Journal of Universal Computer Science, vol. 31, no. 11 (2025), 1222-1247
submitted: 23/8/2024, accepted: 13/4/2025, appeared: 28/9/2025 CC BY 4.0

Fusing Monotonic and One-Class Classification: Elevating SVM with the MC-SVDD Strategy

Ming-Lung Hsu
(Department of Business Administration, College of Management,
National Pingtung University of Science and Technology, Pingtung, Taiwan (R.O.C.)
https://orcid.org/0009-0001-4034-7362, mlhsu@mail.npust.edu.tw)

Yu-Wei Liu
(Department of Industrial and Information Management, College of Management,
National Cheng Kung University, Tainan, Taiwan (R.O.C.)
r36054040@mail.ncku.edu.tw)

Sheng Tun Li
(Department of Industrial and Information Management, College of Management,
Center for Innovative FinTech Business Models, National Cheng Kung University, Tainan,
Taiwan (R.O.C.)
https://orcid.org/0000-0003-1210-3157, stli@mail.ncku.edu.tw)

Abstract: Data mining can be considerably improved with the inclusion of prior domain knowledge; such knowledge reveals complex patterns that might otherwise remain hidden. Among such patterns, monotonic relationships between variables are crucial because of their applicability in real-world contexts. Although considerable growth has occurred in the development of monotonic classification models, many of these models excel in binary or multiclass classification but falter in one-class classification. To address this problem, we developed a monotonicity-constrained support vector domain description (MC-SVDD) model in this study. This model is an innovative evolution of the monotonicity-constrained support vector machine model and is specifically designed for one-class classification with strict adherence to monotonicity constraints. In the developed MC-SVDD model, monotonicity constraints are integrated into the well-established support vector domain description (SVDD) framework. Moreover, methods such as quadratic programming and data visualization are incorporated into the MC-SVDD model. In extensive evaluations, the MC-SVDD model outperformed a conventional SVDD model in prediction performance. This study makes a key contribution to domain-driven data mining.

Keywords: Monotonic classification; One-class classification; Monotonicity-constrained SVM; Support vector data description (SVDD); domain-driven data mining
Categories: I.2.6, I.5
DOI: 10.3897/jucs.135070

1 Introduction

Data mining, which involves extracting valuable patterns and knowledge from large data sets, is a crucial step in knowledge discovery. This process allows hidden patterns and insights to be identified in data. Although traditional data-driven approaches are effective at leveraging organizational data, they often ignore domain knowledge. This omission can lead to discrepancies between the mined knowledge and the expertise of domain-specific professionals, ultimately limiting the practical utility of developed models.

Domain-driven data mining, which involves seamlessly integrating domain knowledge into the data mining process, has attracted considerable research interest. The integration of domain knowledge into the data filtering, data classification, and algorithm adaptation steps of the data mining process enhances the efficiency and generalization capability of this process [Sinha and Zhao, 2008, Cao, 2010, Yu et al., 2010]. Moreover, such integration facilitates the interpretation of mining results and the refining of extracted models, thereby yielding substantial benefits [Eryarsoy et al., 2009].

Monotonic relationships are a crucial form of prior domain knowledge [Kotlowski and Slowinski, 2013]. In these relationships, the output variable varies in one direction with respect to the input variable when all other factors remain constant [Hu et al., 2012b]. The use of such monotonic constraints in real-world applications holds substantial promise. An example of prior domain knowledge in the medical insurance industry is the widespread acknowledgment that healthy applicants should pay lower insurance premiums compared with less healthy applicants. This knowledge indicates the relationship between health and premiums. Similar relationships can be found in a variety of domains, such as economics (where the demand for food increases with population growth) and medicine (where chewing betel nut is associated with an elevated risk of oral cancer).

Leveraging monotonic constraints in data mining offers a multitude of advantages. The use of such constraints shrinks the hypothesis space in the learning process and allows experts to evaluate models using metrics other than accuracy, such as consistency with monotonic constraints. Notably, more reasonable and interpretable information can be extracted by integrating monotonicity constraints into classification techniques [Hu et al., 2012a, Cano et al., 2019].

Monotonic classification can be executed using a diverse range of approaches, including classification trees, rule induction, neural networks, hybridization, instance-based learning, and support vector machine (SVM) techniques—such as monotonicity-constrained SVM (MC-SVM) and regularized monotonic fuzzy support vector machine (RMC-FSVM) [Chen and Li, 2014, Li and Chen, 2015, Cano et al., 2019]. These methods have exhibited high capability in modeling monotonic constraints. However, they are predominantly tailored to binary or multiclass classification problems and are difficult to adapt to one-class classification, which involves unique challenges. One-class classification (OCC) is an intricate problem, primarily because of its theoretical complexities, mainly those related to objective function modification and constraint remodeling. In one-class classification, the primary class (the positive or target class) is well-represented in the training data. By contrast, the other class (the negative or nontarget class) has limited representative samples [Khan and Madden, 2014, Hayashi et al., 2022, Ju et al., 2022, Zhang et al., 2023, Hayashi et al., 2024]. These conditions are typical in outlier detection, novelty detection, and concept learning. In general, one-class classification is more difficult than conventional binary or multiclass classification [Perera et al., 2021].

One-class classification differs fundamentally from unbalanced-class classification in its data assumptions and objectives [Xu, 2017, Dablain et al., 2023]. In unbalanced classification, training data consists of multiple predefined classes, including both majority and minority classes, with the goal of correctly classifying samples despite class imbalance. In contrast, OCC assumes that only data from a single known class (the positive or target class) is available during training, with no explicit representation of other potential classes. The objective is to learn the characteristics of the known class to distinguish it from outliers or anomalies, making it particularly suited for scenarios where data for other classes is unavailable or undefined. These distinctions underscore the unique theoretical and practical challenges of OCC, differentiating it from other classification tasks.

Various one-class SVM-based classifiers with impressive performance have been developed [Mazhelis, 2006], including the one-class SVM [Schölkopf et al., 2001] and support vector domain description (SVDD) classifiers [Tax, 2001]. These classifiers are extensions of the SVM classifier and can perform one-class classification with high accuracy. In particular, the SVDD classifier has attracted considerable academic interest because of its useful characteristics. Studies have demonstrated that the performance of this classifier on a training set is a reliable indicator of its performance on testing data. Moreover, the SVDD classifier is robust against differences in distribution between the training and target sets. This classifier is also highly efficient because of its reliance on a small set of support vectors, which results in minimal computational overhead.

Despite extensive research on domain-driven data mining, most existing studies have focused on incorporating monotonic constraints into binary and multiclass classification. However, the integration of such constraints into OCC remains an unexplored research gap. OCC is widely used in anomaly detection [Hayashi et al., 2022], medical diagnosis [Hayashi et al., 2024], and fraud detection [Peng et al., 2023], where the absence of negative class samples presents a major challenge. Without monotonic constraints, OCC models may fail to align with real-world expert knowledge, potentially leading to inconsistent or less interpretable results.

Motivated by this gap, our study introduces a novel approach that integrates monotonicity constraints into OCC. This integration enhances model interpretability, ensures consistency with domain knowledge, and improves the practical applicability of OCC in real-world scenarios. Specifically, we propose a monotonicity-constrained SVDD (MC-SVDD) model, which embeds prior domain knowledge into the learning process. The objective of this study is to develop a theoretically rigorous and empirically validated MC-SVDD framework that improves classification accuracy while maintaining adherence to monotonic constraints.

To the best of our knowledge, this work is the first to integrate monotonic prior knowledge into one-class data mining. Our approach is built upon SVM-based models, including SVDD and MC-SVM, and extends them to a new domain where monotonicity constraints were previously unexplored. We formally mathematically formulate the one-class classification problem with monotonic constraints and develop a practical algorithm for implementing MC-SVDD. Experimental results demonstrate that the proposed model outperforms conventional SVDD in prediction accuracy while preserving compliance with monotonicity constraints, thereby bridging a critical gap in OCC research.

The rest of this paper is structured as follows. Section 2 presents an in-depth review of the related literature. Section 3 describes the formulation of the MC-SVDD model. Section 4 presents the experimental results. Section 5 clearly delineates the novel contributions of our work. Finally, Section 6 provides a comprehensive discussion on the research results and details the conclusions of this study.

2 Literature Review

This section provides a brief review of monotonic and one-class classification to lay out the theoretical foundation of this study.

2.1 Monotonic Classification

Monotonic classification is a specialized area within data mining and is associated with challenges involving ordinal attributes, with the class attribute exhibiting a discernible increase with specific explanatory attributes [Potharst and Feelders, 2002]. Monotonic classification approaches are either data-based or model-based [Li and Chen, 2015]. In data-based approaches, the objective is to monotonize or relabel the training data set to mitigate the influence of data points that deviate considerably from the desired monotonic relationship. By contrast, model-based approaches involve adherence to monotonicity constraints during modeling, with only monotonic functions being considered in these approaches. Research on these approaches has yielded promising results; particularly regarding the use of prior knowledge to extract more meaningful insights from data (see [Cano et al., 2019]). Several SVM models that use prior knowledge or monotonic settings have been developed [Bartley et al., 2016, Cano et al., 2019], such as MC-SVM, RMC-FSVM, and partially monotone SVM (PM-SVM).

MC-SVM is a credit rating model based on SVMs; in this model, monotonic constraints are derived from the expertise of financial professionals [Chen and Li, 2014]. In RMC-FSVM, monotonicity constraints are formulated as inequalities based on a partial ordering of the training data, with Tikhonov regularization being applied to ensure uniqueness and boundedness [Li and Chen, 2015]. Finally, PM-SVM uses a novel constraint generation technique to efficiently achieve monotonicity [Lee et al., 2006]. Although these monotonic classifiers are highly effective in binary and multiclass classification tasks, they are unsuitable for one-class classification problems. Thus, the present study developed a novel monotonic SVM model that is tailored for one-class classification problems.

2.2 One-Class Classification

SVM is a well-known machine learning model that is based on statistical learning principles [Vapnik, 2013]. Its primary aim is to design an optimal hyperplane that maximizes the separation between negative and positive data sets. However, this model exhibits limitations in one-class classification. The SVDD model, which is based on the SVM model, is a one-class classification model developed by Tax and Duin [Tax and Duin, 1999] that successfully characterizes large data sets by using data description techniques. The aim of SVDD is to identify the optimal hypersphere that encompasses the training set and contains all normal data while excluding abnormal data outside specified tolerances [Ben-Hur et al., 2001, Tax and Duin, 2004, Camastra and Verri, 2005]. This model constructs a decision boundary by using support vectors that effectively delineate the hypersphere's boundary. SVDD achieves a more suitable and precise representation of data through this data transformation process.

