Tampere University

FAHAD SOHRAB

Subspace Support Vector Data Description and Extensions

Tampere University Dissertations 603

Tampere University Dissertations 603

FAHAD SOHRAB

Subspace Support Vector Data Description and Extensions

ACADEMIC DISSERTATION
To be presented, with the permission of
The Faculty of Information Technology and Communication Sciences
of Tampere University,
for public discussion in the auditorium TB109
of the Tietotalo, Korkeakoulunkatu 1, Tampere,
on 27 May 2022, at 14 o'clock.

ACADEMIC DISSERTATION
Tampere University,
Faculty of Information Technology and Communication Sciences
Finland

Responsible supervisor and Custos
Professor
Moncef Gabbouj
Tampere University
Finland

Supervisor
Doctor
Jenni Raitoharju
Tampere University and
Finnish Environment Institute
Finland

Pre-examiners
Associate Professor
Vishal M. Patel
Johns Hopkins University
USA

Associate Professor
Miguel Bordallo López
University of Oulu
Finland

Opponent
Doctor, HDR
Hichem Sahbi
Sorbonne University
Paris, France

Cover design: Roihu Inc.

ISBN 978-952-03-2408-7 (print)
ISBN 978-952-03-2409-4 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-2409-4

PunaMusta Oy – Yliopistopaino
Joensuu 2022

Dare to develop machine learning techniques for circumstances that we cannot imagine.

PREFACE

The research work presented in this thesis has been carried out at the Signal Analysis and Machine Intelligence (SAMI) research group at Tampere University, Finland. The research was funded by the National Science Foundation (NSF), Industry-University Cooperative Research Center (IUCRC), Center for Visual and Decision Informatics (CVDI), and Business Finland. The financial support of Haltian, Tietoevry, and Dead Set Bit is gratefully acknowledged. I would like to acknowledge the Nokia Foundation award as well.

I want to express my gratitude to my supervisors, Prof. Moncef Gabbouj and Dr. Jenni Raitoharju, for their guidance, support, and encouragement during my doctoral studies. I genuinely believe that I would have never made it this far without the mentoring of my supervisors. I was guided in the direction in which I could excel. I extend my gratitude to Prof. Alexandros Iosifidis (Aarhus University, Denmark) for the guidance, mentoring, and support as well.

I am grateful to my pre-examiners, Prof. Vishal M. Patel (Johns Hopkins University, USA) and Prof. Miguel Bordallo López (University of Oulu, Finland), for reviewing this thesis and providing positive feedback and valuable comments. I am also thankful to Dr. Hichem Sahbi (Sorbonne University France) for agreeing to serve as Opponent during the thesis defense.

I would also like to warmly thank Prof. Raju Gottumukkala (University of Louisiana at Lafayette, USA) and his team for their support during my time at the University of Louisiana at Lafayette. I would like to especially thank Matti Vakkuri, Head of Research at Haltian, for the encouragement and support during my doctoral studies.

I would like to show my deepest appreciation to various members of our research group: Anton Muravev, Aysen Degerli, Bilge Can Pullinen, Dat Thanh Tran, Farhad Pakdaman, Firas Laakom, Junaid Malik, Kateryna Chumachenko, Lei Xu, Mehmet Yamaç, Mete Ahıshalı, and Mohammad Fathi Al-Sa'd for their friendship, great dis-

cussions and gossips. I am also thankful to the former members and visitors of our research group for their friendship.

Finally, I would like to express gratitude to my parents and siblings for their continued encouragement and support, I dedicate this thesis to them.

Tampere, April 2022
Fahad Sohrab

ABSTRACT

Machine learning deals with discovering the knowledge that governs the learning process. The science of machine learning helps create techniques that enhance the capabilities of a system through the use of data. Typical machine learning techniques identify or predict different patterns in the data. In classification tasks, a machine learning model is trained using some *training data* to identify the unknown function that maps the input data to the output labels. The classification task gets challenging if the data from some categories are either unavailable or so diverse that they cannot be modelled statistically. For example, to train a model for anomaly detection, it is usually challenging to collect anomalous data for training, but the normal data is available in abundance. In such cases, it is possible to use One-Class Classification (OCC) techniques where the model is trained by using data only from one class.

OCC algorithms are practical in situations where it is vital to identify one of the categories, but the examples from that specific category are scarce. Numerous OCC techniques have been proposed in the literature that model the data in the given feature space; however, such data can be high-dimensional or may not provide discriminative information for classification. In order to avoid the *curse of dimensionality*, standard dimensionality reduction techniques are commonly used as a preprocessing step in many machine learning algorithms. Principal Component Analysis (PCA) is an example of a widely used algorithm to transform data into a subspace suitable for the task at hand while maintaining the meaningful features of a given dataset.

This thesis provides a new paradigm that jointly optimizes a subspace and data description for one-class classification via Support Vector Data Description (SVDD). We initiated the idea of subspace learning for one class classification by proposing a novel Subspace Support Vector Data Description (SSVDD) method, which was further extended to Ellipsoidal Subspace Support Vector Data Description (ESSVDD). ESSVDD generalizes SSVDD for a hypersphere by using ellipsoidal data description and it converges faster than SSVDD. It is important to train a joint model

for multimodal data when data is collected from multiple sources. Therefore, we also proposed a multimodal approach, namely Multimodal Subspace Support Vector Data Description (MSSVDD) for transforming the data from multiple modalities to a common shared space for OCC. An important contribution of this thesis is to provide a framework unifying the subspace learning methods for SVDD. The proposed Graph-Embedded Subspace Support Vector Data Description (GESSVDD) framework helps revealing novel insights into the previously proposed methods and allows deriving novel variants that incorporate different optimization goals.

The main focus of the thesis is on generic novel methods which can be adapted to different application domains. We experimented with standard datasets from different domains such as robotics, healthcare, and economics and achieved better performance than competing methods in most of the cases. We also proposed a taxa identification framework for rare benthic macroinvertebrates. Benthic macroinvertebrate taxa distribution is typically very imbalanced. The amounts of training images for the rarest classes are too low for properly training deep learning-based methods, while these rarest classes can be central in biodiversity monitoring. We show that the classic one-class classifiers in general, and the proposed methods in particular, can enhance a deep neural network classification performance for imbalanced datasets.

