive information of the industrial process. Then, using the SAE feature information data taken from normal samples, SVDD is utilized to determine the control limit of the fluctuation range of the normal functioning state.

A Broad Learning Systems (BLS)-based weighted SVDD algorithm was proposed by Huang et al. [33], this increases the SVDD model's robustness during training by introducing reconstruction of a few atypical samples and error weights. SVDD was utilized to address issues where the proportion of sample categories of data is highly uneven, and the BLS was enhanced to produce a new model for data reconstruction.

On the other hand, a novel approach of a pinball loss SVDD was suggested in order to find outliers in the previous study [34]. In this approach, the sphere classifier uses all the training data, including the ones that are located inside the sphere. Because a little quantity of noisy data has minimal effect regarding the classifier, this technique is more resistive to noise and achieves dispersion minimization in the sphere center. This technique is superior in two ways: first, it employs

the pinball loss, which improves our approach's robustness and reduces scatter in the sphere center, in contrast to the traditional SVDD, which uses the hinge loss function and is noise-sensitive; second, it can be distinguished from the existing weight-varying anti-noise SVDD techniques, which take more pre-processing time to produce.

In industrial anomaly detection, recent improvements to Deep SVDD emphasize robust feature extraction and patch-level representation to handle localized defects, demonstrating improved detection stability in real deployment contexts [35].

3.1.1 Summary

Overall, the reviewed SVDD extensions reveal a clear research trend toward improving robustness and adaptability of the hypersphere boundary under real-world data imperfections. Density-weighted and active-learning-based SVDD variants consistently enhance sensitivity to local data structure, particularly in noisy or imbalanced datasets. Meanwhile, integration with deep learning models enables SVDD to handle high-dimensional and nonlinear feature representations more effectively. However, these improvements often introduce increased computational complexity and additional hyperparameters, which can limit scalability and hinder deployment in real-time systems. Despite performance gains, most SVDD variants remain sensitive to kernel and parameter selection, highlighting the need for adaptive or self-tuning mechanisms in future research.

3.2 One-Class Support Vector Machines

OCSVM is a popular method that is a reliable classifier for unsupervised anomaly detection. However, in some cases, it is not resistant to outliers. A few of its variations have recently been put forth to increase the resilience against outliers. The fundamental concept of OCSVM is to locate in the feature space a hyperplane $(w \cdot \phi(x)) - \rho = 0$ that, with the greatest possible margin, divides sample images from the origin. The following is the primal optimization problem [36].

$$\begin{aligned} & \min_{w, \xi, \rho} \frac{1}{2} \|w\|^2 - \rho + \frac{1}{\nu l} \sum_{i=1}^n \xi_i \\ & \text{s.t. } (w \cdot \phi(x_i)) \geq \rho, \xi_i \geq 0, i = 1, \dots, n \end{aligned} \quad (1)$$

where, l represents all of the training samples, ν is a trade-off parameter, and ξ are slack variables. Meanwhile, w , ρ , and ξ are the weight vector, offset parameter, and i -th training sample, respectively. This is a convex optimization problem, and its dual problem, which is represented below, can be used to solve it.

$$\begin{aligned} & \max_{\alpha} -\frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j \kappa(x_i, x_j) \\ & \text{s.t. } 0 \leq \alpha_i \leq \frac{1}{\nu l}, \sum \alpha_i = 1 \end{aligned} \quad (2)$$

where, $\kappa(x_i, x_j)$ denotes a kernel function. A typical kernel function adopted by researchers is the Gaussian Radial Basis Function (RBF), which is given as follows:

$$\kappa(x_i, x_j) = \exp\left(-\frac{1}{2\sigma^2} \|x_i - x_j\|^2\right) \quad (3)$$

where, σ is a Gaussian function parameter. Because of its capacity to represent intricate and non-linear decision

boundaries, this RBF is employed as the kernel function. The RBF kernel calculates the relationship between two data points' similarity and the Euclidean distance between them [37].

Let's say sample x_i lies outside of the classification hyperplane, ξ_i is greater than 0, and the Karush-Kuhn-Tucker (KKT) condition states that $\alpha_i = 1/\nu l$. This indicates that in traditional OCSVM, samples outside of the surface are defined by SVs. It is well known that the normal vector $w = \sum_{i=1}^l \alpha_i \phi(x_i)$ can be written in terms of the mappings of SVs. When ν is set large to create outliers that are outside the surface, these outliers become SVs and their respective α_i are equal to $1/\nu l$, which gives them more magnitude control over the normal vector and the surface; conversely, when ν is set small, the surface surrounds the outliers, thereby influencing the surface.

There are two components to the strategy suggested by Schölkopf et al. [36]; getting the surface around the target cluster core is the first step; using this surface and find outliers and remove them from the training set so that the final OCSVM model may be trained. It is anticipated that the surface will identify the outliers outside and, more crucially, will not be affected by these outside-located outliers to better contain the cluster core of the target class. By adjusting ν , the first expectation can be fulfilled since this value represents the upper bound for training samples that are situated outside of the surface or hyperplane.

3.2.1 Algorithmic improvements to the OCSVM

Anomaly detection makes considerable use of a OCSVM. To detect anomalies, OCSVM looks for the ideal hyperplane in high-dimensional data that can maximally separate the data from anomalies. However, conventional OCSVM's hinge loss is unlimited, leading to greater loss from outliers and impairing their ability to detect anomalies. Outlier influence is decreased using the bounding hinge loss function [38].

In this section, the most recent development in the fundamental research on OCSVM at the algorithmic level is discussed. The implementation of the OCSVM in the earlier research is always centered around the original formulation, such as in Eqs. (1) and (2). However, recently, more research found in the literature has made significant changes to the OCSVM formulation at the algorithmic level.