SVDD has practical applications in a variety of domains. For instance, Tax and Duin [Tax and Duin, 2002] used SVDD for handwritten digit recognition and concluded that scaling methods, such as domain scaling, variance scaling, and min–max scaling, can reduce the errors of this model for outlier data in most cases. Lee et al. [Lee et al., 2006] proposed a method capable of recognizing individuals in low-resolution images. Park et al. [Park et al., 2007] identified the locations of denoised patterns by deriving a preimage of denoised features. Banerjee et al. [Banerjee et al., 2007] used SVDD to achieve improved performance and a reduced false alarm rate in anomaly detection in hyperspectral images. Jiang et al. [Jiang et al., 2012] applied SVDD to detect mechanical faults and found that this model exhibited superior classification ability and efficiency to conventional neural-network-based models. Liu et al. [Liu et al., 2013] adopted SVDD to detect outliers in uncertain data; they generated a pseudo-training set and constructed a global distinctive classifier.

Wang et al. [Wang et al., 2004] used various one-class classification models, namely the SVDD, Gaussian data description, k-means data description, and principal component analysis data description models, for face classification in human–robot interaction. Their experimentation revealed that the SVDD model outperformed the other models, which underscores its effectiveness in data domain description [Tax and Duin, 1999]. The core objective of SVDD is to identify the minimum spherical volume encapsulating all training data so that an appropriate and precise data description can be provided. This model is widely used across diverse fields.

Although SVDD has advantages in data domain description, this model can be further improved. Zhou et al. [Zhou et al., 2017] incorporated the genetic algorithm (GA) into the SVDD model for aircraft detection. They used the GA to optimize SVDD kernel parameters on the basis of data distribution statistics. Moreover, Yin et al. [Yin et al., 2018] introduced an active-learning-based SVDD model for robust novelty detection; in this model, the influence of noisy data on SVDD is reduced through active learning. However, limited research has incorporated prior domain knowledge, including monotonic relationships, into SVDD. Therefore, the present study developed an MC-SVDD model to address this research gap and enhance the applicability of SVDD.

3 MC-SVDD Model

3.1 Concept of Monotonicity

A common assumption in data mining is that improvements in the evaluation of certain attributes lead to better outcomes in a monotonic manner. In this study, we focused solely on monotonically increasing relationships. Monotonicity can be formally defined as follows.

Definition 1 (partial ordering relation): A partial ordering (denoted by $\preccurlyeq$)) on a set A is a binary relation that is reflexive, antisymmetric, and transitive. These properties are described as follows:

Reflexivity: $a \leq a$ for all $a \in A$

Anti-symmetry: If $a \leq b$ and $b \leq a$ for any $a, b \in A$, then $a = b$

Transitivity: If $a \leq b$ and $b \leq c$ for any $a, b, c \in A$, then $a \leq c$

Definition 2 (comparability): A linear ordering is a partial ordering with comparability, which is defined as follows:

For any $a, b \in A$, either $a \leq b$ or $b \leq a$

Definition 3 (monotonicity): Consider a data set $\Im = \{(x_i, y_i)|i = 1,2,\ldots,N\}$, where input data $x_i \in R^n$ and output data $y_i \in R$. The function $F(x): R^n \to R$ classifies input data into some output variable. We define a partial ordering $\preccurlyeq$ over input space R^n and a linear ordering $\leq$ over space R, with class labels represented as integer values y_i. The classification function is monotonic if it satisfies the following condition:

$$\forall\, x_i \text{ and } x_j, \text{ if } x_i \preccurlyeq x_j \text{ then } F(x_i) \leq F(x_j) \qquad (1)$$

In this study, the partial ordering on input space R^n is defined intuitively as follows: when $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and $\boldsymbol{x}' = (x_1', x_2', \ldots, x_n')$, $\boldsymbol{x} \preccurlyeq \boldsymbol{x}'$ if and only if $x_i \leq x_i^{'}$ for $i = 1,2,\cdots,n$. In a classification problem, a target function is monotonic if experts perceive it as such.

This study adopts a heuristic approach to enhance the monotonicity of an SVDD classifier. To incorporate prior knowledge of monotonicity into a problem, a set of random pairs of virtual examples is denoted as follows:

$$MC = \{(\underset{\sim}{\boldsymbol{x}}_{\boldsymbol{k}}, \tilde{\boldsymbol{x}}_{\boldsymbol{k}})|\text{for all observed } \underset{\sim}{\boldsymbol{x}}_{\boldsymbol{k}} \preccurlyeq \tilde{\boldsymbol{x}}_{\boldsymbol{k}}, k = 1,\ldots M\} \qquad (2)$$

The outcomes predicted by the SVDD classifier should closely adhere to the monotonicity constraint $F(x_i) \leq F(x_j)$ for $k = 1,\ldots,M$.

3.2 Monotonicity Constraints in One-Class Classification

In binary or multiclass classification, the incorporation of monotonicity constraints is relatively straightforward because these constraints often stem from domain-specific knowledge provided by human experts. The expectation is that the predicted classes y and y' should satisfy the condition $y < y'$. To enforce monotonicity, constraints of the form $\boldsymbol{w}^{\boldsymbol{T}}\varphi(\boldsymbol{x}) \leq \boldsymbol{w}^{\boldsymbol{T}}\varphi(\boldsymbol{x}')$ can be added to the model for each pair of input vectors $\boldsymbol{x} \preccurlyeq \boldsymbol{x}'$, where $\varphi(\boldsymbol{x})$ represents the kernel function.

However, in one-class classification, which involves the use of only a single class label, no inherent method exists for comparing data points with class labels. Consequently, the aforementioned constraint cannot be directly applied to one-class classification models. To address this challenge, the class label is transformed into a measure of the distance between the data and the center point of the class.

Specifically, if the distance between x' and the center point of the class is greater than the distance between $\boldsymbol{x}$ and the centroid, then $\boldsymbol{x}$ is considered more likely to belong to the class than is $\boldsymbol{x}'$. Formally, the constraint $(\boldsymbol{x} - \boldsymbol{a})^T(\boldsymbol{x} - \boldsymbol{a}) \leq (\boldsymbol{x}' - \boldsymbol{a})^T(\boldsymbol{x}' - \boldsymbol{a})$ can be introduced into a one-class classification model for each pair of input vectors $(\boldsymbol{x} - \boldsymbol{a}) \preccurlyeq (\boldsymbol{x}' - \boldsymbol{a})$, thereby ensuring the preservation of monotonicity. Here, $\boldsymbol{a}$ represents the center of the data set. This treatment allows the pair of input vectors to be represented in a manner similar to that in multiclass classification, in which $\boldsymbol{x} \preccurlyeq \boldsymbol{x}'$.

3.3 SVDD Model

Consider a training data set $X = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}\}$, where X is a compact subset of R^n. A mapping function ϕ, which is expressed as $\phi: X \to F$ [Schölkopf et al., 2001], is used to map X to a high-dimensional feature space F. The objective is to determine the minimum spherical volume that encompasses data point x_i. The relevant constraint can be expressed as follows:

$$(\boldsymbol{x}_i - \boldsymbol{a})^T(\boldsymbol{x}_i - \boldsymbol{a}) \leq R^2 \qquad (3)$$

The center, radius, and slack variable [Vapnik, 2013] of the minimum spherical volume are denoted by $\boldsymbol{a}$, R, and ξ_i, respectively. To minimize the constraint in Eq. (3), the optimization problem is formulated as follows:

$$\begin{aligned} &\text{Min} \quad F(R, \mathbf{a}, \xi_i) = R^2 + C\sum_i \xi_i \\ &\text{Subject to } (\mathbf{x}_i - \boldsymbol{a})^T(\boldsymbol{x}_i - \boldsymbol{a}) \leq R^2 + \xi_i \\ &\text{where } \ \xi_i \geq 0, \forall i \end{aligned} \tag{4}$$

To solve this problem and find the smallest hypersphere, the Lagrangian is constructed using the following equation:

$$L(R, \boldsymbol{a}; \alpha_i, \gamma_i) = R^2 + C\sum_i \xi_i - \sum_i\{\alpha_i[R^2 + \xi_i - (\boldsymbol{x}_i{}^T\boldsymbol{x}_i - 2\cdot\boldsymbol{a}^T\boldsymbol{x}_i + \boldsymbol{a}^T\boldsymbol{a})]\} - \sum_i(\gamma_i\xi_i) \tag{5}$$

Setting the derivatives of L with respect to R, $\boldsymbol{a}$, and ξ_i to 0 leads to the following constraints:

$$\sum_i \alpha_i = 1\,, \boldsymbol{a} = \sum_i \alpha_i \boldsymbol{x}_i\,, C - \alpha_i - \gamma_i = 0, \forall i \tag{6}$$

Eq. (5) is reformulated, and Eq. (6) is substituted into Eq. (5). Consequently, the objective to maximize L with respect to α_i is obtained. This objective is expressed as follows:

$$\begin{gathered} L = \sum_i \alpha_i\,(\boldsymbol{x}_i{}^T\cdot\boldsymbol{x}_i) - \sum_{i,j}\alpha_i\,\alpha_j(\boldsymbol{x}_i{}^T\cdot\boldsymbol{x}_j) \\ \text{with constraints } 0 \leq \alpha_i \leq C, \sum_i \alpha_i = 1 \end{gathered} \tag{7}$$

To determine whether test point $\boldsymbol{z}$ lies within smallest sphere, the distance from test point $\boldsymbol{z}$ to the sphere's center is calculated. Test point $\boldsymbol{z}$ is accepted if this distance is less than the radius, which indicates that this point satisfies the constraint in Eq. (3). When the sphere's center is expressed in terms of support vectors, an object is classified into the normal class when the following condition holds:

$$(\boldsymbol{z}^T\cdot\boldsymbol{z}) - 2\cdot\sum_i \alpha_i\,(\boldsymbol{z}^T\cdot\boldsymbol{x}_i) + \sum_{i,j}\alpha_i\,\alpha_j(\boldsymbol{x}_i{}^T\cdot\boldsymbol{x}_j) \leq R^2 \tag{8}$$

Because data are often not spherically distributed, Tax and Duin [Tax and Duin, 1999] introduced a proper kernel function $K(\boldsymbol{x_i}, \boldsymbol{x_j})$ to replace all inner products $(\boldsymbol{x_i}^T\cdot\boldsymbol{x_j})$. This function enhances the method's flexibility, and the problem of finding a data domain description can then be defined as follows:

$$L = \sum_i \alpha_i\,K(\boldsymbol{x}_i, \boldsymbol{x}_i) - \sum_{i,j}\alpha_i\,\alpha_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{9}$$

$$\text{subject to } 0 \leq \alpha_i \leq C, \sum_i \alpha_i = 1.$$

The condition expressed in Eq. (8) is modified to obtain the following equation:

$$K(\boldsymbol{z}, \boldsymbol{z}) - 2\sum_i \alpha_i\,K(\boldsymbol{z}, \boldsymbol{x}_i) + \sum_{i,j}\alpha_i\,\alpha_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) \leq R^2 \tag{10}$$

3.4 Derivation of the MC-SVDD Model

The MC-SVDD model is created by integrating monotonicity constraints into the conventional SVDD model and employing a regularization process to ensure a globally optimal solution. The model derivation is presented for nonlinear cases, and the linear model can be easily derived as a special case.