CONTENTS

NOMENCLATURE

$\mathbf{1}$	Vector containing all ones
$\mathbf{1}_c$	Vector having ones at elements corresponding to instances that belong to class c and zeros elsewhere
$\mathbf{a}$	Center of hypersphere
$\mathbf{A}$	Graph weight matrix
$\mathbb{A}$	Diagonal matrix containing α values in its diagonal
c	Class index
C	Hyper-parameter to regulate number of outliers
$\mathscr{C}$	Total number of classes
d	Subspace dimensionality
D	Dimensionality of original feature space
D_m	Dimensionality of original feature space of modality m
$\mathbf{D}$	Diagonal degree matrix
$\mathbf{E}$	Covariance of data in d-dimensional space
$\mathscr{F}$	Kernel space
$\mathscr{G}$	Undirected weighted graph
$\mathbf{I}$	Identity matrix
$\mathbf{J}$	Matrix with all values identically $\frac{1}{N}$
$\mathbf{k}_*$	Test instance kernel vector
$\hat{\mathbf{k}}_*$	Centralized test instance kernel vector
$\mathbf{K}$	Kernel matrix
$\hat{\mathbf{K}}$	Centralized kernel matrix

$\mathbf{L}$	Laplacian matrix of intrinsic graph
$\mathbf{L}_b$	Laplacian matrix of between-class scatter
$\mathbf{L}_p$	Laplacian matrix of penalty graph
$\mathbf{L}_t$	Laplacian matrix of total scatter
$\mathbf{L}_w$	Laplacian matrix of within-class scatter
$\mathbf{L}_x$	Laplacian matrix
M	Total number of modalities
N	Number of training instances
N_c	Total number of samples that belong to class c
$\mathbf{q}$	Eigenvector used for forming the projection matrix $\mathbf{Q}$
$\mathbf{Q}$	Projection matrix
$\mathbf{Q}_m$	Projection matrix used for modality m
R	Radius of hypersphere
s	Support vector
$\mathbf{S}_b$	Between-class scatter matrix
$\mathbf{S}_Q$	Matrix encoding geometric data relationship in subspace
$\mathbf{S}_t$	Total scatter matrix
$\mathbf{S}_x$	Matrix encoding geometric data relationship in original feature space
$\mathbf{S}_\alpha$	Matrix encoding data relationship based on SVDD graph
$\mathbf{S}_w$	Within-class scatter matrix
t	Eigenvector used in spectral regression
$\mathrm{Tr}(\cdot)$	Trace operator
$\mathbf{u}$	Center of data description relative to hypersphere
$\mathbf{U}$	Matrix forming eigenvectors of $\hat{\mathbf{K}}$
$\mathbf{x}$	D-dimensional input vector
$\mathbf{x}_*$	Test vector in its original feature space
$\mathbf{x}_m$	Input vector in modality m

$\mathbf{X}$	Matrix containing training data of D dimensions by N samples
$\mathbf{X}_m$	Data matrix of modality m
$\mathbf{y}$	Data vector in d-dimensional space
$\mathbf{y}_*$	Vector of test instance in d-dimensional space
$\mathbf{y}_m$	Data vector in d-dimensional space coming from modality m
$\mathbf{z}$	Vector in d-dimensional space used for denoting projected data in graph-embedded subspace SVDD
$\mathbf{Z}_m$	Data representation for linear and NPT case in MSSVDD
α	Lagrange multiplier associated with data vector in SVDD
$\boldsymbol{\alpha}$	Vector containing α values
α_m	Lagrange multiplier associated with data vector $\mathbf{x}_m$ of modality m
$\boldsymbol{\alpha}_m$	Vector containing α_m values
β	Hyper-parameter for controlling the importance of regularization term(s)
γ	Lagrange multiplier
γ_m	Lagrange multiplier used for data in modality m
ϵ	Small non-zero constant
η	Learning rate parameter
$\boldsymbol{\lambda}$	Vector used for selecting different regularization strategies
$\boldsymbol{\Lambda}$	Diagonal matrix containing non-negative eigenvalues in its diagonal
ξ	Slack variable to account for outliers
ξ_m	Slack variable to account for outliers in modality m
σ	Hyper-parameter to determine the width of Gaussian kernel
υ	Eigenvalue
Υ	Regularization term introduced in ESSVDD
ψ	Regularization term introduced in SSVDD
ω	Regularization term introduced in MSSVDD

LIST OF ABBREVIATIONS

Acc	Accuracy
CNN	Convolutional Neural Network
EOCC	Ensemble One-Class Classification
ESSVDD	Ellipsoidal Subspace Support Vector Data Description
ESVDD	Ellipsoidal Support Vector Data Description
FP	False Positive
GM	Geometric Mean
GESSVDD	Graph-Embedded Subspace Support Vector Data Description
GESVDD	Graph-Embedded Support Vector Data Description
GESVM	Graph-Embedded Support Vector Machine
LDA	Linear Discriminant Analysis
MSSVDD	Multimodal Subspace Support Vector Data Description
NPT	Non-Linear Projection Trick
OCC	One-Class Classification
OCSVM	One-Class Support Vector Machine
PCA	Principal Component Analysis
PD	Positive Detected
Pre	Precision
RBF	Radial Basis Function
SSVDD	Subspace Support Vector Data Description

SVDD	Support Vector Data Description
SVM	Support Vector Machine
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate

LIST OF PUBLICATIONS

[P1] F. Sohrab, J. Raitoharju, M. Gabbouj and A. Iosifidis, "Subspace Support Vector Data Description," International Conference on Pattern Recognition, 2018, pp. 722-727, doi: 10.1109/ICPR.2018.8545819.

[P2] F. Sohrab, J. Raitoharju, A. Iosifidis and M. Gabbouj, "Ellipsoidal Subspace Support Vector Data Description," IEEE Access, vol. 8, pp. 122013-122025, 2020, doi: 10.1109/ACCESS.2020.3007123.

[P3] F. Sohrab, J. Raitoharju, A. Iosifidis and M. Gabbouj, "Multimodal Subspace Support Vector Data Description," Pattern Recognition, vol. 110, p. 107648, 2021. doi: 10.1016/j.patcog.2020.107648.

[P4] F. Sohrab, A. Iosifidis, M. Gabbouj and J. Raitoharju, "Graph-Embedded Subspace Support Vector Data Description, arXiv:2104.14370.

[P5] F. Sohrab and J. Raitoharju, "Boosting Rare Benthic Macroinvertebrates Taxa Identification With One-Class Classification," IEEE Symposium Series on Computational Intelligence (SSCI), 2020, pp. 928-933, doi: 10.1109/SSCI47803.2020.9308359.

Author's contribution

Publication I	The publication proposes SSVDD, which maps the data to a subspace optimized for OCC. The candidate contributed to the mathematical formulation, carried out the simulations, and wrote the paper overseen by the co-authors.
Publication II	The publication proposes ESSVDD, which is an improved version of SSVDD. ESSVDD takes into account the covariance of the data in the subspace. The candidate contributed to the mathematical formulation, carried out the simulations, and wrote the paper.
Publication III	The publication proposes MSSVDD, which is a multimodal version of SSVDD. The candidate carried out the mathematical formulation, implementation, and writing of the paper.
Publication IV	The publication proposes GESSVDD, which is a graph-embedded formulation covering all the SSVDD variants. The candidate contributed to the mathematical formulation, simulations, and writing of the paper.
Publication V	The publication proposes a pipeline, boosting rare benthic macroinvertebrates taxa identification with OCC. The candidate contributed to enhancing the proposed idea, carried out the implementations, and wrote the paper.