For example, a new OCSVM algorithm based on hidden information was proposed by Zhu and Zhong [39]. They proposed a modified OCSVM that exploits so-called "hidden" or group information present in the training data. Unlike the standard OCSVM, which models all training samples using a single set of slack variables, their approach assumes that samples can be meaningfully partitioned into groups (e.g., based on demographic, acquisition, or contextual attributes). The key motivation is that deviations from the learned boundary may have different significance across groups, and a single global slack mechanism may bias the novelty boundary, particularly when training data are limited or heterogeneous. To address this, the method introduces a second learning space, referred to as the correcting space, in addition to the conventional decision space. While all samples contribute to defining the decision boundary in the decision space, group-specific correcting functions are learned in the correcting space to model slack variables. These correcting functions regulate how much each group is allowed to violate the boundary, effectively imposing group-dependent constraints on the slack variables. Importantly, the correcting

space does not perform classification; instead, it modulates the influence of training errors during optimization. This two-space formulation enables the OCSVM to incorporate auxiliary group information, improving generalization by preventing dominant or noisy groups from disproportionately shaping the novelty boundary.

In another approach, an algorithm for resilient and sparse anomaly detection was put forth by Tian et al. [40]. They have added the ramp loss function with comparison to the initial implementation of OCSVM. The primary goal of this work is to create a sparse and robust semi-supervised method by using the non-convexity aspects of the ramp loss function. Additionally, the derived model was solved using the Concave-Convex Procedure (CCCP), which solves a non-differentiable non-convex optimization problem. The authors have introduced a ramp loss function in this manner:

$$R_{\rho,s} = \begin{cases} 0, & z \geq \rho \\ \rho - z, & \rho - s < z < \rho \\ s, & z \leq \rho - s \end{cases} \quad (4)$$

where, z and s are the loss function score and pre-defined parameter, respectively. Therefore, the ramp loss function as in Eq. (4) was employed to strengthen the resilience of OCSVM and prevent the outliers from being excluded as support vectors. Based on this function, for those $z \leq \rho - s$, the ramp loss will be flat, and its value will be a constant s . Ramp loss OCSVM can be reformulated as:

$$\min_{w,\rho,s} \frac{1}{2} \|w\|_2^2 + \frac{1}{\nu l} \sum_{i=1}^l R_{\rho,s}(w^T \phi(x_i)) \quad (5)$$

To accurately categorize samples that resemble training data, OCSVM seeks to choose an appropriate region from an unknown probability distribution that includes the majority of the input samples. This algorithm can be used to locate the hypersphere with the smallest radius to discover outliers. However, a conventional OCSVM's hinge loss is unlimited, leading to greater loss from outliers and impairing their ability to detect anomalies, which in turn results in the lessened outlier's influence [40].

On the other hand, a new method for detecting anomalies in big amounts of data using randomized nonlinear features in SVM was proposed [38]. This method reduces computing complexity by doing away with the requirement to handle enormous kernel matrices for large datasets. Instead of looking for support vectors that are optimized using an unlimited loss function, the authors have proposed an iterative approach with a bounded loss function using a half-quadratic optimization technique.

In this case, the random projection can reduce the optimization method's computational cost needed for large nonlinear kernels. Hence, a similar effect can be observed when using nonlinear features. Considering these issues, the authors in the research [38] proposed OCSVM's optimization issue with a bounded loss function as:

$$\max_{w,\rho} \frac{1}{2} \|w\|_2^2 - \frac{1}{\nu l} \sum_{i=1}^l \xi_i^r - \rho \quad (6)$$

where, randomized slack variables, $\xi_i^r = \beta(1 - e^{-\eta \xi_i})$, with normalization constant, β given as follows:

$$\beta = \frac{1}{1 - e^{-\eta}} \quad (7)$$

where, the scaling constant is $\eta \geq 0$. The fundamental goal of constraining the loss function of $\xi_i^r = 1$ is ensured by the normalizing constant β . The scale constant, η , sets the upper bound, and the loss function becomes conventional hinge loss when $\eta = 0$. The above formula demonstrates that the bounded loss function OCSVM is convex and monotonic, just like conventional OCSVM. It is evident that η regulates the limit of the loss function, substituting the unbounded loss function of the conventional SVM with a bounded one. Greater η values indicate higher degrees of scaling, and vice versa [38].

In terms of kernel parameters, for example σ in Eq. (3), whose selection is not simple for anomaly detection situations, has a significant role in OCSVM's performance. Furthermore, in certain complex and unequal data distributions, the densities and shapes of distinct data regions can differ significantly, which makes it challenging for OCSVM to use a global kernel parameter to generate good boundaries in every region. Hence, in the study by Pang et al. [41], they have suggested a hybrid technique combining vector quantization and OCSVM, known as VQ-OCSVM, to address the problems.

To be more precise, distribution information about normal data is extracted via vector quantization, and the resulting information is utilized to build a clear mapping function that creates a high-dimensional feature space from data. The classifier is then constructed within the feature space using OCSVM. Through the addition of the explicit mapping to OCSVM, two regularization hyperparameters exist, λ_1 and λ_2 have been introduced in VQ-OCSVM. For λ_2 , the traditional OCSVM algorithm is referred to, and it is suggested to be fixed to the value of 1. As for λ_1 , they have introduced the subsequent practical yet empirical approach to determine its value.

In order to achieve feature auto-selection and model sparsity, keep in mind that λ_1 regulates the weight of the L_1 -norm regularization in the objective function. It makes sense that a model with a λ_1 that is too small will overfit the typical training data and be too complex, whereas a model with a λ_1 that is too large will be too simple and underfit the typical training data [41].

The dual problem, which is a quadratic programming problem, is solved using the conventional OCSVM in order to solve the primal problem. Nevertheless, quadratic programming is ineffective for training large-scale problems since its computation is cubic and its storage complexity increases quadratically with problem scale. Therefore, Zhu et al. [42] proposed to directly train OCSVM in primordial space. Unfortunately, the gradient-based optimization approach, a first-order method that converges quickly, cannot solve it because of the non-differentiability of hinge loss utilized in OCSVM.