For a set of observed data points $\{\boldsymbol{x_i}|i = 1,2,\ldots,N\}$, the MC-SVDD model can be formulated as follows:

$$\begin{aligned} &\text{Min } F(R, \boldsymbol{a}, \xi_i) = R^2 + C\sum_{i=1}^{N} \xi_i \\ &\text{Subject to } (\boldsymbol{x}_i - \boldsymbol{a})^T(\boldsymbol{x}_i - \boldsymbol{a}) \le R^2 + \xi_i \\ &\qquad (\underset{\sim}{\boldsymbol{x}}_k - \boldsymbol{a})^T(\underset{\sim}{\boldsymbol{x}}_k - \boldsymbol{a}) \le (\widetilde{\boldsymbol{x}}_k - \boldsymbol{a})^T(\widetilde{\boldsymbol{x}}_k - \boldsymbol{a}) \\ &\qquad \text{for the observation } \underset{\sim}{\mathbf{x}}_k \preccurlyeq \widetilde{\boldsymbol{x}}_k,\ k = 1,2,\ldots,M \\ &\qquad \forall i, \xi_i \ge 0 \end{aligned} \tag{11}$$

where R represents the radius of the hypersphere, $\boldsymbol{a}$ is the center of the data, C is a parameter that balances margin maximization and classification errors, and ξ_i denotes the distance between outlier point i and the corresponding margin hyperplane.

The objective function and constraints in the optimization problem expressed in Eq. (11) are nonlinear. Similar to the original SVDD problem, the aforementioned problem can be solved in a dual space by using Lagrangian multipliers. The Lagrangian for the problem expressed in Eq. (11) is given as follows:

$$\begin{aligned} &L(R, \boldsymbol{a}, \xi_i; \alpha_i, \beta_k, \gamma_i) = R^2 + C\sum_i \xi_i \\ &-\sum_i \{\alpha_i[R^2 + \xi_i - (\boldsymbol{x}_i{}^T \cdot \boldsymbol{x}_i - 2 \cdot \boldsymbol{a}^T \cdot \boldsymbol{x}_i + \boldsymbol{a}^T \cdot \boldsymbol{a})]\} \\ &-\sum_k \{\beta_k[(\widetilde{\boldsymbol{x}}_k - \boldsymbol{a})^T(\widetilde{\boldsymbol{x}}_k - \boldsymbol{a}) - (\underset{\sim}{\boldsymbol{x}}_k - \boldsymbol{a})^T(\underset{\sim}{\boldsymbol{x}}_k - \boldsymbol{a})]\} - \sum_i (\gamma_i \xi_i) \end{aligned} \tag{12}$$

where the Lagrange multipliers $\alpha_i, \gamma_i \ge 0$ for $i = 1, \ldots, N$ and $\beta_k \ge 0$ for $k = 1, \ldots, M$. The optimal solution can be found at the saddle point of the Lagrangian by conducting minimization over the primal variables R and $\boldsymbol{a}$ [Witten et al., 2016] and then performing maximization over the dual multipliers $\alpha_i, \gamma_i \ge 0$ for $i = 1, \ldots, N$ and over $\beta_k \ge 0$ for $k = 1, \ldots, M$.

$$\underset{\alpha,\beta,\gamma}{Max}\ \underset{R,\boldsymbol{a},\xi}{Min}\ L(R, \boldsymbol{a}, \xi_i; \alpha, \beta, \gamma) \tag{13}$$

Solving the problem expressed in Eq. (13) involves taking derivatives with respect to R, ξ_i, and $\boldsymbol{a}$ as follows:

$$\begin{cases} \frac{\partial L}{\partial R} = 0 \longrightarrow 2R - 2R\sum_i \alpha_i = 0; \sum_i \alpha_i = 1\ \forall_i \\ \frac{\partial L}{\partial \xi_i} = 0 \longrightarrow C - \alpha_i - \gamma_i = 0\ \forall_i, \alpha_i \ge 0,\ \gamma_i \ge 0 \\ \frac{\partial L}{\partial \boldsymbol{a}} = 0 \longrightarrow \boldsymbol{a} = \sum_i \alpha_i \boldsymbol{x}_i - \sum_k \beta_k(\widetilde{\boldsymbol{x}}_k - \underset{\sim}{\boldsymbol{x}}_k) \\ \qquad i = 1, \ldots, N,\ k = 1, \ldots, M \end{cases}$$

(14)

The problem expressed in Eq. (13) can be transformed into a quadratic programming problem (dual problem) with the following form:

$$\begin{aligned} \text{Max}_{\alpha,\beta} Q(\alpha, \beta) = &\sum_i \alpha_i\, (\boldsymbol{x}_i{}^T \boldsymbol{x}_i) - \sum_k \beta_k[\widetilde{\boldsymbol{x}}_k{}^T \widetilde{\boldsymbol{x}}_k - \underset{\sim}{\boldsymbol{x}}_k{}^T \underset{\sim}{\boldsymbol{x}}_k] \\ &- \frac{1}{2}\sum_{i,j} \alpha_i\, \alpha_j \cdot (2 \cdot \boldsymbol{x}_i{}^T \boldsymbol{x}_j) \end{aligned}$$

$$
\begin{aligned}
& -\frac{1}{2}\sum_{i,k} \alpha_i \beta_k \cdot [-2 \cdot (\tilde{\boldsymbol{x}}_k - \underset{\sim}{\boldsymbol{x}}_k)^T \boldsymbol{x}_i] \\
& -\frac{1}{2}\sum_{k,i} \alpha_i \beta_k \cdot [-2 \cdot \boldsymbol{x}_i{}^T (\tilde{\boldsymbol{x}}_k - \underset{\sim}{\boldsymbol{x}}_k)] \\
& -\frac{1}{2}\sum_{k,l} \beta_k \beta_l \cdot [4 \cdot (\tilde{\boldsymbol{x}}_k - \underset{\sim}{\boldsymbol{x}}_k)^T (\tilde{\boldsymbol{x}}_l - \underset{\sim}{\boldsymbol{x}}_l)] \\
\text{Subject to} \quad & \sum_i \alpha_i = 1\,, \\
& 0 \le \alpha_i \le C, \beta_k \ge 0, \\
& \quad i,j = 1,\ldots,N \;\; k,l = 1,\ldots,M
\end{aligned} \tag{15}
$$

A kernel trick can be applied to this quadratic form. Given a symmetric continuous function $K: R^n \times R^n \to R$ that satisfies Mercer's condition [Mercer, 1909, Reid, 1970, Schölkopf and Smola, 2002], there exists a mapping function $\phi(x)$ such that $K(\boldsymbol{x}, \boldsymbol{x}') = \phi(\boldsymbol{x})\phi(\boldsymbol{x}')$. When an appropriate kernel K is selected, the nonlinear MC-SVDD classifier can be defined as follows:

$$
\begin{aligned}
& f_{MC-SVDD}(\boldsymbol{z}; \boldsymbol{\alpha}, \boldsymbol{\beta}, R) = I(\|\boldsymbol{z} - \boldsymbol{a}\|^2 \le R^2) \\
& = I\big((\boldsymbol{z}^T\boldsymbol{z}) - 2 \cdot [\textstyle\sum_i \alpha_i(\boldsymbol{z}^T\boldsymbol{x}_i) - \sum_k \beta_k \boldsymbol{z}^T(\tilde{\boldsymbol{x}}_k - \underset{\sim}{\boldsymbol{x}}_k)] + \big[\sum_{i,j} \alpha_i\alpha_j \boldsymbol{x}_i{}^T\boldsymbol{x}_j - \\
& \quad 2\textstyle\sum_{i,k} \alpha_i\beta_k \boldsymbol{x}_i{}^T(\tilde{\boldsymbol{x}}_k - \underset{\sim}{\boldsymbol{x}}_k) + \sum_{k,l} \beta_k\beta_l(\tilde{\boldsymbol{x}}_k - \underset{\sim}{\boldsymbol{x}}_k)^T(\tilde{\boldsymbol{x}}_k - \underset{\sim}{\boldsymbol{x}}_k)\big] \le R^2\big)
\end{aligned} \tag{16}
$$

where $\alpha_i, \alpha_j, \beta_k$, and β_l are solutions to the quadratic programming problem expressed in Eq. (15). The indicator function I is defined as follows:

$$
I(A) = \begin{cases} 1, & \text{if } A \text{ is true} \\ 0, & \text{otherwise} \end{cases} \tag{17}
$$

The incorporation of a kernel in a learning model typically expands the function class, thereby enhancing the model's learning capacity. Furthermore, the use of a Mercer kernel can reduce the number of computationally intensive calculations associated with the derivation of the inner products $\phi(\boldsymbol{x})\phi(\boldsymbol{x}')$ during the optimization of the objective function.

Three common choices for kernel K are expressed as follows:

$$
\begin{aligned}
& K(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{x}^T\boldsymbol{x}' \text{ (linear)}, \\
& K(\boldsymbol{x}, \boldsymbol{x}') = (\tau + \boldsymbol{x}^T\boldsymbol{x}') \text{ (Polynomial of degreed)}, \\
& K(\boldsymbol{x}, \boldsymbol{x}') = exp(-\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2/\sigma^2) \text{ (Gaussian or RBF kernel)}
\end{aligned} \tag{18}
$$

Selecting the most suitable kernel for a specific application is an open problem. In classification problems, linear, polynomial, and radial basis function (RBF) kernels are frequently used; in nonlinear function estimation and modeling problems, linear and RBF kernels are commonly employed. Therefore, in this study, linear and RBF kernels are adopted.