1 INTRODUCTION

In data classification, the overall goal is to create a predictive data model that can accurately classify unseen input data into a set of pre-defined classes. In supervised learning, the class labels of a part of the data are used for training. In contrast, in unsupervised learning, possible patterns in the data are identified without any data labels. Traditional classification algorithms aim at learning a classification model for several pre-defined categories. However, it can be challenging to gather data from some of the categories in certain situations. For example, in fraud detection [1], abnormal event detection in videos [2], and nosocomial infection detection [3], most of the possible situations representing fraud, abnormal events in videos, and nosocomial infections in clinical data, respectively, cannot be observed for training the model. Nevertheless, detecting all sorts of abnormal situations is crucial.

Unavailability of data from one or several of the categories led to the inception of machine learning methods that require data only from one class during the training process. One-Class Classification (OCC) methods are used to create a model for predicting whether an unseen sample comes from this class of interest. The class/ category of data used for obtaining the data description is referred to as the target class or the positive class, while all other categories combined are referred to as the negative class or outliers. No outlier data are needed for training the methods.

OCC techniques are helpful when it is costly to collect the data from the negative class, or the data are so diverse that it cannot be modelled statistically. In scenarios like predicting the unexpected breakdown of an electric induction motor [4], monitoring the gearbox of helicopters [5], or monitoring the structural health of a bridges [6], the data instances of the catastrophic situations are not feasible to collect. However, the normal data corresponding to the positive class is available in abundance.

OCC methods have been thoroughly studied in the literature for more than two decades. The improvement of OCC methods has led to their usage in many technological applications, such as in bot detection on Twitter [7]. In [2], OCC is used

for abnormal event detection in videos, while in [8] it is employed to detect and recognize low-flying, small unmanned aerial vehicles efficiently. In [9], it is used for railway vehicle detection from audio recordings. OCC is also used in the medical field for anomaly detection in biomedical images [10]. Other common applications of OCC are document classification [11], disease diagnosis [12], intrusion detection [13], and novelty detection [14].

OCC problems are traditionally tackled mainly by three approaches: density estimation, reconstruction-based, and border-based description [15]. The density estimation and reconstruction-based approaches are suitable when the target data follows a selected distribution and assumptions about the data generating process are accurate. The challenge in density estimation and reconstruction-based approaches is identifying an appropriate distribution for the data at hand [16]. On the other hand, border-based approaches do not require prior information about the density or the data generation process. Border-based approaches are also referred to as support vector-based approaches, where *support vectors* are instances in the training data that are crucial for defining the data description and influence the decision boundary of the model.

While practical large-scale OCC applications are nowadays based on deep learning [17], many of the fundamental ideas are still based on traditional OCC algorithms. For example, a deep OCC method, which maps the data representations into a hypersphere characterized by center and radius of minimum volume, is proposed in [18]. The loss function used for training the neural network is a modified version of the objective function traditionally used in OCC [15]. Thus, deepening the understanding of traditional OCC algorithms remains an important topic and has been the main focus of this thesis.

The different approaches for OCC in the literature mainly focus on data from a single modality. Moreover, most traditional methods model the data in the given feature space and do not find more discriminative features in lower-dimensional space. These methods lack the capability of integrating different data relations in the optimization process. Thus, there is a need to create new methods and an adequate framework that fills the identified gaps and provides a more detailed analysis into OCC literature.

In this thesis, we seek answers to the following research questions: Can we create a methodology where the subspace learning and model learning complement each

other for improved OCC performance? Can we further improve the results by replacing the traditional hyperspherical description with a more flexible one? Can we find an optimized subspace and create a methodology for training a common model for multimodal data in the case of OCC? Can we formulate subspace learning for OCC in a graph-embedded framework? To answer these questions, we set our objectives as described in the following subsection.

1.1 Objectives

In this thesis, we focus on improving the performance of support vector-based data descriptions for OCC. The assumption is that a data description for OCC in an optimized subspace would improve the performance during inference compared to the data description in the original feature space. The more specific objectives of the thesis are as follows:

- To propose a new subspace learning paradigm for OCC which incorporates the model information in the optimization process of subspace learning for OCC.
- To further optimize the data description by considering alternatives for the traditional hyperspherical description.
- To infer a shared latent representation for multimodal data and exploit the relationship between the modalities in OCC.
- To integrate the proposed solutions in a unified framework and highlight the similarities and differences between the solutions therein.
- To assess the performance of OCC methods over different domain applications.

1.2 Contributions and Publications

A significant contribution of this thesis is providing a new paradigm of jointly optimizing subspace learning and OCC via Support Vector Data Description (SVDD). The initial work, which forms the basis of this thesis, was introduced in [P1]. In [P2], we introduced a generalized solution as compared to [P1]. We proposed a novel mul-

timodal extension of [P1] in [P3]. As a major contribution of this thesis, we built a unified framework of subspace learning for OCC which was proposed in [P4]. Finally, we demonstrated the capability of OCC methods to improve the performance of a deep Convolutional Neural Network (CNN)-based algorithm used to identify rare benthic macroinvertebrates in [P5]. The candidate implemented the proposed methods, contributed to the mathematical formulation, carried out the simulations, and wrote the manuscripts overseen by the co-authors. The relationship between the different contributions proposed in this thesis is depicted in Figure 1.1.

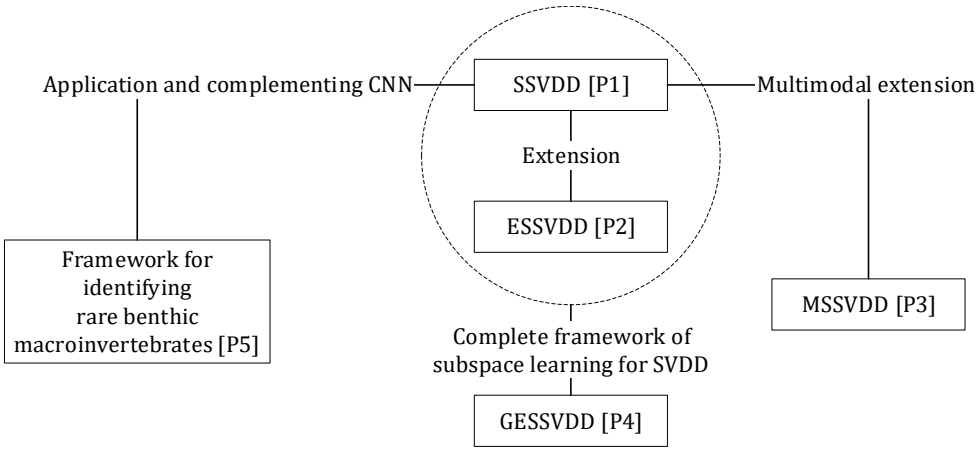

Figure 1.1 Schematic view showing relations between different contributions of the thesis.

In [P1], we proposed a gradient-based iterative subspace learning method for OCC. The proposed method, namely Subspace Support Vector Data Description (SSVDD), maps the data into a lower-dimensional space optimized for OCC. In the optimized lower-dimensional space, the target data is encapsulated in a hypersphere. We proposed linear and non-linear versions and four different regularization strategies for the proposed SSVDD in this publication. SSVDD forms the basis of the thesis, and we further enhanced, extended, and analyzed it in the other publications.