Furthermore, the OCSVM is less resilient to outliers because of the unbounded hinge loss. Because of the outliers, the decision boundary will diverge greatly from the ideal hyperplane. A nonconvex differentiable function, a Huberized truncated loss function, was suggested as a solution to the problems to increase the OCSVM's resilience [42]. As an alternative to hinge loss in conventional OCSVM, the Huberized truncated loss function was proposed to be used due to its insensitivity to outliers. Unlike regular OCSVM, robust OCSVM has a differentiable primal objective function.

The primordial space's resilient OCSVM can be solved using a method for rapid proximal gradients. The following are the two Huber loss functions that the authors have suggested:

$$H_1(r) = \begin{cases} \rho - r - \frac{\delta}{2}, & r \leq \rho - \delta \\ \frac{(\rho-r)^2}{2\delta}, & \rho - \delta < r \leq \rho \\ 0, & r > \rho \end{cases} \quad (8)$$

$$\kappa H_s(r) = \begin{cases} s - r - \frac{\delta}{2}, & r \leq \rho - \delta \\ \frac{(s-r)^2}{2\delta}, & s - \delta < r \leq s \\ 0, & r > s \end{cases} \quad (9)$$

The continuous and differentiable nature of $H_1(r)$ and $H_s(r)$ was readily apparent. It can be observed that these functions diminish to the hinge loss when $\delta \rightarrow 0$ and expand linearly as δ lowers. Following the combination of $H_1(r)$ and $-H_s(r)$, the Huberized truncated loss function, $T_s(r) = H_1(r) - H_s(r)$ can be written as follows:

$$T_s(r) = H_1(r) - H_s(r) = \begin{cases} \rho - s, & r \leq s - \delta \\ \rho - r - \frac{\delta}{2} - \frac{(s-r)^2}{2\delta}, & s - \delta < r \leq s \\ \rho - r - \frac{\delta}{2}, & s < r \leq \rho - \delta \\ \frac{(\rho-r)^2}{2\delta}, & \rho - \delta < r \leq \rho \\ 0, & r > \rho \end{cases} \quad (10)$$

As presented above, the development of OCSVM began with a modification that added further restrictions to the slack variables for the instances belonging to various groups. Then, later, the ramp loss function was introduced, where this loss function's non-convexity properties were used to create a robust, sparse semi-supervised method. After that bounded loss function was integrated utilizing OCSVM's randomized nonlinear characteristics, removing the requirement to handle big kernel matrices for big datasets, which lowers the complexity in terms of time and space. In another method, it was developed by utilizing the vector quantization technique combined with the OCSVM, which is more capable of generalization complex data distribution. Later, a Huberized truncated loss function was proposed to overcome the hinge loss function's drawbacks and increase the OCSVM's robustness.

(1) Synthesis of algorithmic advancements in OCSVM

Algorithmic improvements to OCSVM can be broadly categorized into three dominant directions: loss-function modification, integration with auxiliary learning models, and computational acceleration. Loss-based approaches, such as ramp, bounded, and Huberized loss functions, consistently improve robustness against outliers by limiting the influence of extreme samples; however, they often introduce non-convex optimization challenges and heightened parameter sensitivity. Hybrid approaches that combine OCSVM with deep learning or vector quantization enhance representational power but increase model complexity and training cost. Efforts aimed at accelerating OCSVM through randomized features or primal-space optimization improve scalability for large datasets, yet may trade off accuracy in highly heterogeneous data distributions. Overall, existing methods exhibit a

recurring tension between robustness, computational efficiency, and generalization capability.

3.2.2 Applications of OCSVM

There are many areas or fields of applications that have made use of OCSVM in dealing with the challenge of novelty detection. In this section, some of the recent research that has applied OCSVM is discussed. For example, Delgado-Prieto et al. [20] proposed a multi-modal strategy to enhance the effectiveness of novelty detection. The approach used in this work is broken down into three primary steps. The first step is a specific feature calculation and reduction over all accessible physical magnitudes. Next, a group of OCSVM-based novelty detection models was developed to detect previously unconsidered events. Lastly, a diagnosis model that consists of a feature fusion scheme was introduced to classify faults accurately.

A new scheme was proposed to be utilized for denoising, which displays the optimal thresholding rule and mother wavelet combination for every signal [43]. This scheme requires features to be collected from the time-frequency and time-domains, and the energy-to-frequency analysis shows the optimal mother wavelet for feature extraction from each signal. The criterion of the Shannon entropy ratio. In this work, it is also demonstrated because by making the signals more nonlinear, the statistical traditional or mixtures of statistical traditional and solely nonlinear features have the ability to fully classify data; nonlinear features are unable to do so. Hence, three rotating systems' vibration data are detected using OC-SVM; nevertheless, the focus is on data pre-processing techniques such denoising, dimension reduction, vectorization, feature extraction, and normalization. In particular, for the first time, the impact of both classical and nonlinear statistical feature extraction on novelty detection.

Saari et al. [44] have used OCSVM to use fault-specific characteristics taken from vibration signals in order to detect and identify wind turbine bearing problems. By training models using these attributes as input for a OCSVM, automatic identification was accomplished. By adjusting the model tuning settings, detection models with varying sensitivity were trained concurrently. Additionally, efforts were made to determine a process for choosing the model tuning parameters by first determining the system's criticality and then using that information to estimate the detection model's accuracy.

An unsupervised deep learning method was proposed founded on an OCSVM-based deep auto-encoder, where the classifier was used in the measurement of the reaction data obtained from baseline or intact structures as training data, making it possible to identify structural degradation in the future [45]. The well-crafted deep auto-encoder, which is utilized as an extractor to extract damage-sensitive features from the observed acceleration response data, and the OCSVM, which is employed as a damage detector, are the main innovations and contributions of the suggested approach.

On the other hand, Cardoso and Poppi [46] suggested a novel technique that uses one-class modeling and Raman spectroscopy to examine the adulteration in cassava starch. OCSVM was found to have achieved higher accuracy than other techniques. This study evaluated two methods for one-class classifier models: OC-SVM and soft independent modelling by class analogy (SIMCA).