3.5 Learning Algorithm for the MC-SVDD Model

This section presents a learning algorithm for the MC-SVDD model expressed in Eq. (15). The objective function $Q(\alpha, \beta)$ is reformulated in a matrix form as follows:

$$
Q(\alpha, \beta) = -\tfrac{1}{2}[\alpha^T \quad \beta^T]\boldsymbol{G}\begin{bmatrix}\alpha \\ \beta\end{bmatrix} + (\mathbf{x}_i{}^T\boldsymbol{x}_i)_1^N\alpha - \big[\tilde{\boldsymbol{x}}_k{}^T\tilde{\boldsymbol{x}}_k - \underset{\sim}{\boldsymbol{x}}_k{}^T\underset{\sim}{\boldsymbol{x}}_k\big]_1^M\beta \tag{19}
$$

The matrix $\boldsymbol{G}$ is defined as follows: $\boldsymbol{G} = \begin{bmatrix} \boldsymbol{G}^{11} & \boldsymbol{G}^{12} \\ \boldsymbol{G}^{21} & \boldsymbol{G}^{22} \end{bmatrix}$. Moreover, the submatrices of this matrix are expressed as follows:

$$\begin{aligned} \boldsymbol{G}_{i,j}^{11} &= 2 \cdot \boldsymbol{x}_i{}^T \boldsymbol{x}_j \\ \boldsymbol{G}_{i,k}^{12} &= (-2) \cdot [\tilde{\boldsymbol{x}}_k - \underset{\sim}{\boldsymbol{x}}_k]^T \boldsymbol{x}_i \\ \boldsymbol{G}^{21} &= (\boldsymbol{G}^{12})^T \\ \boldsymbol{G}_{k,l}^{22} &= 4 \cdot (\tilde{\boldsymbol{x}}_k - \underset{\sim}{\boldsymbol{x}}_k)^T (\tilde{\boldsymbol{x}}_k - \underset{\sim}{\boldsymbol{x}}_k) \end{aligned} \tag{20}$$

This representation can be used to transform the optimization problem into the following quadratic programming form:

$$\begin{aligned} &\mathrm{Max}_{\alpha,\beta} Q(\alpha,\beta) = -\frac{1}{2}[\alpha^T \quad \beta^T]\boldsymbol{G}\begin{bmatrix}\alpha \\ \beta\end{bmatrix} + (\boldsymbol{x}_i{}^T \boldsymbol{x}_i)_1^N \alpha - \left[\tilde{\boldsymbol{x}}_k{}^T \cdot \tilde{\boldsymbol{x}}_k - \underset{\sim}{\boldsymbol{x}}_k{}^T \cdot \underset{\sim}{\boldsymbol{x}}_k\right]_1^M \beta \\ &\text{Subject to} \sum_i \alpha_i = 1\,, \\ &\qquad 0 \le \alpha_i \le C, \beta_k \ge 0 \\ &\qquad i = 1, \dots, N,\ k = 1, \dots, M. \end{aligned} \tag{21}$$

If $\boldsymbol{G}$ is a positive semidefinite matrix, a global solution exists for this matrix. When $\boldsymbol{G}$ is a positive definite matrix, a global and unique solution exists for this matrix. Moreover, when $\boldsymbol{G}$ is an indefinite matrix, multiple local solutions might exist. The aforementioned variations can be addressed using optimization techniques—such as interior point and conjugate gradient methods—or optimization software—such as CPLEX, LINDO, or MATLAB quadprog. For simplicity, we used MATLAB quadprog in our implementation.

The algorithm of the MC-SVDD model first selects an appropriate kernel function. If an RBF kernel is selected, the hyperparameter C and kernel parameter σ are optimized through k-fold cross-validation on the training and validation data, with grid search being employed to determine the optimal values.

Figure 1 illustrates the structured methodology employed in this study for developing the MC-SVDD model. The process begins with the collection and normalization of data, ensuring numerical stability and minimizing biases that could impact the model's performance. Once the data is prepared, monotonic relationships between features and the target variable are identified based on domain knowledge. These relationships are then translated into monotonicity constraints, providing a structured framework that integrates expert insights into the learning process.

With the constraints defined, the methodology progresses to embedding them within the SVDD framework, modifying the objective function to enforce monotonicity while optimizing the hypersphere volume. This step ensures the model learns decision boundaries that are consistent with domain-specific knowledge. To enhance generalization, a grid search is used to fine-tune critical parameters, such as the kernel function and regularization coefficients.

The optimization problem, now constrained by monotonicity, is solved using the interior-point method, which guarantees convergence to an optimal solution. Once the model is trained, its performance is evaluated using metrics such as accuracy, precision, and recall, ensuring it meets both predictive and interpretive standards. The final step involves selecting the best-performing MC-SVDD model based on the optimization results and validating it against unseen data. This validation confirms the model's robustness, interpretability, and practical applicability.

This carefully designed methodology ensures an optimal balance between predictive accuracy and monotonicity compliance, making the MC-SVDD model a reliable tool for applications in anomaly detection, fraud detection, and medical diagnosis.

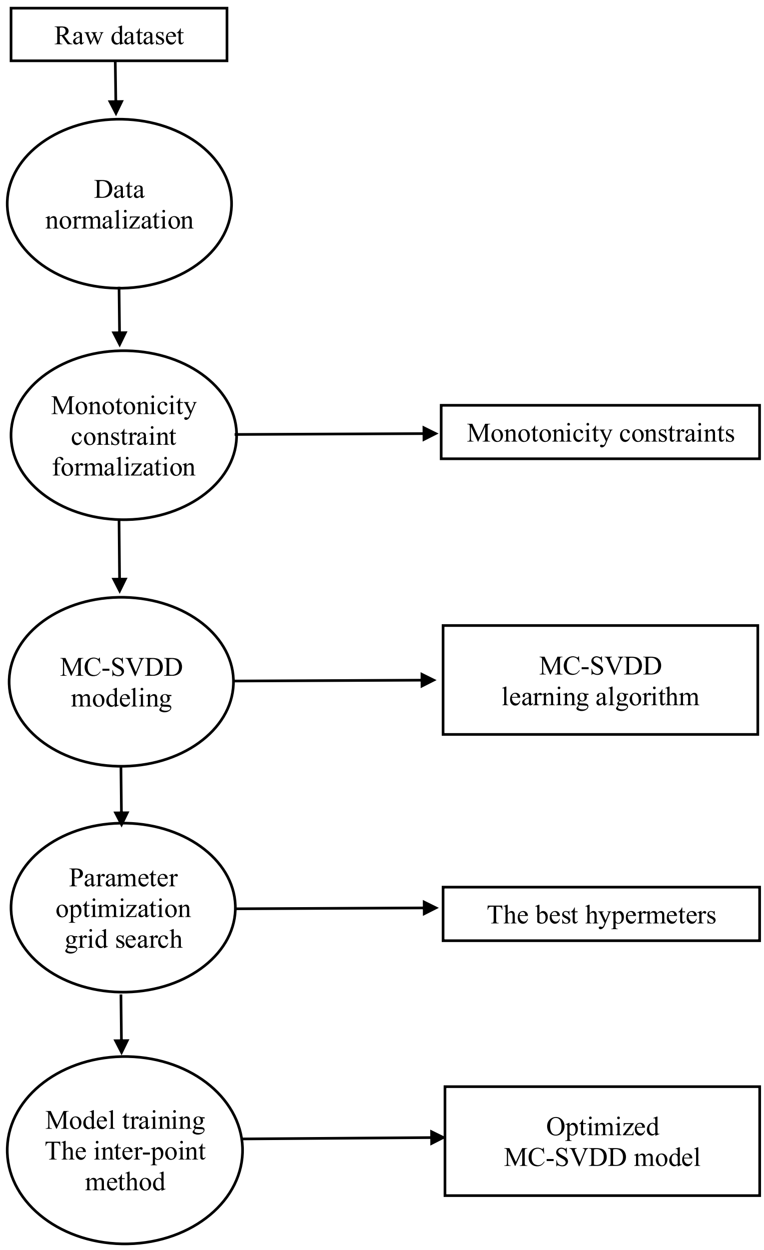


Figure 1: Data flow diagram of this study

4 Experiments and Analysis

This section presents results from experiments where the proposed MC-SVDD model was evaluated against a vanilla SVDD model.

4.1 Experimental Design

In our experiments, we employed four synthetic and nine real-world datasets, which are detailed in Section 4.2, to ensure a robust evaluation across both controlled and practical conditions. The MC-SVDD model was implemented using MATLAB R2017b in the Windows Server 2008 operating system on a system with an Intel Core i7-3770 CPU clocked at 3.4-GHz with 32 GB of RAM.

The preprocessing steps resulted in the data attributes being normalized to fall within the range [0, 1]. Subsequently, we partitioned the data into two sets: 80% of the data served as the training set, and the remaining 20% of the data constituted the test set. To ensure robustness, we divided the data into groups based on attribute values to achieve roughly equivalent group sizes and similar attribute distributions within each group.

We conducted five-fold cross-validation to assess the classification performance; this technique is widely used for validating a model's effectiveness and generalization capabilities.

We employed a grid search method to identify the most suitable model parameters. The hyperparameter C and kernel parameter σ were examined across the following range of candidate values:

$$C = \{2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^{0}, 2^{1}, 2^{2}\},$$
$$\sigma = \{2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^{0}, 2^{1}, 2^{2}\}.$$

For constraint selection, we randomly selected 10% of the nonrepeated data from the training set and conducted experimentation with the following numbers of constraints on the basis of recommendations in prior research [Chen and Li, 2014, Li and Chen, 2015, Chuang et al., 2020]:

$$MC = \{30,60,90,120,150,180\}.$$

After establishing the optimal models, we used the test data to evaluate the prediction performance of these models in terms of standard metrics, such as accuracy, F score, area under the receiver operating characteristic curve (AUC), and Matthews correlation coefficient (MCC). MCC, which is a widely used measure for assessing the quality of binary classifications [Matthews, 1975], indicates the correlation between the observed and predicted values. This coefficient is defined as follows:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}. \tag{22}$$

where TP, TN, FP, and FN represent the numbers of true positives, true negatives, false positives, and false negatives, respectively, within a confusion matrix. MCC ranges from −1 to 1, with 1 indicating a perfect prediction that completely agrees with the observation and −1 indicating a prediction that entirely contradicts the observed data.