In [P2], we proposed a novel method, namely Ellipsoidal Subspace Support Vector Data Description (ESSVDD), to transform data from the original feature space to a lower-dimensional space optimized for OCC. ESSVDD considers the covariance of data in the optimization process and yields a more generalized solution than SSVDD. In ESSVDD, class variance in the projected space is used to express the regularization terms. We presented both linear and non-linear formulations of the proposed ESSVDD, and the simulation results show that ESSVDD converges faster

and performs better than SSVDD.

In [P3], we proposed a novel method for mapping data from multiple modalities to a common shared subspace along with forming a common model. The proposed method, namely Multimodal Subspace Support Vector Data Description (MSSVDD), defines a separate projection matrix for each modality and transforms the data to a common shared subspace, while the classification model is optimized jointly for data from all the modalities. During the iterative optimization process, data from only the class of interest is utilized. We presented both linear and non-linear formulations, along with seven different regularization strategies. We also carried out complexity analysis of the proposed method. The obtained results show that MSSVDD yields better results in majority of cases than traditional concatenation of multimodal data on five multimodal datasets.

In [P4], we proposed a framework that unifies the previously proposed variants and several novel variants of SSVDD under a graph-embedded formulation called Graph-Embedded Subspace Support Vector Data Description (GESSVDD). The proposed framework allowed us to reveal and analyze connections to traditional subspace learning algorithms and led to novel spectral and spectral regression-based solutions as alternatives to the gradient-based technique. The novel SSVDD variants derived from the framework led to improved performance against the baselines and other competing methods.

In [P5], we proposed a framework for identifying rare benthic macroinvertebrates where the specimens possibly representing rare species are directed for further human expert inspection. We analyzed the capability of traditional OCC methods to complement and improve a deep CNN-based algorithm for dividing the tasks between humans and machines. In the experiments, regularized linear SSVDD produced the best performance among the applied OCC methods. The experiments also support the idea of moving from fully manual to semi-automated taxa identification.

1.3 Thesis Outline

The rest of this thesis is organized as follows. Chapter 2 provides an overview of the related work relevant to subspace learning for OCC. An overview of graph embedding and non-linear transformations is briefly discussed in Chapter 2 as well.

The contributions of the thesis are described in Chapter 3, summarizing the main findings in the publications. The conclusions are drawn in Chapter 4, where future work is also proposed. The publications are included at the end of this thesis.

2 RELATED WORK

This chapter provides an overview of concepts needed for understanding the contributions of this thesis. We first discuss OCC and elaborate on SVDD [19]. The second section provides an overview of dimensionality reduction focusing on subspace learning. Subsequently, this chapter discusses graph embedding along with spectral regression and summarizes Graph-Embedded Support Vector Data Description (GESVDD) [20]. The last section of the chapter briefly discusses non-linear transformations used in deriving the non-linear versions of the proposed methods in this thesis.

2.1 One-Class Classification

OCC algorithms aim to build a model by considering data from a single class [21]. The class of interest used for modeling is referred to as the positive class or the target class. All other categories not included in the training process are referred to as the negative class or outliers. OCC aims to capture the information from a single class to form a data description that can accurately classify unseen data into either the positive or negative class during inference. Figure 2.1 depicts the basic idea of OCC.

OCC techniques are mainly solved by density-based [22], reconstruction-based [23], or border-based approaches [15]. The density-based approaches for OCC rely on estimating the density of the target class by using a density estimation method, such as Parzen density [22], Gaussian model [15], or a mixture of Gaussians [24]. The density-based approaches are applicable in cases when a high number of training samples are available. In reconstruction-based approaches, the underlying function for data description is formed by using some prior information, such as data clustering characteristics [25]. In reconstruction-based methods, domain-specific prior information is essential for training a predictive model. In border-based approaches,

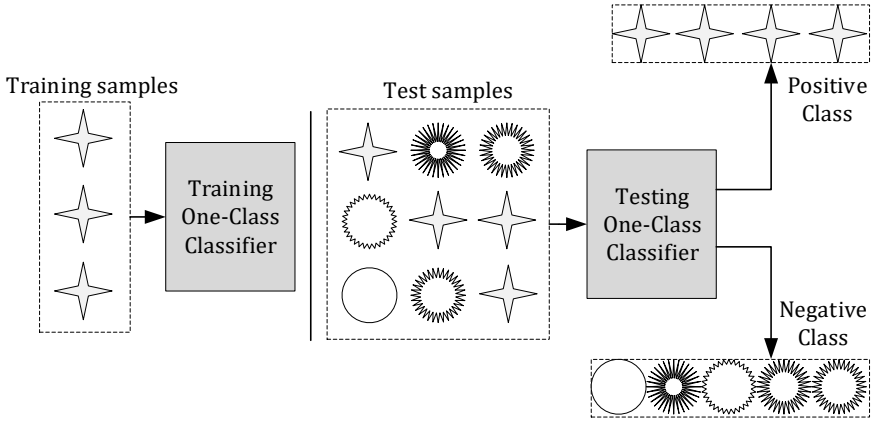

Figure 2.1 In One-Class Classification, a model is trained by using samples of the positive class only. During testing, the model is used to classify samples also from the negative class [P4].

data description is obtained by defining a closed boundary around the target class without estimating its density [19]. One-Class Support Vector Machine (OCSVM) [26], and SVDD [19] are classic examples of border-based techniques for OCC. In OCSVM, a hyperplane is constructed, which separates the target class from the origin. In SVDD, a hypersphere with minimum volume enclosing the target class in a given feature space is formed. The border-based approaches are generally easier to apply than density-based and reconstruction-based approaches as they do not require estimating density nor domain-specific prior information of the data generating process. However, the common border-based approaches operate in a feature space that is not optimized for OCC. Thus, an approach for transforming data into an optimized subspace suited for OCC can help improve the overall classification accuracy.

In the literature, deep learning techniques for OCC have been recently proposed as well. Many of these techniques are based on the traditional approaches that find a closed tight boundary around the target class data. For example, in [18], deep SVDD, which jointly trains a neural network along with optimizing a hypersphere in the output space, is proposed. In [27], a deep learning-based approach for one-class transfer learning is proposed. The benefit of deep learning-based techniques for OCC is end-to-end learning of feature representation and modeling. The challenging aspect of deep learning-based techniques for OCC is the requirement of large amounts of target class data during training. Moreover, a deep learning-based approach typically relies on minimizing a classical one-class loss function on the learned final layer representations, which suffers from the fundamental drawback of representation collapse

[28]. Representation collapse refers to inferring a degenerate solution where all the points are mapped to one single point. In [29], generative adversarial networks are used to generate a typical data distribution for deep anomaly detection; however, generating an entire data distribution is challenging, and inaccurate in practice [28].