A new hybrid model was proposed in the previous study [47] that combines an unsupervised deep belief network (DBN) and

OCSVM to solve the difficulty of detecting high-dimensional geochemical abnormalities. This model uses DBN to initially extract the pertinent features, which are then fed into the OCSVM. The hybrid method's decision function values are used to map the geochemical patterns associated with iron deposits. The OCSVM is then trained using the features discovered from the original data in order to successfully distinguish the multivariate geochemical anomaly from the geochemical background. Ramp OC-SVM, a robust and nonconvex semi-supervised technique, was proposed by Tian et al. [48]. OCSVM was used to handle the lack of labelled data for deceptive opinions and by utilizing the Ramp loss functions, non-convex features to its advantage, where it removes the impact of non-reviewed opinions and outliers.

In the other work, the large building land area was able to be retrieved from Landsat image data using an OCSVM-based approach [49]. Their works have demonstrated that OCSVM is appropriate for the research of remote sensing image classification when only one class of features is extracted, and expanding the use of OCSVM in remote sensing picture building recognition, demonstrating its effectiveness and precision in this context. In the study by Yin and Wang [50], a new strategy was put forward to detect the outliers of unhealthy leaf images from the massive leaf images dataset using OCSVM. They have established that their modified OCSVM has produced very efficient and robust results in detecting the anomalous data from the extensive dataset of leaf images.

The decision tree classifier and OCSVM were used to create a hybrid intrusion detection system [51]. This system is designed to detect both known and highly accurate zero-day cyberattacks with low false alarm rates. Using OCSVM as the classifier inside their ensemble learning model, the experimental results have shown that this system outperforms traditional single classifier approaches and proves to be more effective than other machine learning techniques.

On the other hand, a novel personalized federated learning method based on OCSVM was proposed by Anaissi et al. [52]. In this work, the method aims to address the transmission and data privacy concerns raised by central machine learning models by customizing the support vectors that are generated at each client in the distributed learning model structure. Their results have shown that the proposed OCSVM-based personalized federated learning method has achieved significant accuracy and can precisely generalize the clients' models compared to the other techniques.

Recently, an OCSVM-based algorithm was proposed by Karami and Niaki [53] to overcome the limited sensitivity of fast change point detection, high computing costs, poor scalability with large networks, and excessive reliance on case-based features. It is adaptable to a variety of social network applications and efficiently identifies network disruptions by utilizing both nodal and network-level characteristics. The approach reduces the processing of input data by using a well-defined training data dictionary with an evolutionary network update procedure, improving memory and time efficiency.

Recent application studies also demonstrate the practicality of one-class classification for structural health monitoring using online anomaly detection, emphasizing real-time deployment under limited anomaly labels and evolving operating conditions [54].

(1) Summary of application-oriented implementation

Across diverse application domains, including fault diagnosis, medical analysis, remote sensing, and cybersecurity, OCSVM-based novelty detection demonstrates strong adaptability and consistent performance when labeled anomaly data are scarce. Application-driven studies highlight the importance of domain-specific feature engineering and preprocessing pipelines, which often play a decisive role in detection accuracy. Nevertheless, reliance on handcrafted features or expert knowledge reduces portability across domains. Furthermore, many application-oriented models require careful parameter tuning and exhibit sensitivity to operating conditions, limiting robustness under data drift or evolving environments. These observations suggest that future application-focused research should emphasize adaptive feature learning and automated parameter calibration to improve long-term reliability.

Recent comparative studies continue to include OC-SVM as a primary baseline for unsupervised anomaly detection. For example, Agyemang [55] evaluated OC-SVM against Isolation Forest, Local Outlier Factor, Robust Covariance, and an SGD-based OC-SVM variant, reporting that OC-SVM remains among the most effective methods for identifying outliers, achieving strong recall performance, while highlighting trade-offs with precision depending on the dataset characteristics.

4. SUMMARY

Table 1 summarizes the review of the OCSVM and SVDD algorithms in the context of novelty detection. The works are arranged in this table from the most recent implementation to the oldest research. Consistent improvements of fundamental algorithms like OCSVM and SVDD are revealed in this review of novelty detection techniques. Using methods like ramp loss functions, deep belief networks (DBNs), ensemble classifiers, and metaheuristic optimizations like Grasshopper or Sparrow Search, researchers have enhanced these algorithms. These integrations are intended to increase the accuracy of fault detection, strengthen resilience to outliers, and support complex or high-dimensional datasets from spam filtering and plant health monitoring to electromechanical faults and geochemical analyses. Dynamic adaptations like quarter-sphere methods and variable-radius SVDD further highlight the fields' emphasis on efficiency and generalization across a range of applications.

Even though there has been a lot of progress, there are still issues mostly with computational cost scalability and parameter sensitivity. Many models rely on fine-tuning kernels, loss parameters, or feature extraction architectures, which can make them difficult to use in large-scale or real-time scenarios. Additionally, some methods are less adaptable to different environments because they mainly rely on domain-specific preprocessing pipelines or signature databases. Furthermore, techniques such as empirical pinball losses or fixed threshold autoencoders show vulnerability when dealing with noisy inputs or changing data distributions. When considered collectively, the developments show promise but also highlight the need to strike a careful balance in novelty detection between innovation and generalizability.