The experimental results were averaged over 30 trials to ensure fair comparisons between all methods. Furthermore, we conducted one-tailed paired t-tests to assess the similarity between each pair of measurements. A p value lower than 0.05 indicated a statistically significant difference.

4.2 Data Sets

Four synthetic datasets were used in the experiments, each generated from distinct functions—exponential, horizontal line, sigmoid, and sine. These datasets were specifically designed to control key properties such as class distribution, noise levels, and feature correlations, allowing systematic evaluation under well-defined conditions. Importantly, the test data in these experiments follows OCC principles, where only positive class samples are available during training, and outlier detection is conducted without explicit negative class labels.

Complementarily, nine real-world datasets were employed to validate the practical applicability of the proposed approach. Unlike synthetic datasets, real-world datasets introduce greater complexity and variability, reflecting the unpredictability encountered in practical OCC tasks such as anomaly detection and fraud detection. In all cases, the evaluation does not assume predefined class distributions as in unbalanced classification, reinforcing the methodology's adherence to OCC principles.

To provide a clear understanding of the synthetic data sets, Table 1 presents the target functions used to generate the data points for the target class (1) and

Hyperplane function	Target function for one-class data
Exponential	$f(x_1, x_2) = \begin{cases} 1, & if \;\; x_2 < e^{x_1} \\ -1, & if \;\; x_2 > e^{x_1} \end{cases}$
Horizontal line	$f(x_1, x_2) = \begin{cases} 1, & if \;\; x_2 > 0.5 \\ -1, & if \;\; x_2 < 0.5 \end{cases}$
Sigmoid	$f(x_1, x_2) = \begin{cases} 1, & if \;\; x_2 > \dfrac{1}{1+e^{-x_1}} \\ -1, & if \;\; x_2 < \dfrac{1}{1+e^{-x_1}} \end{cases}$
Sine	$f(x_1, x_2) = \begin{cases} 1, & if \;\; x_2 > sin\, x_1 \\ -1, & if \;\; x_2 < sin\, x_1 \end{cases}$

Table 1: Target functions for the four synthetic one-class data sets

Data set	Source	No. of Instance	No. of Attribute	Attribute Type	Output Classes
Auto MPG	UCI	392	7	Numeric	high (> 28): 106 (+) low (≤ 28): 286
Car Acceptability	UCI	390	6	Qualitative	Acceptable:117 (+) Unacceptable:273
Cleveland heart disease	UCI	299	18	Numeric	> 50% diameter:138 (+) < 50% diameter:161
German	UCI	1000	20	Qualitative	good:700 (+) bad:300
Haberman BC Survival	UCI	306	3	Numeric	Survived ≥ 5 years: 225 (+) Survived < 5 years:81
PD600	Taiwan Local bank	600	17	Qualitative	Non-default: 300 (+) default: 300
Pima	UCI	768	8	Numeric	diabetes: 268 (+) normal: 500
South African Heart Disease	KEEL	462	9	Numeric	> 50% diameter:160 (+) < 50% diameter:302
WDBC	UCI	683	9	Numeric	Benign: 444 (+) Malignant: 239

UCI: University of California, Irvine

Table 2: Information on the real-world experimental data sets

nontarget class (−1). Notably, for the exponential and sigmoid functions, x_1 and x_2 exhibited increasing trends, whereas for the horizontal line and sine functions, only x_2 exhibited a monotonic increase.

The real-world data sets (Table 2) comprised multiple data sets obtained from the University of California, Irvine [Bache and Lichman, 2013], the PD600 data set obtained from a local bank in

Data set	Monotone Increasing	Monotone Decreasing
Auto Mileage	Origin, Year	Displacement, Weight
Car Acceptability	Persons, Safety	Price, Maintenance Required
Cleveland heart disease	Age, Type Angina, Atypical Angina, Male, BP, Cholestoral, ECG=1, ECG=2, Exercise Angina, Exercise ST depression, Number of Vessels Fluoroscopy	Max Heart Rate
German	Loan During, Single Appliance (vs Guarantor), Co-	Cheque Balance, Savings, Mouths in Job, Age, Other Installment Plans, Owns House

	Appliance (vs Guarantor), Renting.	
Haberman BC Survival	Year	Age, Nodes
PD600	Real estate, Monthly salary, Real income, Record of payment	Years in the current job, Credit
Pima	Plasma glucose concentration a 2 hours in an oral glucose tolerance test	-
South African	Systolic BP, Cum Tobacco, Cholesterol, Adiposity, Family History, Type A Behavior, Age, Obesity	-
WDBC	All increasing	-

Table 3: Monotonic relationships in the real-world data sets

Taiwan [Li et al., 2006, Chen and Li, 2014], and a data set obtained from the KEEL repository [Alcalá-Fdez et al., 2011]. Instances with missing values were carefully excluded. In the aforementioned data sets, the output class indicated by "(+)" is the target class. Monotonic relationships were established through consultation with domain experts and references to previous studies ([Li and Chen, 2015]; Table 3). Subsequently, monotonic constraints were constructed on the basis of these relationships.

4.3 Experimental Evaluation: Synthetic Data Sets

Table 4 details the performance results of the proposed MC-SVDD model and vanilla SVDD model for comparison, both of which use RBF and linear kernels. Notable improvements in performance metrics are highlighted in bold. As presented in Table 4, the MC-SVDD model outperformed the SVDD model and exhibited consistently higher accuracy scores under both kernel settings. This result indicates that the MC-SVDD model can effectively segregate testing data into target and outlier data.

Kernel function	Data sets	MC-SVDD					SVDD				
		Accuracy	Specificity	F-Score	MCC	AUC	Accuracy	Specificity	F-Score	MCC	AUC
RBF	Exponential	**0.7917**	**0.6456**	**0.8192**	**0.6358**	**1.0000**	0.6156	0.3074	0.7033	0.3258	0.8389
	Line	**0.9557**	**0.9402**	**0.9558**	**0.9157**	**1.0000**	0.8649	0.8197	0.8639	0.7511	1.0000
	Sigmoid	**0.8918**	**0.8374**	**0.8979**	**0.8008**	**1.0000**	0.7759	0.6441	0.7848	0.5929	0.9706
	Sine	**0.8767**	**0.8013**	**0.8853**	**0.7727**	**1.0000**	0.8257	0.7035	0.8442	0.6853	0.9936

Linear	Exponential	**0.7583**	**0.5691**	**0.7929**	**0.5826**	**0.9760**	0.5195	0.1177	0.6549	0.1223	0.7785
	Line	**0.8390**	**0.7458**	0.8463	**0.7067**	**0.9887**	0.8347	0.7182	**0.8486**	0.6990	0.9706
	Sigmoid	**0.6962**	**0.4893**	**0.7470**	**0.4491**	**0.9367**	0.6533	0.3709	0.7295	0.3922	0.8245
	Sine	**0.7920**	**0.6346**	**0.8183**	**0.6313**	**0.9418**	0.7813	0.5953	0.8135	0.6153	0.8556

Table 4: Performance of the MC-SVDD and SVDD models on the synthetic data sets

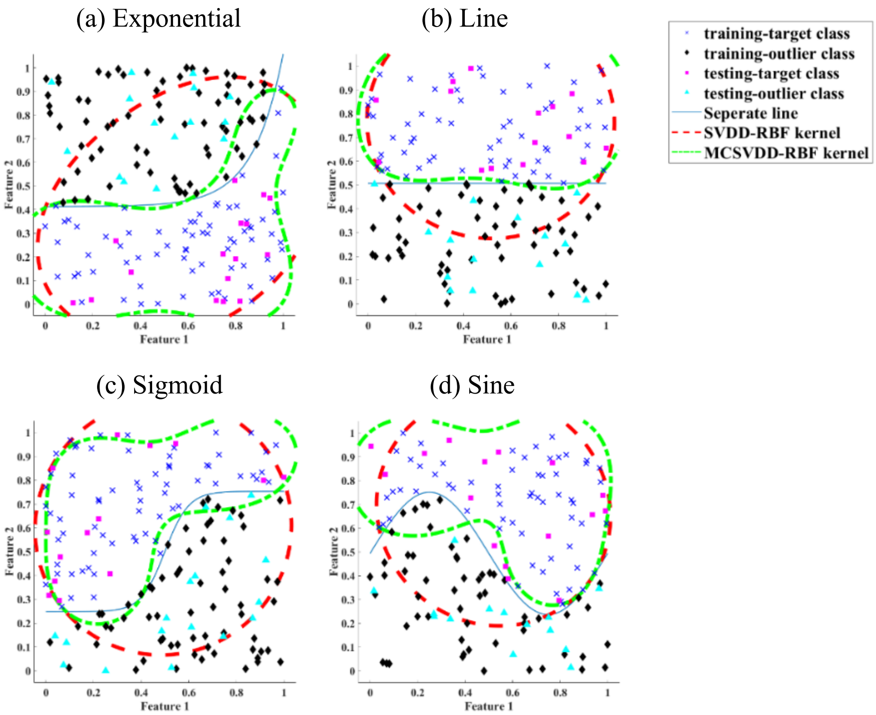


Figure 2: Hyperspheres of the MC-SVDD an SVDD models for the synthetic data sets

Furthermore, the MC-SVDD model significantly outperformed the SVDD model in terms of specificity and F score; this difference was found to be significant in a paired t-test. The mean improvement in specificity across synthetic datasets was 74.31%, while the F-score improvement was 8.77%. This result suggests that the proposed model more accurately excluded data from the outlier class within the hypersphere than did the SVDD model.

The results of the MC-SVDD and SVDD models for the synthetic data sets are depicted in Figure 2. This figure indicates that the MC-SVDD outperformed the SVDD model in classifying continuously distributed target classes, namely those in the data sets generated from the line, exponential, sine, and sigmoid functions. The monotonic constraints imposed on the MC-SVDD model allowed for a more effective and adaptive decision boundary, enhancing classification precision.

Conventional one-class classifiers often incorporate outlier data into the target class, which leads to diminished model performance during training. However, the results of this study indicate that the introduction of monotonic characteristics into a one-class classifier allows it to achieve effective and precise hypersphere fitting and sorting in classification tasks. This phenomenon underscores the crucial role of the information hidden in monotone constraints in data set classification methods. The MC-SVDD model emulated cognitive characteristics and effectively handled all target classes. Consequently, it consistently yielded superior results to the SVDD model in one-class classification tasks.