2.1.1 Support Vector Data Description

SVDD [19] finds a spherically shaped description of the data in a given feature space. The training data points used for obtaining a tight boundary are denoted by a matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_N], \mathbf{x}_i \in \mathbb{R}^D$, where N is the total number of samples in D-dimensional feature space. The samples used for obtaining the data description belong to only one class. The function to be minimized in SVDD is as follows:

$$\begin{aligned} \min \quad & F(R, \mathbf{a}) = R^2 \\ \text{s.t.} \quad & ||\mathbf{x}_i - \mathbf{a}||_2^2 \leq R^2, \ \forall i \in \{1, \ldots, N\}, \end{aligned} \tag{2.1}$$

where R denotes the radius of the hypersphere and $\mathbf{a} \in \mathbb{R}^D$ denotes the center of the hypersphere. In order to accommodate for outliers in the data points, slack variables ξ_i, $i = 1, \ldots, N$ are introduced into the optimization problem:

$$\begin{aligned} \min \quad & F(R, \mathbf{a}) = R^2 + C \sum_{i=1}^{N} \xi_i \\ \text{s.t.} \quad & ||\mathbf{x}_i - \mathbf{a}||_2^2 \leq R^2 + \xi_i, \\ & \xi_i \geq 0, \ \forall i \in \{1, \ldots, N\}. \end{aligned} \tag{2.2}$$

The hyperparameter $C > 0$ is used for controlling the trade-off between the volume of the hypersphere and the amount of target data outside the hypersphere. The constraints are incorporated into the minimization function by using Lagrange mul-

tipliers as

$$\begin{aligned}
L =& R^2 + C\sum_{i=1}^{N}\xi_i + \sum_{i=1}^{N}\alpha_i\big(\|\mathbf{x}_i-\mathbf{a}\|_2^2 - R^2 - \xi_i\big) - \sum_{i=1}^{N}\gamma_i\xi_i \\
=& R^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\big(R^2 + \xi_i - (\mathbf{x}_i-\mathbf{a})^\intercal(\mathbf{x}_i-\mathbf{a})\big) - \sum_{i=1}^{N}\gamma_i\xi_i \\
=& R^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\big(R^2 + \xi_i - \mathbf{x}_i^\intercal\mathbf{x}_i + 2\mathbf{a}^\intercal\mathbf{x}_i - \mathbf{a}^\intercal\mathbf{a}\big) - \sum_{i=1}^{N}\gamma_i\xi_i, \qquad (2.3)
\end{aligned}$$

where $\alpha_i \geq 0$ and $\gamma_i \geq 0$ are Lagrange multipliers. The Lagrangian (2.3) should be minimized with respect to R, $\mathbf{a}$, ξ_i and maximized with respect to the Lagrange multipliers α_i and γ_i [19]. Now taking partial derivatives with respect to variables R, $\mathbf{a}$, and ξ_i and setting them to zero, we get

$$\begin{aligned}
&\frac{\partial L}{\partial R} = 2R - 2R\sum_{i=1}^{N}\alpha_i = 2R(1-\sum_{i=1}^{N}\alpha_i) \\
&\frac{\partial L}{\partial R} = 0 \Rightarrow \sum_{i=1}^{N}\alpha_i = 1, \qquad (2.4)
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial L}{\partial \mathbf{a}} = -\sum_{i=1}^{N}\alpha_i\big(2\mathbf{x}_i - 2\mathbf{a}\big) = -2\sum_{i=1}^{N}\alpha_i\mathbf{x}_i + 2\sum_{i=1}^{N}\alpha_i\mathbf{a} = -2\sum_{i=1}^{N}\alpha_i\mathbf{x}_i + 2\mathbf{a} \\
&\frac{\partial L}{\partial \mathbf{a}} = 0 \Rightarrow \mathbf{a} = \sum_{i=1}^{N}\alpha_i\mathbf{x}_i. \qquad (2.5)
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial L}{\partial \xi_i} = \frac{\partial}{\partial \xi_i}C\sum_{i=1}^{N}\xi_i - \frac{\partial}{\partial \xi_i}\sum_{i=1}^{N}\alpha_i\xi_i - \frac{\partial}{\partial \xi_i}\sum_{i=1}^{N}\gamma_i\xi_i = C - \alpha_i - \gamma_i \\
&\frac{\partial L}{\partial \xi_i} = 0 \Rightarrow C - \alpha_i - \gamma_i = 0. \qquad (2.6)
\end{aligned}$$

From (2.6), it can be seen that $\alpha_i = C - \gamma_i$ and, since $\alpha_i \geq 0$ and $\gamma_i \geq 0$, it is demanded that $0 \leq \alpha_i \leq C$. After substituting (2.4), (2.5), and (2.6) into (2.3), the Lagrangian

dual of (2.2) can be simplified as

$$
\begin{aligned}
L = R^2 - \sum_{i=1}^{N} \alpha_i R^2 + C\sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \xi_i \alpha_i - \sum_{i=1}^{N} \xi_i \gamma_i + \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^{\intercal}\mathbf{x}_i - \sum_{i=1}^{N} \alpha_i 2\mathbf{a}^{\intercal}\mathbf{x}_i + \\
\sum_{i=1}^{N} \alpha_i \mathbf{a}^{\intercal}\mathbf{a} = R^2 - \sum_{i=1}^{N} \alpha_i R^2 + \sum_{i=1}^{N} \xi_i \big(C - \alpha_i - \gamma_i\big) + \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^{\intercal}\mathbf{x}_i - \sum_{i=1}^{N} \alpha_i 2\mathbf{a}^{\intercal}\mathbf{x}_i + \\
\sum_{i=1}^{N} \alpha_i \mathbf{a}^{\intercal}\mathbf{a} = R^2 - R^2 + \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^{\intercal}\mathbf{x}_i - 2\sum_{i=1}^{N} \alpha_i \mathbf{x}_i \sum_{j=1}^{N} \alpha_j \mathbf{x}_j + \sum_{i=1}^{N} \alpha_i \mathbf{x}_i \sum_{j=1}^{N} \alpha_j \mathbf{x}_j \\
= \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^{\intercal}\mathbf{x}_i - 2\sum_{i=1}^{N} \alpha_i \mathbf{x}_i \sum_{j=1}^{N} \alpha_j \mathbf{x}_j + \sum_{i=1}^{N} \alpha_i \mathbf{x}_i \sum_{j=1}^{N} \alpha_j \mathbf{x}_j.
\end{aligned}
\tag{2.7}
$$

Hence, the Lagrangian dual of (2.2) can be written as

$$
\begin{aligned}
\max \quad & L = \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^{\intercal}\mathbf{x}_i - \sum_{i}^{N}\sum_{j}^{N} \alpha_i \alpha_j \mathbf{x}_i^{\intercal}\mathbf{x}_j, \\
\text{s.t.} \quad & 0 \le \alpha_i \le C.
\end{aligned}
\tag{2.8}
$$