Table 1. Summary of different approaches used in novelty detection

Methods	Improvements	Limitations
DW-SVDD (SVDD and k -NN) [22]	Enhances SVDD by incorporating density weights from k -NN, prioritizing high-density regions, shifting boundary toward dense clusters, and improving accuracy without assuming data distribution.	May underperform with symmetric or multi-modal data where density weights exert less influence, resulting in less accurate descriptions and potential misclassification.
Ramp-OCSVM [40]	Replaces hinge loss with ramp loss to improve robustness against outliers; solves non-convex optimization via CCCP.	Non-convex and non-differentiable optimization requires iterative CCCP, increasing computational cost and limiting scalability on large datasets.
OCSVM [20]	Multimodal novelty detection methods are used to ensemble OCSVMs and separate feature reduction, enhance detection accuracy, fault diagnosis reliability, and adaptability to unknown electromechanical faults.	Risks diagnosis errors without novelty detection, struggles with overlapping fault features, and relies on separate feature reductions, which may introduce complexity and affect generalizability across varied electromechanical systems.
OCSVM [43]	Enhanced OC-SVM by using nonlinear features and a systematic preprocessing pipeline, which includes advanced denoising and optimized wavelet selection.	Despite high accuracy, the algorithm depends on tailored preprocessing for each signal, increasing computational burden and limiting scalability; traditional statistical features may fail under rising nonlinearity.
OCSVM [44]	The algorithm enhances fault detection by combining fault-specific vibration features with parallel OCSVM models tuned for system criticality, enabling earlier detection than traditional methods with minimal false alarms.	The algorithm fails to reliably identify fault locations without auxiliary methods, is sensitive to feature and parameter selection, and cannot always distinguish between similar fault signatures from different components.
OCSVM and C5 decision tree classifier [51]	Hybrid algorithm combines decision tree and OCSVM using a stacking ensemble, enabling accurate detection of both known and zero-day intrusions with reduced false alarm rates.	Hybrid system adds complexity, depends heavily on signature database updates and kernel selection, and may struggle to adapt quickly to evolving malware behaviors or large-scale data.
Deep Belief Networks (DBN) and OCSVM [47]	The hybrid model boosts anomaly detection by extracting nonlinear, high-level geochemical features via deep belief networks, improving accuracy and scalability for complex, high-dimensional mineral datasets.	The model relies heavily on optimal parameter tuning, involves increased computational cost, and may struggle with generalization across datasets with diverse geochemical patterns.
OCSVM [38]	The algorithm improves OCSVM by integrating randomized nonlinear features and a bounded loss function, reducing sensitivity to outliers, and significantly decreasing computational complexity for large-scale IoT anomaly detection.	The performance depends heavily on parameter tuning, may struggle with highly imbalanced or noisy IoT data, and randomized features could dilute the discriminative power.
Ramp-OCSVM [48]	The algorithm enhances spam detection by combining semi-supervised learning with ramp loss, improving robustness to outliers and non-review noise while maintaining strong generalization with limited labeled deceptive data.	Requires careful ramp parameter tuning, involves non-convex optimization, and may struggle with evolving linguistic styles or sparse labelled deceptive data in large-scale opinion datasets.
OCSVM [46]	The algorithm integrates Raman spectroscopy with OCSVM for rapid, non-destructive cassava starch adulteration detection, achieving higher accuracy and detecting adulterants at low concentrations.	The detection of lower adulteration concentration remains challenging due to Raman sensitivity and sample heterogeneity. OCSVM also depends on careful parameter tuning and effective preprocessing.
SVM, SVDD, and Grasshopper Optimization Algorithm (GOA) [27]	The algorithm introduces a two-stage fault diagnosis framework using GOA-optimized SVDD and SVM, with entropy-based feature selection tailored to normal data, boosting early detection and classification accuracy.	The algorithm depends heavily on entropy indicator selection, requires careful parameter tuning via GOA, and may be less effective with severely imbalanced or noisy datasets.
SVDD [29]	The improvement generalizes SVDD by replacing the fixed linear slack penalty with a tuneable ℓ_p -norm, allowing nonlinear error weighting, better sparsity control, and improved generalization performance in one-class classification tasks.	The selection of optimal p and kernel parameters is challenging, and the performance may degrade with poor tuning, and computational cost increases with complex optimization in high-dimensional data.
DR-SVDD [30]	The method improves fault diagnosis by combining hypersphere radius with local data distribution, enabling dynamic radius adjustment and multi-class fault identification with higher accuracy in the systems.	Requires careful tuning of radius weight and neighbour count, may struggle with noisy data, and adds complexity with dynamic boundary calculations per test sample.
DBN, SVDD, and Sparrow Search Algorithm (SSA) [31]	The algorithm uses a DBN for automatic feature extraction, optimized SVDD via SSA for accurate degradation modelling, eliminating reliance on fault data and manual parameter tuning.	Limited testing beyond experimental datasets may cause potential challenges in real-world variability, like changing speed and load. Only the amplitude spectrum was used while the other inputs were not explored, and the anti-noise capability needs deeper analysis.

Methods	Improvements	Limitations
SVDD and Part Interval Stacked Autoencoder (PISAE) [32]	The model improves SAE loss function to handle partial uncertainty, extracts robust nonlinear features via PISAE, and combines with SVDD for accurate process monitoring without misclassifying normal fluctuations as faults.	The method relies on predefined uncertainty ranges; lacks dynamic estimation of uncertainty levels across variables, and may underperform with highly imbalanced or noisy data due to fixed thresholding in error processing.
Broad Learning System (BLS) and SVDD [33]	Incorporates BLS-based reconstruction error as sample weighting, introduces a few anomalous samples during SVDD training, enhancing robustness, generalization, and domain boundary accuracy in highly imbalanced datasets.	The model depends heavily on reconstruction quality, and it may underperform if reconstruction errors do not reflect anomaly likelihood accurately, resulting in limited effectiveness in high-noise data or datasets with densely overlapping normal-abnormal distributions.
pin-SVDD [34]	The algorithm introduces pinball loss to SVDD, making all training data influential in sphere formation, enhancing robustness to noise and resampling, while preserving computational efficiency without extra preprocessing.	Lacks dynamic adaptability to evolving data distributions, such as assuming a static penalty structure. The pinball parameter tuning is empirical and may risk excluding valid targets if set improperly.
OCSVM [49]	The method applies OCSVM to extract construction land from multi-temporal satellite imagery data with high accuracy, enabling efficient monitoring, expansion analysis, and driving factor evaluation in urban environments.	Only construction land is classified, ignoring other land types, and therefore lacks granularity in mixed-use areas. It requires careful sample selection, and parameter tuning may impact accuracy across time or terrain variations.
OCSVM [50]	A modified OCSVM was proposed to detect leaf image anomalies, integrates expert-scored labels, extracts feature via neural networks, and assigns outlier scores for fine-grained plant health evaluation.	The algorithm relies on expert-labelled data, limiting scalability; dataset diversity and fuzzy labels affect accuracy; feature extraction and scoring are dataset-dependent, potentially reducing generalization to unseen leaf types.
DBN, Improved Quarter-Sphere Based OCSVM (IQSSVM) [56]	The model combines DBN with IQSSVM for real-time anomaly detection; reduces complexity via sorting, removes kernel parameter dependency, and supports accurate unsupervised processing of high-dimensional sequential data.	The model requires careful DBN tuning for feature extraction, and hence, performance may vary with DBN architecture. It assumes anomaly distribution characteristics in which the accuracy may drop if anomalies are not one-sided or if data noise increases.