Figure 3 displays the receiver operating characteristic (ROC) curves for the synthetic data sets; these curves highlight the relationship between classifier performance and the selected threshold. In practical applications of one-class classifiers, the selection of an appropriate threshold is essential.

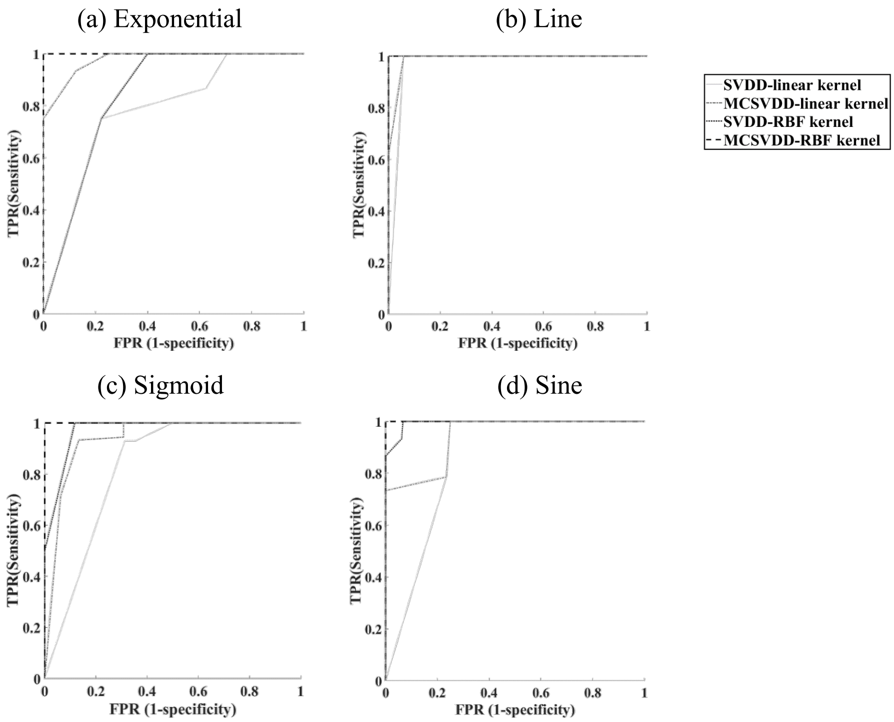


Figure 3: ROC curves of the MC-SVDD and SVDD models for the synthetic data sets

However, accurately estimating the optimal threshold during training is challenging because no outlier samples are available for assessing specificity. Consequently, one-class classifiers assume a fraction (ε) of legitimate target data to be outliers for defining the boundary of the target data, with the typical ε value being set as 10%. Therefore, in this study, we set ε as 10%. The MCC and AUC values for the

MC-SVDD and SVDD models are presented in Table 4. These results demonstrated that the MC-SVDD model had superior generalization and predictive capabilities relative to the SVDD model for the target class. The MCC and AUC values of the MC-SVDD model indicated that it had robust performance, particularly in datasets with continuous distributions.

4.4 Experimental Evaluation: Real-World Data Sets

The results obtained with the MC-SVDD and SVDD models for the real-world data sets are presented in Table 5. These results indicate the robust generalization capacity of the MC-SVDD model for the target class. One-class classification tasks for real-world data sets are inherently more complex than are those for synthetic data, primarily because of the varying degrees of overlap between the outlier class and the target class in real-world data. On average, one-class classifiers exhibit lower accuracy for real-world data sets than for synthetic data sets. Achieving high accuracy in one-class classification for real-world data is a formidable challenge, especially in scenarios characterized by significant class overlaps.

Kernel function	Data sets	MC-SVDD					SVDD				
		Accuracy	Specificity	F-Score	MCC	AUC	Accuracy	Specificity	F-Score	MCC	AUC
RBF	Auto MPG	**0.7320**	**0.6676**	**0.6381**	**0.5107**	**0.8847**	0.6587	0.5479	0.6030	0.4379	0.8721
	Car Acceptability	**0.8047**	**0.7823**	**0.7173**	**0.6140**	**0.9755**	0.7814	0.7487	0.6873	0.5763	0.9557
	Cleveland heart disease	**0.7248**	**0.6323**	**0.7264**	**0.4790**	**0.8622**	0.6492	0.5511	0.6225	0.3484	0.7915
	German	**0.7067**	**0.0676**	**0.8240**	**0.1277**	**0.5521**	0.6951	0.0505	0.8168	0.0558	0.5432
	Haberman BC Survival	**0.7347**	0.0407	**0.8450**	**0.0691**	**0.5656**	0.7207	**0.0451**	0.8352	0.0338	0.5231
	PD600	**0.6456**	**0.7934**	0.5503	**0.3838**	**0.7735**	0.5749	0.2142	**0.6877**	0.2158	0.6740
	Pima	**0.6714**	**0.0910**	**0.7956**	**0.1715**	**0.6310**	0.6550	0.0390	0.7881	0.0583	0.6290
	South African	**0.6873**	**0.1344**	**0.8040**	**0.2261**	**0.6475**	0.6633	0.0988	0.7889	0.1218	0.5533
	WDBC	**0.9496**	**0.9465**	**0.9621**	**0.8908**	**0.9972**	0.9263	0.8255	0.9455	0.8389	0.9855
Linear	Auto MPG	**0.6580**	**0.5588**	**0.6085**	**0.4516**	**0.8456**	0.6005	0.4625	0.5744	0.4003	0.8113
	Car Acceptability	**0.7831**	**0.7435**	0.6910	0.5836	**0.9409**	0.7653	0.6864	**0.7094**	**0.5844**	0.9000
	Cleveland heart disease	**0.6969**	**0.5756**	**0.7185**	**0.4380**	**0.8320**	0.6548	0.4439	0.7089	0.3780	0.8141
	German	**0.7065**	**0.0478**	**0.8251**	**0.1148**	**0.5276**	0.6941	0.0323	0.8173	0.0362	0.5207
	Haberman BC Survival	**0.7336**	0.0306	**0.8448**	**0.0556**	**0.5374**	0.7228	**0.0390**	0.8369	0.0318	0.5231
	PD600	**0.5801**	**0.1852**	**0.6989**	**0.2639**	**0.6176**	0.5363	0.1142	0.6740	0.1363	0.5837
	Pima	**0.6623**	**0.0446**	**0.7929**	**0.1311**	**0.5598**	0.6548	0.0316	0.7885	0.0692	0.5412
	South African	**0.6823**	**0.1094**	**0.8023**	**0.2064**	0.5789	0.6663	0.0832	0.7925	0.1336	**0.6250**
	WDBC	0.8601	0.8031	0.9382	0.8152	**0.9663**	**0.9236**	**0.8130**	**0.9437**	**0.8329**	0.9649

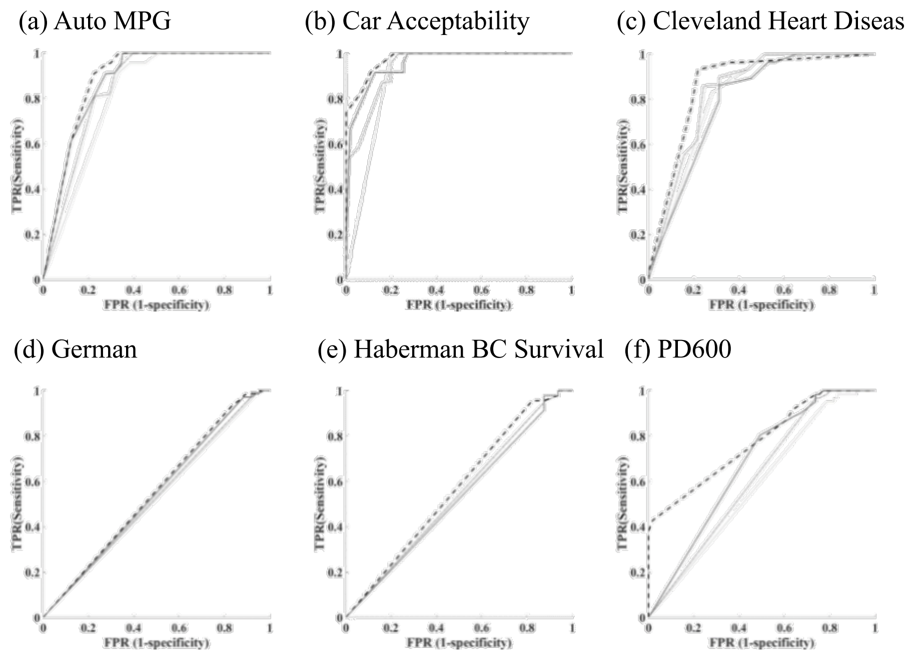
Table 5: Performance of the MC-SVDD and SVDD models for the real-world data sets

We conducted pairwise comparisons between the results obtained with the MC-SVDD and SVDD models for further analysis. When the RBF kernel was used, the MC-SVDD model achieved significantly higher specificity than did the SVDD model (57.73% improvement). However, as presented in Table 6, the MC-SVDD model had significantly lower sensitivity than did the SVDD model. Nevertheless, this result was acceptable because the MC-SVDD model exhibited high accuracy and MCC values. This trade-off between specificity and sensitivity suggests that the MC-SVDD model effectively minimizes false positives while maintaining robust classification boundaries.

When a linear kernel was used, the MC-SVDD model exhibited significantly higher specificity than did the SVDD model for the Auto MPG, German, PD600, Pima, and South African data sets. Thus, the hypersphere of the MC-SVDD model effectively captured the characteristics of these data sets. For the Cleveland Heart Disease and Car Acceptability data sets, the MC-SVDD model achieved higher accuracy and specificity than did the SVDD model. Moreover, the F scores and MCC values of the MC-SVDD model indicated that it developed a suitable decision boundary for these data sets.