By solving (2.8), a set of α_i for each data point is obtained in the training process. This can be solved by using the sequential minimal optimization method [30] as implemented in the LIBSVM toolbox [31]. When an instance falls inside the hypersphere, the corresponding Lagrange multiplier will be zero. On the other hand, if $||\mathbf{x}_i - \mathbf{a}||_2^2 \ge R^2$, the Lagrange multiplier will be greater than zero. Thus, the obtained α_i values define the position of each instance in the description. The samples corresponding to $\alpha_i = 0$ lie inside the sphere, the samples corresponding to $0 < \alpha_i < C$ lie on the boundary of the hypersphere, and those with $\alpha_i = C$ lie outside the boundary of the hypersphere. To summarize

$$||\mathbf{x}_i - \mathbf{a}||_2 < R \rightarrow \alpha_i = 0, \tag{2.9}$$

$$||\mathbf{x}_i - \mathbf{a}||_2 = R \rightarrow 0 < \alpha_i < C, \tag{2.10}$$

$$||\mathbf{x}_i - \mathbf{a}||_2 > R \rightarrow \alpha_i = C. \tag{2.11}$$

In some previous studies [15], the term support vector is used for all the samples with a positive α_i value. The discrimination between support vectors lying on

the boundary and those outside the boundary is depicted using different notations. For the sake of simplicity, in this thesis, we only refer to the data points lying on the boundary as support vectors. The data points lying outside the boundary are referred to as outliers.

During testing, the decision to classify a test instance $\mathbf{x}_*$ as positive or negative depends on its distance from the center of the data description. The distance is calculated as follows:

$$||\mathbf{x}_* - \mathbf{a}||_2^2 = \mathbf{x}_*^\intercal \mathbf{x}_* - 2\mathbf{x}_*^\intercal \mathbf{a} + \mathbf{a}^\intercal \mathbf{a}. \tag{2.12}$$

If the above distance is greater than the radius R, the test instance is classified as an outlier; otherwise, it is accepted as a positive instance. The radius R can be calculated as follows:

$$R^2 = \mathbf{s}^\intercal \mathbf{s} - 2\mathbf{s}^\intercal \mathbf{a} + \mathbf{a}^\intercal \mathbf{a}, \tag{2.13}$$

where $\mathbf{s}$ is any support vector with $0 < \alpha_i < C$.

Various techniques have been proposed in the literature for improving the performance of SVDD [32, 33, 34, 35]. In [36], a technique for handling uncertain data in the training set is proposed. In the proposed technique, a confidence score is incorporated into SVDD where each training instance contributes differently to the construction of the decision boundary. In [34], additional terms for expressing global and local geometric information are proposed for SVDD. In [37], an ellipsoidal boundary is used instead of a hyperspherical boundary for better encapsulation of data. A method for removing voids in the hypersphere of SVDD is proposed in [33]. The proposed method tackles the issue of voids by identifying different clusters and training SVDD for each cluster and uses feature scaling and data centering techniques to make the scatter more spherical for the classifier. A similar approach is used in [38], which forms a multi-sphere SVDD for tackling multi-distribution data. A method for creating two boundaries simultaneously for target data containing two classes in the training data is proposed in [39].

The above-mentioned methods enhance the overall performance of SVDD for lower dimensional data; however, it does not address the *curse of dimensionality* [40]. The curse of dimensionality indicates that the number of instances needed to estimate the mapping function with a given accuracy grows exponentially with respect

to the number of input variables. Moreover, the multi-sphere SVDD approaches assume that the available data shares the common space, and hence multiple data descriptions are obtained in a given space. This assumption is not true when the data comes from multiple sources with different dimensionality. Creating an optimized subspace for unimodal data and a joint subspace for multimodal data has not been addressed in the literature.

2.2 Dimensionality Reduction

Dimensionality reduction refers to the process of reducing the dimensionality of a given dataset. It aims to represent the data in a lower-dimensional space in a way that the critical characteristics of the data in its original dimensions are preserved. Dimensionality reduction techniques can be broadly divided into two categories: feature selection and subspace learning.

2.2.1 Feature Selection

Feature selection is the process of selecting relevant features of the data according to some criteria [41], [42], [43]. Two main approaches are typically used to carry out feature selection for dimensionality reduction, namely the *filter* approach, and the *wrapper* approach [44]. Algorithms following the filter approach use the intrinsic characteristics of the data by performing statistical analysis and filtering out the most discriminative features. The filter approaches are model agnostic, as the features are selected without inferring any information from the model for the task at hand. The algorithms that follow the wrapper approach use some mechanism (e.g., a greedy search approach) to select features for modeling in low-dimensional space by leveraging information from the classifier [45].

Hybrid approaches combining numerous aspects of the filter and wrapper approaches have also been proposed [46, 47]. Embedded methods are another type of feature selection, which are akin to wrapper methods, but in embedded methods, the classification algorithms contain a built-in ability to select features [48]. While the wrapper and embedded approaches generally perform better in terms of classification accuracy, they are computationally more expensive than the filter approaches. The model-specific approaches (wrapper and embedded) for feature selection map

the data to a new subspace that is suitable for the particular model to be learned.

2.2.2 Subspace Learning

Subspace learning is an approach for learning a new lower-dimensional representation of a high-dimensional space [49] [50]. In subspace learning for classification, the aim is to transform the features from a given space into a lower-dimensional space optimized for better classification accuracy. Linear Discriminant Analysis (LDA) is an example of a typical subspace learning algorithm for classification, which aims at minimizing the within-class scatter of the data and maximizing the between-class scatter by optimizing the following Fisher-Rao's criterion:

$$\mathbf{Q}^* = \arg\min \frac{\mathrm{Tr}(\mathbf{Q}\mathbf{S}_w\mathbf{Q}^\intercal)}{\mathrm{Tr}(\mathbf{Q}\mathbf{S}_b\mathbf{Q}^\intercal)}, \tag{2.14}$$

where $\mathbf{Q}$ is the projection matrix and $\mathrm{Tr}(\cdot)$ is the trace operator. The within-class scatter matrix $\mathbf{S}_w$ and between-class scatter matrix $\mathbf{S}_b$ are defined as follows:

$$\begin{aligned}
\mathbf{S}_w &= \sum_{c=1}^{\mathscr{C}}\sum_{i=1}^{N_c}(\mathbf{x}_{ci}-\mu_c)(\mathbf{x}_{ci}-\mu_c)^\intercal = \sum_{c=1}^{\mathscr{C}}\sum_{i=1}^{N_c}\left(\mathbf{x}_{ci}\mathbf{x}_{ci}^\intercal - 2\mathbf{x}_{ci}\mu_c^\intercal + \mu_c\mu_c^\intercal\right) \\
&= \sum_{c=1}^{\mathscr{C}}\left(\sum_{i=1}^{N_c}(\mathbf{x}_{ci}\mathbf{x}_{ci}^\intercal) - N_c\mu_c\mu_c^\intercal\right) = \sum_{c=1}^{\mathscr{C}}\left(\mathbf{X}\mathrm{diag}(\mathbf{1}_c)\mathbf{X}^\intercal - \frac{1}{N_c}\mathbf{X}\mathbf{1}_c\mathbf{1}_c^\intercal\mathbf{X}^\intercal\right) \\
&= \mathbf{X}\left(\mathbf{I} - \sum_{c=1}^{\mathscr{C}}\frac{1}{N_c}\mathbf{1_c}\mathbf{1_c}^\intercal\right)\mathbf{X}^\intercal = \mathbf{X}\mathbf{L}_w\mathbf{X}^\intercal,
\end{aligned} \tag{2.15}$$