5. CONCLUSION

This review examined novelty detection from the perspective of domain-description learning, focusing on SVM-based formulations, particularly SVDD and OCSVM. Unlike probabilistic, distance-based, and reconstruction-based methods, domain-description approaches directly model the boundary of normal data, which explains their broad adoption in scenarios where abnormal samples are rare, unknown, or difficult to label.

The reviewed literature shows that recent progress in SVDD/OCSVM development follows several consistent directions. First, robust loss designs (e.g., ramp, bounded, pinball, or Huberized losses) reduce the influence of outliers and contaminated training data; however, they often introduce non-convex optimization or additional hyperparameters, increasing training complexity and sensitivity. Second, hybrid and deep integration strategies enhance representation learning and improve detection accuracy in high-dimensional domains, but they commonly require extensive tuning and can increase computational cost, limiting practicality in real-time applications. Third, acceleration and scalability techniques, including randomized features and primal-space optimization, reduce computational and memory burdens for large datasets; nonetheless, these improvements may trade off boundary precision in heterogeneous or dynamically evolving environments.

Despite these developments a number of unresolved issues are still present in all of the studies that were surveyed. The sensitivity of OCSVM/SVDD performance to kernel parameters loss settings and threshold selection is a major problem that frequently necessitates expert-driven tuning and decreases portability across domains. The decreased dependability of static decision boundaries under concept drift

data drift and non-stationary circumstances particularly in streaming or long-term monitoring settings is another persistent drawback. Lastly while many studies show good empirical performance, they offer few theoretical assurances about generalization and robustness under contamination and distribution shift.

Future research should therefore prioritize: (i) self-tuning and adaptive domain-description models that can update boundaries under drift while maintaining stability, (ii) scalable optimization frameworks that preserve robustness without high computational overhead, and (iii) hybrid methods that combine domain-description learning with modern feature learning while minimizing parameter sensitivity. Addressing these directions will significantly improve the deployability of OCSVM/SVDD-based novelty detection in real-world systems.

REFERENCES

- [1] Li, S., Tung, W.L., Ng, W.K. (2014). A novelty detection machine and its application to bank failure prediction. *Neurocomputing*, 130: 63-72. <https://doi.org/10.1016/j.neucom.2013.02.043>
- [2] Wu, H., Prasad, S., Priya, T. (2014). Detecting new classes via infinite warped mixture models for hyperspectral image analysis. In 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, pp. 5027-5031. <https://doi.org/10.1109/icip.2014.7026018>
- [3] Duong, P., Nguyen, V., Dinh, M., Le, T., Tran, D., Ma, W. (2015). Graph-based semi-supervised Support Vector Data Description for novelty detection. In 2015 International Joint Conference on Neural Networks

(IJCNN), pp. 1-6. <https://doi.org/10.1109/IJCNN.2015.7280565>

[4] Clifton, D.A., Clifton, L.A., Bannister, P.R., Tarassenko, L. (2008). Automated novelty detection in industrial systems. In *Advances of Computational Intelligence in Industrial Systems*, pp. 269-296. https://doi.org/10.1007/978-3-540-78297-1_13

[5] Yadav, B., Devi, V.S. (2014). Novelty detection applied to the classification problem using Probabilistic Neural Network. In *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, Orlando, FL, USA, pp. 265-272. <https://doi.org/10.1109/CIDM.2014.7008677>

[6] Alsuwaidi, A., Grieve, B., Yin, H. (2018). Feature-ensemble-based novelty detection for analyzing plant hyperspectral datasets. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(4): 1041-1055. <https://doi.org/10.1109/JSTARS.2017.2788426>

[7] Ding, X., Li, Y., Belatreche, A., Maguire, L.P. (2014). An experimental evaluation of novelty detection methods. *Neurocomputing*, 135: 313-327. <https://doi.org/10.1016/j.neucom.2013.12.002>

[8] Hoffmann, H. (2007). Kernel PCA for novelty detection. *Pattern Recognition*, 40(3): 863-874. <https://doi.org/10.1016/j.patcog.2006.07.009>

[9] ZareMoodi, P., Beigy, H., Siahroudi, S.K. (2015). Novel class detection in data streams using local patterns and neighborhood graph. *Neurocomputing*, 158: 234-245. <https://doi.org/10.1016/j.neucom.2015.01.037>

[10] ZareMoodi, P., Siahroudi, S.K., Beigy, H. (2016). A support vector based approach for classification beyond the learned label space in data streams. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing (SAC '16)*, Pisa, Italy, pp. 910-915. <https://doi.org/10.1145/2851613.2851652>

[11] Farid, D.M., Rahman, C.M. (2012). Novel class detection in concept-drifting data stream mining employing decision tree. In *2012 7th International Conference on Electrical and Computer Engineering*, Dhaka, Bangladesh, pp. 630-633. <https://doi.org/10.1109/ICECE.2012.6471629>

[12] Masud, M.M., Gao, J., Khan, L., Han, J., Thuraisingham, B. (2009). Integrating novel class detection with classification for concept-drifting data streams. In *Machine Learning and Knowledge Discovery in Databases*, pp. 79-94. https://doi.org/10.1007/978-3-642-04174-7_6

[13] Faria, E.R., Gama, J., Carvalho, A.C.P.L.F. (2013). Novelty detection algorithm for data streams multi-class problems. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing (SAC '13)*, Coimbra, Portugal, pp. 795-800. <https://doi.org/10.1145/2480362.2480515>