RBF kernel			
MC-SVDD		SVDD	
Sensitivity	Specificity	Sensitivity	Specificity
0.4978	**0.7934**	**0.9355**	0.2142

Table 6: Sensitivity and specificity of the MC-SVDD and SVDD models for the PD600 data set when the RBF kernel was used



(g) Pima (h) South African (i) WDBC

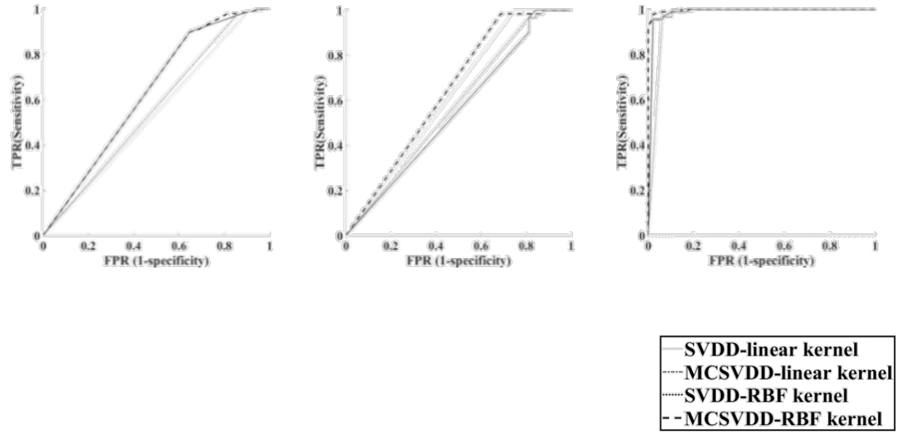


Figure 4: ROC curves of the MC-SVDD and SVDD models for the real-world data sets

However, for the Haberman BC Survival data set, the MCC and specificity of the MC-SVDD model approached 0, which suggested that its hypersphere was possibly too large to encompass the outlier data effectively. This highlights the challenge of using hypersphere-based classifiers on datasets with extreme class imbalances. Furthermore, for the WDBC data set, the SVDD model outperformed the MC-SVDD model. Nevertheless, the MC-SVDD model outperformed the SVDD model on the other data sets.

In general, the MC-SVDD and SVDD models exhibited better performance with the RBF kernel than with the linear kernel. The difference was particularly pronounced for specificity, which led to higher accuracy, F score, and MCC values with the RBF kernel than with the linear kernel. This result can be attributed to the fact that the RBF kernel can classify data more effectively than can the linear kernel, especially when nonlinear patterns are present in the dataset.

However, for both models, the sensitivity achieved with the RBF kernel was lower than that achieved with the linear kernel, especially for the PD600 data set. As presented in Table 6, a common characteristic for these the data set was that the hypersphere trained by the RBF kernel was too small to encompass all the target data. Moreover, the hypersphere trained by the linear kernel was too large and included excessive outlier data in the target data. This situation is reflected in the sensitivity and specificity values, which indicate the challenges faced when data close to the hypersphere contain a mixture of target and outlier data.

The ROC curves of the two models for the real-world data sets are shown in Figure 4. The AUC values in Table 5 confirmed that the MC-SVDD model had superior generalization and prediction abilities compared to the SVDD model for the target classes of all the real-world data sets, which highlighted the consistent prediction performance of the MC-SVDD model. Moreover, the prediction performance of both models worsened as the overlap between the two classes increased. Consequently, further improvements are required to the proposed model to obtain accurate predictions for the German and Haberman BC Survival data sets. Future research should explore adaptive hypersphere sizing techniques to address this issue.

4.5 Comparison of MC-SVDD with Recent Works

To further contextualize the MC-SVDD model within the landscape of monotonic classification and one-class classification, we compare its theoretical underpinnings with recent studies in the field. This comparison underscores the distinctive contributions of MC-SVDD in achieving a balance between interpretability, computational efficiency, and classification accuracy, while addressing key challenges that existing methods do not fully resolve.

One of the fundamental challenges in OCC is defining a hypersphere that effectively encloses the target class while minimizing its volume. Traditional SVDD models often rely on a fixed kernel function, making it difficult to adapt to different data distributions. Guo et al. [Guo et al., 2021] introduced Multi-Kernel Learning SVDD (MKL-SVDD) to address this limitation by incorporating boundary sample weights and kernel optimization. While this approach improves classification performance through better kernel selection, it does not enforce monotonicity constraints. In contrast, MC-SVDD explicitly integrates monotonic constraints into the SVDD framework, ensuring that classification decisions adhere to domain-specific monotonic relationships, which is particularly important in regulated environments such as credit scoring, medical diagnosis, and risk assessment.

Monotonic classification has also been studied extensively in deep learning models to improve fairness, interpretability, and generalization. Zhao et al. [Zhao et al., 2024] proposed the Deep Isotonic Embedding Network (DIEN), a neural network-based method that enforces monotonicity through isotonic embedding vectors. While DIEN effectively handles monotonic and non-monotonic features, deep learning models generally require large training datasets and significant computational resources, making them impractical for certain real-world applications. By contrast, MC-SVDD leverages kernel methods, which allow it to enforce monotonicity constraints efficiently while remaining computationally lightweight.

The combination of monotonic constraints with imbalanced classification has also been explored in the literature. Zhu et al. [Zhu et al., 2021] introduced Weighted Multi-Objective Classification SLFN (WMCS-SLFN), which applies a multi-objective genetic algorithm to optimize monotonicity and classification performance in imbalanced datasets. While their approach achieves high accuracy, it requires evolutionary optimization, which substantially increases computational complexity. In contrast, MC-SVDD embeds monotonic constraints directly within the hypersphere optimization framework, maintaining both interpretability and efficiency without the need for computationally intensive optimization procedures.

From a probabilistic perspective, Zhang and Jatowt [Zhang and Jatowt, 2020] explored Bayesian OCC using configurable one-class Naïve Bayes models. Their method enhances classification performance by refining clustering quality metrics but lacks explicit geometric constraints for hypersphere-based classification. MC-SVDD, by contrast, introduces explicit monotonicity constraints into the hypersphere formulation, enabling structured decision-making in domains where expert knowledge dictates monotonic trends (e.g., disease progression, financial risk models).

Rule-based monotonic classification techniques have also been studied as a means to enforce monotonicity constraints in classification models. Verbeke et al. [Verbeke et al., 2017] proposed RULEM, a post-processing heuristic that modifies rule-based models (e.g., decision trees) to enforce monotonicity. While RULEM ensures that the model output remains monotonic, it introduces additional complexity by requiring post-hoc rule modifications. MC-SVDD, in contrast, integrates monotonicity constraints directly into its optimization framework, eliminating the need for additional rule-processing steps and making it inherently more scalable and adaptable to various datasets.

By addressing limitations inherent in kernel optimization, deep learning, evolutionary optimization, Bayesian models, and rule-based methods, MC-SVDD achieves a unique balance of classification accuracy, interpretability, and computational efficiency. This makes it a highly viable solution for real-world applications requiring both monotonicity constraints and one-class classification capabilities.

5 Novel Contributions

This study introduces a novel approach to integrating monotonicity constraints into OCC, addressing a significant research gap. While monotonic constraints have been widely studied in binary and multiclass classification, their application to one-class classification remains unexplored. This work presents the MC-SVDD model, which extends the SVDD framework by embedding monotonicity constraints into the classification process. This ensures that decision boundaries remain consistent with domain-specific knowledge, enhancing both interpretability and practical applicability.

To achieve this, we propose a rigorous mathematical formulation that transforms one-class classification into a constrained quadratic optimization problem, enabling structured decision-making under monotonic constraints. We further develop a practical training algorithm that efficiently enforces these constraints using quadratic programming solvers, ensuring computational feasibility.

The empirical evaluation of MC-SVDD demonstrates its superiority over conventional SVDD in classification accuracy, while preserving monotonicity across diverse datasets. This work not only improves the theoretical foundations of one-class classification but also bridges the gap between monotonic learning and outlier detection, extending the benefits of structured domain knowledge to previously unexplored classification tasks.

By incorporating monotonicity constraints into one-class classification, this study provides a more interpretable, theoretically sound, and computationally efficient solution for real-world applications where structured domain knowledge plays a critical role. These contributions lay a foundation for future research on integrating domain-driven constraints into machine learning models, opening new avenues for theoretical and applied advancements.

6 Conclusions and Future Work

In this study, we developed a novel one-class classification model that considers monotonic constraints. In many knowledge engineering and machine learning applications, domain experts often provide prior knowledge regarding monotonic relationships between response and distance variables. Monotonicity is a critical modeling requirement for enhancing decision explanations and support in specific scenarios.

The developed MC-SVDD model is an extension of the SVDD model, which has exhibited excellent performance in one-class classification problems involving real-world and synthetic data sets. In contrast to the SVDD model, the MC-SVDD model can consider the monotonic characteristics of one-class classification problems, thereby formulating them as quadratic programming problems in a dual space. We devised an algorithm to train the MC-SVDD model. To evaluate the developed model, we applied it and the SVDD model to nine benchmark real-world data sets and four 2-dimensional synthetic data sets. The decision boundaries of the MC-SVDD model for each data set indicated its excellent prediction performance for target classes with a continuous distribution. These decision boundaries effectively excluded outlier data, which led to highly accurate predictions. The proposed model outperformed the SVDD model for almost all the examined data sets, with the differences in the evaluation parameters between the two models being statistically significant ($p < 0.05$). Moreover, the ROC curves and AUC values of the MC-SVDD and SVDD models indicated that the MC-SVDD model had superior prediction performance. Thus, the results of this study underscore the substantial contribution of prior domain knowledge in one-class classification.

In this study, the conjugate gradient algorithm was used to solve the relevant quadratic programming problem. However, the computational efficiency of this algorithm notably decreases with an increase in the number of monotonicity constraints. Although this algorithm can proficiently solve quadratic programming problems, future studies should investigate whether other algorithms, such as the stochastic gradient descent algorithm, can improve the scalability of the developed MC-SVDD model under monotonicity constraints and accelerate its training phase.

Acknowledgements

This study was supported by the National Science and Technology Council, Taiwan, under the grants MOST 108-2410-H-006-107 and MOST 109-2410-H-006-044.

References

[Alcalá-Fdez et al., 2011] Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., & Herrera, F. (2011). KEEL Data-Mining Software Tool: Data set repository, integration of algorithms and experimental analysis framework. Journal of Multiple-Valued Logic and Soft Computing, 17, 255-287.

[Bache and Lichman, 2013] Bache, K., & Lichman, M. (2013). UCI Machine Learning Repository. University of California, School of Information and Computer Science. [Online]. Available: http://archive.ics.uci.edu/ml.

[Banerjee et al., 2007] Banerjee, A., Burlina, P., & Meth, R. (2007). Fast hyperspectral anomaly detection via SVDD. In Proceedings of the IEEE International Conference on Image Processing (ICIP) (Vol. 4, pp. IV101-IV104).