$$\begin{aligned}
\mathbf{S}_b &= \sum_{c=1}^{\mathscr{C}}N_c(\mu_c-\mu)(\mu_c-\mu)^\intercal = \sum_{c=1}^{\mathscr{C}}N_c\left(\mu_c\mu_c^\intercal - 2\mu_c\mu^\intercal + \mu\mu^\intercal\right) \\
&= \sum_{c=1}^{\mathscr{C}}(N_c\mu_c\mu_c^\intercal) - N\mu\mu^\intercal = \mathbf{X}\left(\sum_{c=1}^{\mathscr{C}}\frac{1}{N_c}\mathbf{1_c}\mathbf{1_c}^\intercal - \frac{1}{N}\mathbf{1}\mathbf{1}\right)\mathbf{X}^\intercal = \mathbf{X}\mathbf{L}_b\mathbf{X}^\intercal,
\end{aligned} \tag{2.16}$$

where $\mathscr{C}$ and N denote the total number of classes and samples respectively, N_c represents the total number of samples belonging to class c, $\mathbf{x}_{ci}$ is the ith sample of class c, μ denotes the mean of the data, and μ_c represents the mean of class c. $\mathbf{I}$ denotes the identity matrix, and $\mathbf{1}$ represents a vector containing all values as ones. The vector $\mathbf{1}_c$ has ones at elements corresponding to instances that belong to class c and zeros elsewhere.

The trace ratio problem in (2.14) does not have a closed-form solution, but it is commonly approximated with the corresponding ratio trace problem [51]:

$$\mathbf{Q}^* = \arg\min \ \mathrm{Tr}\big((\mathbf{Q}\mathbf{S}_b\mathbf{Q}^\intercal)^{-1}\mathbf{Q}\mathbf{S}_w\mathbf{Q}^\intercal\big), \tag{2.17}$$

which can be efficiently solved via solving the generalized eigenvalue problem

$$\mathbf{S}_w\mathbf{q} = \upsilon\mathbf{S}_b\mathbf{q}, \tag{2.18}$$

where υ denotes an eigenvalue and $\mathbf{q}$ denotes an eigenvector. The projection matrix $\mathbf{Q}$ is formed by keeping the eigenvectors corresponding to the d smallest non-zero eigenvalues as its rows.

The sum of the within-class scatter and the between-class scatter is equal to the total-class scatter, i.e., $\mathbf{S}_t = \mathbf{S}_w + \mathbf{S}_b$. Thus, LDA can be also solved via an equivalent optimization problem

$$\mathbf{Q}^* = \arg\min \ \frac{\mathrm{Tr}(\mathbf{Q}\mathbf{S}_w\mathbf{Q}^\intercal)}{\mathrm{Tr}(\mathbf{Q}\mathbf{S}_t\mathbf{Q}^\intercal)}, \tag{2.19}$$

where

$$\begin{aligned}\mathbf{S}_t &= \sum_{c=1}^{\mathscr{C}}\sum_{i=1}^{N_c}(\mathbf{x}_{ci}-\mu)(\mathbf{x}_{ci}-\mu)^\intercal = \sum_{c=1}^{\mathscr{C}}\sum_{i=1}^{N_c}\big(\mathbf{x}_{ci}\mathbf{x}_{ci}^\intercal - 2\mathbf{x}_{ci}\mu^\intercal + \mu\mu^\intercal\big)\\ &= \sum_{c=1}^{\mathscr{C}}\sum_{i=1}^{N_c}(\mathbf{x}_{ci}\mathbf{x}_{ci}^\intercal) - N\mu\mu^\intercal = \mathbf{X}\left(\mathbf{I}-\frac{1}{N}\mathbf{1}\mathbf{1}^\intercal\right)\mathbf{X}^\intercal.\end{aligned} \tag{2.20}$$

Note that for centered data $\mathbf{S}_t = \mathbf{X}\mathbf{X}^\intercal$.

LDA is an example of supervised learning [52]. In supervised learning, the transformation is learned by leveraging the class labels. LDA followed by random projections also helps random projections to yield a more discriminative subspace [53]. The advantage of random projection mainly includes approximate distance preservation [54]. Random projections, however, do not exploit any information from the data inherently and are not helpful for joint optimization of the model. Other categories of subspace learning methods are unsupervised and semi-supervised methods. In unsupervised methods, the transformation is carried out without the use of class labels. For example, Principal Component Analysis (PCA) finds a linear combina-

tion of the features to find a new representation of the data without exploiting the label information. PCA has been extended to Kernel-PCA using the kernel method and used in different applications. In [55], Kernel-PCA is performed using a triangular kernel for shape description.

The third category of subspace learning is semi-supervised dimensionality reduction which finds a lower-dimensional space from a combination of both labeled and unlabeled data. Supervised methods usually perform better than unsupervised methods if sufficient data are available [46, 56, 57].

OCC techniques are usually studied and developed for features living in their original feature space or higher-dimensional kernel space. On the other hand, dimensionality reduction has been an active research area primarily for challenges with data available from all categories. In this thesis, we mainly focus on developing subspace learning techniques for OCC. Moreover, the relationship and similarities of the traditional subspace learning methods with the proposed subspace learning methods for OCC are analyzed.

2.3 Graph Embedding

Graph embedding is an approach for converting high-dimensional graphs into low-dimensional vectors while maximally preserving the properties of the graph. Traditionally, graph embedding algorithms are associated with two research problems [58]: graph analytics [59] and representation learning [60]. In graph analytics, the aim is to extract useful information from a graph to find out the strength and direction of the relationships between objects represented by the graph. In representation learning, data representations are obtained for modeling the data in a lower-dimensional space. The general framework of graph embedding for dimensionality reduction was proposed in [61].