[14] Masud, M.M., Chen, Q., Khan, L., Aggarwal, C.C., Gao, J., Han, J., Srivastava, A., Oza, N.C. (2013). Classification and adaptive novel class detection of feature-evolving data streams. *IEEE Transactions on Knowledge and Data Engineering*, 25(7): 1484-1497. <https://doi.org/10.1109/TKDE.2012.109>

[15] Pimentel, M.A.F., Clifton, D.A., Clifton, L., Tarassenko, L. (2014). A review of novelty detection. *Signal Processing*, 99: 215-249. <https://doi.org/10.1016/j.sigpro.2013.12.026>

[16] Domingues, R., Michiardi, P., Barlet, J., Filippone, M. (2020). A comparative evaluation of novelty detection algorithms for discrete sequences. *Artificial Intelligence Review*, 53(5): 3787-3812. <https://doi.org/10.1007/s10462-019-09779-4>

[17] Entezami, A., Shariatmadar, H., Mariani, S. (2020). Early damage assessment in large-scale structures by innovative statistical pattern recognition methods based on time series modeling and novelty detection. *Advances in Engineering Software*, 150: 102923. <https://doi.org/10.1016/j.advengsoft.2020.102923>

[18] Tang, J., Tian, Y., Liu, X. (2019). LGND: A new method for multi-class novelty detection. *Neural Computing and Applications*, 31(8): 3339-3355. <https://doi.org/10.1007/s00521-017-3270-7>

[19] Oliveira, M.A., Simas Filho, E.F., Albuquerque, M.C.S., Santos, Y.T.B., da Silva, I.C., Farias, C.T.T. (2020). Ultrasound-based identification of damage in wind turbine blades using novelty detection. *Ultrasonics*, 108: 106166. <https://doi.org/10.1016/j.ultras.2020.106166>

[20] Delgado-Prieto, M., Carino, J.A., Saucedo-Dorantes, J.J., Osornio-Rios, R.A., Romeral, L., Romero Troncoso, R.J. (2018). Novelty detection based condition monitoring scheme applied to electromechanical systems. In *2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*, Turin, Italy, pp. 1213-1216. <https://doi.org/10.1109/ETFA.2018.8502503>

[21] Buczak, A.L., Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys and Tutorials*, 18(2): 1153-1176. <https://doi.org/10.1109/COMST.2015.2494502>

[22] Cha, M., Kim, J.S., Baek, J.G. (2014). Density weighted support vector data description. *Expert Systems with Applications*, 41(7): 3343-3350. <https://doi.org/10.1016/j.eswa.2013.11.025>

[23] Ahsan, M., Khusna, H., Wibawati, Lee, M.H. (2023). Support vector data description with kernel density estimation (SVDD-KDE) control chart for network intrusion monitoring. *Scientific Reports*, 13(1): 19779. <https://doi.org/10.1038/s41598-023-46719-3>

[24] Yin, L., Wang, H., Fan, W. (2018). Active learning based support vector data description method for robust novelty detection. *Knowledge-Based Systems*, 153: 40-52. <https://doi.org/10.1016/j.knosys.2018.04.020>

[25] Kim, S., Choi, Y., Lee, M. (2015). Deep learning with support vector data description. *Neurocomputing*, 165: 111-117. <https://doi.org/10.1016/j.neucom.2014.09.086>

[26] Wang, Z., Cha, Y.J. (2020). Unsupervised deep learning approach using a deep auto-encoder with a one-class support vector machine to detect damage. *Structural Health Monitoring*, 20(1): 406-425. <https://doi.org/10.1177/1475921720934051>

[27] Zhang, J., Zhang, Q., Qin, X., Sun, Y. (2022). A two-stage fault diagnosis methodology for rotating machinery combining optimized support vector data description and optimized support vector machine. *Measurement*, 200: 111651. <https://doi.org/10.1016/j.measurement.2022.111651>

[28] Wang, J., Liu, P., Lu, S., Zhou, M., Chen, X. (2023). Decentralized plant-wide monitoring based on mutual information-Louvain decomposition and support vector

data description diagnosis. *ISA Transactions*, 133: 42-52. <https://doi.org/10.1016/j.isatra.2022.07.017>

[29] Rahimzadeh Arashloo, S. (2022). ℓ_p -Norm support vector data description. *Pattern Recognition*, 132: 108930. <https://doi.org/10.1016/j.patcog.2022.108930>

[30] Lu, J., Gao, Y., Zhang, L., Deng, H., Cao, J., Bai, J. (2022). A novel dynamic radius support vector data description based fault diagnosis method for proton exchange membrane fuel cell systems. *International Journal of Hydrogen Energy*, 47(84): 35825-35837. <https://doi.org/10.1016/j.ijhydene.2022.08.145>

[31] Pan, Y., Cheng, D., Wei, T., Jia, Y. (2022). Rolling bearing performance degradation assessment based on deep belief network and improved support vector data description. *Mechanical Systems and Signal Processing*, 181: 109458. <https://doi.org/10.1016/j.ymssp.2022.109458>

[32] Wu, Q., Lu, W., Yan, X. (2022). Process monitoring of nonlinear uncertain systems based on Part Interval Stacked Autoencoder and Support Vector Data Description. *Applied Soft Computing*, 129: 109570. <https://doi.org/10.1016/j.asoc.2022.109570>

[33] Huang, Q., Zheng, Z., Zhu, W., Fang, X., Fang, R., Sun, W. (2022). Anomaly detection algorithm based on broad learning system and support vector domain description. *Mathematics*, 10(18): 3292. <https://doi.org/10.3390/math10183292>

[34] Zhong, G., Xiao, Y., Liu, B., Zhao, L., Kong, X. (2022). Pinball loss support vector data description for outlier detection. *Applied Intelligence*, 52(14): 16940-16961. <https://doi.org/10.1007/s10489-022-03237-5>

[35] Huang, W., Li, Y.J., Xu, Z.N., Yao, X.W., Wan, R.C. (2025). Improved deep support vector data description model using feature patching for industrial anomaly detection. *Sensors*, 25(1): 67. <https://doi.org/10.3390/s25010067>