[Bartley et al., 2016] Bartley, C., Liu, W., & Reynolds, M. (2016). Effective monotone knowledge integration in kernel support vector machines. In Proceedings of the 12th International Conference on Advanced Data Mining and Applications (ADMA) (pp. 3-18). Springer.

[Ben-Hur et al., 2001] Ben-Hur, A., Horn, D., Siegelmann, H. T., & Vapnik, V. (2001). Support vector clustering. Journal of Machine Learning Research, 2, 125-137.

[Camastra and Verri, 2005] Camastra, F., & Verri, A. (2005). A novel kernel method for clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(5), 801-805.

[Cano et al., 2019] Cano, J.-R., Gutiérrez, P. A., Krawczyk, B., Woźniak, M., & García, S. (2019). Monotonic classification: An overview on algorithms, performance measures and data sets. Neurocomputing, 341, 168-182.

[Cao, 2010] Cao, L. (2010). Domain-driven data mining: Challenges and prospects. IEEE Transactions on Knowledge and Data Engineering, 22(6), 755-769.

[Chen and Li, 2014] Chen, C. C., & Li, S. T. (2014). Credit rating with a monotonicity-constrained support vector machine model. Expert Systems with Applications, 41(16), 7235-7247.

[Chuang et al., 2020] Chuang, H. C., Chen, C. C., & Li, S. T. (2020). Incorporating monotonic domain knowledge in support vector learning for data mining regression problems. Neural Computing and Applications, 32, 11791-11805.

[Dablain et al., 2023] Dablain, D., Krawczyk, B., & Chawla, N. V. (2023). DeepSMOTE: Fusing deep learning and SMOTE for imbalanced data. IEEE Transactions on Neural Networks and Learning Systems, 34(9), 6390-6404. https://doi.org/10.1109/TNNLS.2021.3136503.

[Eryarsoy et al., 2009] Eryarsoy, E., Koehler, G. J., & Aytug, H. (2009). Using domain-specific knowledge in generalization error bounds for support vector machine learning. Decision Support Systems, 46(2), 481-491.

[Guo et al., 2021] Guo, W., Wang, Z., Hong, S., Li, D., Yang, H., & Du, W. (2021). Multi-kernel support vector data description with boundary information. Engineering Applications of Artificial Intelligence, 102, Article 104254. https://doi.org/10.1016/j.engappai.2021.104254.

[Hayashi et al., 2022] Hayashi, T., Cimr, D., Studnička, F., Fujita, H., Bušovský, D., & Cimler, R. (2022). OCSTN: One-class time-series classification approach using a signal transformation network into a goal signal. Information Sciences, 614, 71-86. https://doi.org/10.1016/j.ins.2022.09.027

[Hayashi et al., 2024] Hayashi, T., Cimr, D., Studnička, F., Fujita, H., Bušovský, D., & Cimler, R. (2024). Patient deterioration detection using one-class classification via cluster period estimation subtask. Information Sciences, 657, 119975. https://doi.org/10.1016/j.ins.2023.119975

[Hu et al., 2012a] Hu, Q., Che, X., Zhang, L., Guo, M., & Yu, D. (2012). Rank entropy-based decision trees for monotonic classification. IEEE Transactions on Knowledge and Data Engineering, 24(11), 2052-2064.

[Hu et al., 2012b] Hu, Q., Pan, W., Song, Y., & Yu, D. (2012). Large-margin feature selection for monotonic classification. Knowledge-Based Systems, 31, 8-18.

[Jiang et al., 2012] Jiang, Z., Feng, X., Feng, X., & Li, L. (2012). A study of SVDD-based algorithm to the fault diagnosis of mechanical equipment system. Physics Procedia, 33, 1068-1073.

[Ju et al., 2022] Ju, H., Lee, D., Kang, S. K., & Yu, H. (2022). Mitigating viewpoint sensitivity of self-supervised one-class classifiers. Information Sciences, 611, 225-242. https://doi.org/10.1016/j.ins.2022.08.042

[Khan and Madden, 2014] Khan, S. S., & Madden, M. G. (2014). One-class classification: Taxonomy of study and review of techniques. Knowledge Engineering Review, 29(3), 345-374.

[Kotlowski and Slowinski, 2013] Kotlowski, W., & Slowinski, R. (2013). On nonparametric ordinal classification with monotonicity constraints. IEEE Transactions on Knowledge and Data Engineering, 25(11), 2576-2589.

[Lee et al., 2006] Lee, S.-W., Park, J., & Lee, S.-W. (2006). Low-resolution face recognition based on support vector data description. Pattern Recognition, 39(9), 1809-1812.

[Li and Chen, 2015] Li, S. T., & Chen, C. C. (2015). A regularized monotonic fuzzy support vector machine model for data mining with prior knowledge. IEEE Transactions on Fuzzy Systems, 23(5), 1713-1727.

[Li et al., 2006] Li, S.-T., Shiue, W., & Huang, M.-H. (2006). The evaluation of consumer loans using support vector machines. Expert Systems with Applications, 30(4), 772-782.

[Liu et al., 2013] Liu, B., Xiao, Y., Cao, L., Hao, Z., & Deng, F. (2013). SVDD-based outlier detection on uncertain data. Knowledge and Information Systems, 34(3), 597-618.

[Matthews, 1975] Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta: Protein Structure, 405(2), 442-451.

[Mazhelis, 2006] Mazhelis, O. (2006). One-class classifiers: A review and analysis of suitability in the context of mobile-masquerader detection. South African Journal of Science, 6(36), 29-48.

[Mercer, 1909] Mercer, J. (1909). Functions of positive and negative type, and their connection with the theory of integral equations. Philosophical Transactions of the Royal Society A, 209, 415-446.

[Park et al., 2007] Park, J., Kang, D., Kim, J., Kwok, J. T., & Tsang, I. W. (2007). SVDD-based pattern denoising. Neural Computation, 19(7), 1919-1938.

[Peng et al., 2023] Peng, H., Zhao, J., Li, L., Ren, Y., & Zhao, S. (2023). One-class adversarial fraud detection nets with class-specific representations. IEEE Transactions on Network Science and Engineering, 10(6), 3793-3803. https://doi.org/10.1109/TNSE.2023.3273543.

[Perera et al., 2021] Perera, P., Oza, P., & Patel, V. M. (2021). One-class classification: A survey. arXiv preprint arXiv:2101.03064.

[Potharst and Feelders, 2002] Potharst, R., & Feelders, A. J. (2002). Classification trees for problems with monotonicity constraints. SIGKDD Explorations, 4(1), 1-10.

[Reid, 1970] Reid, C. (1970). Hilbert-Courant. Springer Science & Business Media.

[Schölkopf and Smola, 2002] Schölkopf, B., & Smola, A. J. (2002). Learning with kernels: Support vector machines, regularization, optimization, and beyond. MIT Press.

[Schölkopf et al., 2001] 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.

[Sinha and Zhao, 2008] Sinha, A. P., & Zhao, H. (2008). Incorporating domain knowledge into data mining classifiers: An application in indirect lending. Decision Support Systems, 46(1), 287-299.

[Tax and Duin, 1999] Tax, D. M., & Duin, R. P. (1999). Support vector domain description. Pattern Recognition Letters, 20(11), 1191-1199.

[Tax and Duin, 2002] Tax, D. M., & Duin, R. P. (2002). Using two-class classifiers for multiclass classification. In Proceedings of the International Conference on Pattern Recognition (ICPR) (Vol. 2, pp. 124-127).

[Tax and Duin, 2004] Tax, D. M., & Duin, R. P. (2004). Support vector data description. Machine Learning, 54(1), 45-66.

[Tax, 2001] Tax, D. M. J. (2001). One-class classification: Concept-learning in the absence of counter-examples (Ph.D. dissertation). TU Delft.

[Vapnik, 2013] Vapnik, V. (2013). The nature of statistical learning theory. Springer Science & Business Media.

[Verbeke et al., 2017] Verbeke, W., Martens, D., & Baesens, B. (2017). RULEM: A novel heuristic rule learning approach for ordinal classification with monotonicity constraints. Applied Soft Computing, 60, 858-873. https://doi.org/10.1016/j.asoc.2017.01.042.

[Wang et al., 2004] Wang, Q., Lopes, L. S., & Tax, D. J. (2004). Visual object recognition through one-class learning. In Proceedings of the International Conference on Image Analysis and Recognition (ICIAR) (pp. 463-470).

[Witten et al., 2016] Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools and Techniques (4th ed.). Morgan Kaufmann Publishers Inc.

[Xu, 2017] Xu, Y. (2017). Maximum margin of twin spheres support vector machine for imbalanced data classification. IEEE Transactions on Cybernetics, 47(6), 1540-1550. https://doi.org/10.1109/TCYB.2016.2551735.

[Yin et al., 2018] 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.

[Yu et al., 2010] Yu, T., Simoff, S., & Jan, T. (2010). VQSVM: A case study for incorporating prior domain knowledge into inductive machine learning. Neurocomputing, 73(13), 2614-2623.

[Zhang and Jatowt, 2020] Zhang, Y., & Jatowt, A. (2020). Estimating a one-class naive Bayes text classifier. Intelligent Data Analysis, 24(3), 567-579. https://doi.org/10.3233/IDA-194669.

[Zhang et al., 2023] Zhang, W., Wu, Q. M. J., Zhao, W. G. W., Deng, H., & Yang, Y. (2023). Hierarchical one-class model with subnetwork for representation learning and outlier detection. IEEE Transactions on Cybernetics, 53(10), 6303-6316. https://doi.org/10.1109/TCYB.2022.3166349

[Zhao et al., 2024] Zhao, J., Zhang, H., Wang, Y., Zhai, Y., & Yang, Y. (2024). Deep isotonic embedding network: A flexible monotonic neural network. Neural Networks, 171, 457-465. https://doi.org/10.1016/j.neunet.2023.12.026.

[Zhou et al., 2017] Zhou, Y., Wu, K., Meng, Z., & Tian, M. (2017). Fault detection of aircraft based on support vector domain description. Computers & Electrical Engineering, 61, 80-94.

[Zhu et al., 2021] Zhu, H., Liu, H., & Fu, A. (2021). Class-weighted neural network for monotonic imbalanced classification. International Journal of Machine Learning and Cybernetics, 12, 1191-1201. https://doi.org/10.1007/s13042-020-01228-x.