[36] Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7): 1443-1471. <https://doi.org/10.1162/089976601750264965>

[37] Panza, M.A., Pota, M., Esposito, M. (2023). Anomaly detection methods for industrial applications: A comparative study. *Electronics*, 12(18): 3971. <https://doi.org/10.3390/electronics12183971>

[38] Razzak, I., Zafar, K., Imran, M., Xu, G. (2020). Randomized nonlinear one-class support vector machines with bounded loss function to detect of outliers for large scale IoT data. *Future Generation Computer Systems*, 112: 715-723. <https://doi.org/10.1016/j.future.2020.05.045>

[39] Zhu, W., Zhong, P. (2014). A new one-class SVM based on hidden information. *Knowledge-Based Systems*, 60: 35-43. <https://doi.org/10.1016/j.knosys.2014.01.002>

[40] Tian, Y., Mirzabagheri, M., Bamakan, S.M.H., Wang, H., Qu, Q. (2018). Ramp loss one-class support vector machine: A robust and effective approach to anomaly detection problems. *Neurocomputing*, 310: 223-235. <https://doi.org/10.1016/j.neucom.2018.05.027>

[41] Pang, J., Pu, X., Li, C. (2022). A hybrid algorithm incorporating vector quantization and one-class support vector machine for industrial anomaly detection. *IEEE Transactions on Industrial Informatics*, 18(12): 8786-8796. <https://doi.org/10.1109/TII.2022.3145834>

[42] Zhu, W., Song, Y., Xiao, Y. (2022). Huberized one-class support vector machine with truncated loss function in the primal space. *Advances in Engineering Software*, 173: 103208. <https://doi.org/10.1016/j.advengsoft.2022.103208>

[43] Sadooghi, M.S., Khadem, S.E. (2018). Improving one class support vector machine novelty detection scheme using nonlinear features. *Pattern Recognition*, 83: 14-33. <https://doi.org/10.1016/j.patcog.2018.05.002>

[44] Saari, J., Strömbergsson, D., Lundberg, J., Thomson, A. (2019). Detection and identification of windmill bearing faults using a one-class support vector machine (SVM). *Measurement*, 137: 287-301. <https://doi.org/10.1016/j.measurement.2019.01.020>

[45] Wang, K., Lan, H. (2020). Robust support vector data description for novelty detection with contaminated data. *Engineering Applications of Artificial Intelligence*, 91: 103554. <https://doi.org/10.1016/j.engappai.2020.103554>

[46] Cardoso, V.G.K., Poppi, R.J. (2021). Cleaner and faster method to detect adulteration in cassava starch using Raman spectroscopy and one-class support vector machine. *Food Control*, 125: 107917. <https://doi.org/10.1016/j.foodcont.2021.107917>

[47] Xiong, Y., Zuo, R. (2020). Recognizing multivariate geochemical anomalies for mineral exploration by combining deep learning and one-class support vector machine. *Computers and Geosciences*, 140: 104484. <https://doi.org/10.1016/j.cageo.2020.104484>

[48] Tian, Y., Mirzabagheri, M., Tirandazi, P., Bamakan, S.M.H. (2020). A non-convex semi-supervised approach to opinion spam detection by ramp-one class SVM. *Information Processing and Management*, 57(6): 102381. <https://doi.org/10.1016/j.ipm.2020.102381>

[49] Nie, J., Dong, Y., Zuo, R. (2022). Construction land information extraction and expansion analysis of Xiaogan City using one-class support vector machine. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 3519-3532. <https://doi.org/10.1109/JSTARS.2022.3170495>

[50] Yin, M., Wang, L. (2022). Outlier detection of leaf images based on one-class support of vector machine. *Journal of Physics: Conference Series*, 2179(1): 012040. <https://doi.org/10.1088/1742-6596/2179/1/012040>

[51] Khraisat, A., Gondal, I., Vamplew, P., Kamruzzaman, J., Alazab, A. (2020). Hybrid intrusion detection system based on the stacking ensemble of C5 decision tree classifier and one class support vector machine. *Electronics*, 9(1): 173. <https://doi.org/10.3390/electronics9010173>

[52] Anaissi, A., Suleiman, B., Alyassine, W. (2022). A personalized federated learning algorithm for one-class support vector machine: An application in anomaly detection. In *Lecture Notes in Computer Science*, pp. 373-379. https://doi.org/10.1007/978-3-031-08760-8_31

[53] Karami, A., Niaki, S.T.A. (2024). An online support vector machine algorithm for dynamic social network monitoring. *Neural Networks*, 171: 497-511. <https://doi.org/10.1016/j.neunet.2023.12.024>

[54] Abdrabo, A. (2024). Application of online anomaly detection using one-class classification to the Z24 Bridge. ETH Zurich. <https://doi.org/10.3929/ethz-b-000712475>

[55] Agyemang, E.F. (2024). Anomaly detection using unsupervised machine learning algorithms: A simulation

study. *Scientific African*, 26: e02386.
<https://doi.org/10.1016/j.sciaf.2024.e02386>

[56] Qiao, Y., Wu, K., Jin, P. (2023). Efficient anomaly detection for high-dimensional sensing data with one-class support vector machine. *IEEE Transactions on Knowledge and Data Engineering*, 35(1): 404-417.
<https://doi.org/10.1109/TKDE.2021.3077046>

NOMENCLATURE

A_P	average precision
F_N	false negative
F_P	false positives
H	Huber-loss function
k	number of neighbor
l	total training samples
R	ramp-loss function
r	group's index
roc	receiver operating characteristic
s	pre-defined parameter
T	Huberized truncated-loss function

T_P	true positives
w	weight vector
x	data point
z	function score

Greek symbols

α	positive Lagrange multiplier
β	normalization constant
δ	Hinge loss
η	scaling constant
ℓ	norm
κ	kernel function
λ	regularization parameter
ν	trade-off parameter
ϕ	transformation function
ρ	offset parameter
Z	correcting space
σ	Gaussian function parameter
τ	training data
ξ	slack variable