Let $\mathcal{G} = \{\mathbf{X}, \mathbf{A}\}$ represent an undirected weighted graph with data points in $\mathbf{X}$ representing the graph nodes and the similarity matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ representing the relations between these data points. The Laplacian matrix $\mathbf{L}$ of the graph can be defined as

$$\mathbf{L} = \mathbf{D} - \mathbf{A}, \;\; [\mathbf{D}]_{ii} = \sum_{j \neq i} [\mathbf{A}]_{ij}, \forall i \in \{1, \ldots, N\}, \tag{2.21}$$

where $\mathbf{D}$ is a diagonal matrix. Graph embedding [61] was proposed as a general framework for various subspace learning methods under the graph preserving criterion

$$\begin{aligned}\mathbf{Q}^* &= \underset{\mathrm{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L}_\mathbf{p}\mathbf{X}^\intercal\mathbf{Q}^\intercal)=m}{\arg\min} \sum_{i\neq j}(\mathbf{Q}\mathbf{x}_i-\mathbf{Q}\mathbf{x}_j)^2\mathbf{A}_{ij} \\ &= \underset{\mathrm{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L}_p\mathbf{X}^\intercal\mathbf{Q}^\intercal)=m}{\arg\min} \mathrm{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L}\mathbf{X}^\intercal\mathbf{Q}^\intercal), \\ &= \arg\min \frac{\mathrm{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L}\mathbf{X}^\intercal\mathbf{Q}^\intercal)}{\mathrm{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L}_p\mathbf{X}^\intercal\mathbf{Q}^\intercal)},\end{aligned} \tag{2.22}$$

where $\mathbf{L}$ and $\mathbf{L}_p$ denote the Laplacians of *intrinsic* and *penalty* graphs, respectively. The intrinsic graph Laplacians correspond to data relations to be preserved while the penalty graph Laplacians correspond to data relations to be penalized. As explained in Section 2.2.2, the solution to (2.22) can be approximated by solving the generalized eigenvalue problem

$$\mathbf{X}\mathbf{L}\mathbf{X}^\intercal\mathbf{q} = \upsilon\mathbf{X}\mathbf{L}_p\mathbf{X}^\intercal\mathbf{q}. \tag{2.23}$$

Different formulations of the intrinsic and penalty graph lead to different subspace learning algorithms. As derived in (2.20), (2.15), and (2.16), the Laplacians for commonly used total-scatter, within-class, and between-classes matrices can be given as

$$\mathbf{L}_t = \mathbf{I} - \frac{1}{N}, \tag{2.24}$$

$$\mathbf{L}_w = \mathbf{I} - \sum_{c=1}^{\mathscr{C}} \frac{1}{N_c}\mathbf{1}_c\mathbf{1}_c^\intercal, \tag{2.25}$$

$$\mathbf{L}_b = \sum_{c=1}^{\mathscr{C}} N_c(\frac{1}{N_c}\mathbf{1}_c - \frac{1}{N}\mathbf{1})(\frac{1}{N_c}\mathbf{1}_c - \frac{1}{N}\mathbf{1})^\intercal. \tag{2.26}$$

2.3.1 Spectral Regression

Spectral regression [62] provides an alternative way to solve (2.23). It was introduced to speed up the linear discriminant analysis [63]. Recently in [64], it has been used for speeding up subclass discriminant analysis. Let $\mathbf{t}$ be the eigenvector of the fol-

lowing problem

$$\mathbf{L}\mathbf{t} = \upsilon \mathbf{L}_p \mathbf{t}, \tag{2.27}$$

where υ is the corresponding eigenvalue. If $\mathbf{X}^\intercal \mathbf{q} = \mathbf{t}$, then $\mathbf{q}$ is the eigenvector of the generalized eigen-decomposition problem in (2.23) with the same eigenvalue. This is because $\mathbf{X}\mathbf{L}\mathbf{X}^\intercal \mathbf{q} = \mathbf{X}\mathbf{L}\mathbf{t} = \upsilon \mathbf{X}\mathbf{L}_p \mathbf{t} = \upsilon \mathbf{X}\mathbf{L}_p \mathbf{X}^\intercal \mathbf{q}$.

In order to estimate the projection matrix $\mathbf{Q}$, first the vector $\mathbf{t}$ is obtained from (2.27), then the vector $\mathbf{q}$ is estimated so that $\mathbf{X}^\intercal \mathbf{q} = \mathbf{t}$ is satisfied. The solution is estimated using regularized least squares:

$$\begin{aligned} \mathbf{q} &= \arg\min\left(||\mathbf{X}^\intercal \mathbf{q} - \mathbf{t}||^2 + \eta ||\mathbf{q}||^2 \right) \\ &= (\mathbf{X}\mathbf{X}^\intercal + \epsilon \mathbf{I})^{-1} \mathbf{X}\mathbf{t}. \end{aligned} \tag{2.28}$$

Because the graph Laplacian matrices used in our work are sparse by construction, spectral regression bypasses the need to compute the eigendecomposition of dense matrices.

2.3.2 Graph-Embedded Support Vector Data Description

Most of the proposed graph embedding algorithms for classification assume that data from all categories is available during training. However, in reality, this assumption may not be true. In [20], a graph-embedded approach for OCC was proposed. Graph-Embedded Support Vector Data Description (GESVDD) is an extension of SVDD, which incorporates generic graph structures in the optimization process of SVDD. The generic graph structures express geometric data relationships of interest in the optimization process of SVDD. Exploiting such information in the optimization process of SVDD improves the classification performance. GESVDD solves the following optimization problem:

$$\begin{aligned} \min \quad & R^2 + C \sum_{i=1}^{N} \xi_i \\ \text{s.t.} \quad & (\mathbf{x}_i - \mathbf{a})^\intercal \mathbf{S}_x^{-1} (\mathbf{x}_i - \mathbf{a}) \leq R^2 + \xi_i, \\ & \xi_i \geq 0, \forall i \in \{1, \ldots, N\}, \end{aligned} \tag{2.29}$$

where the matrix $\mathbf{S}_x = \mathbf{X}\mathbf{L}_x\mathbf{X}^\intercal$ encodes the geometric data relationships and $\mathbf{L}_x$ represents the graph Laplacian expressing the geometric data relationship of any graph. The Lagrangian of GESVDD is

$$L = \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^\intercal \mathbf{S}_x^{-1} \mathbf{x}_i - \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \mathbf{x}_i^\intercal \mathbf{S}_x^{-1} \mathbf{x}_j \alpha_j. \tag{2.30}$$

The optimization problem of GESVDD is equivalent to SVDD in transformed feature space. In [20], much effort has been put into optimizing the data description by incorporating generic graph structures in the feature or kernel space, but no effort has been made to combine such graph structures with projecting the data into an optimized subspace. We address this issue by optimizing a subspace projection along with incorporating geometric data relationships of interest in the optimization process of GESVDD in this thesis.

2.4 Non-linear Transformations

Non-linear transformations are used to transform data from its original input space non-linearly to another higher-dimensional feature space (*kernel space*) to achieve more flexible decision boundaries for classification. The data representations $\mathbf{x}_i \in \mathbb{R}^D$, $i = 1, \ldots, N$ are mapped to the kernel space $\mathscr{F}$ using a non-linear function $\phi(\cdot)$, such that $\mathbf{x}_i \in \mathbb{R}^D \rightarrow \phi(\mathbf{x}_i) \in \mathscr{F}$. Figure 2.2 depicts the obtained boundaries before and after the non-linear transformation for SVDD where the hypersphere obtained in the kernel space has a more flexible boundary for the data in the original feature space.

2.4.1 Kernel Trick

Non-linear transformations are used for transforming data to a higher-dimensional space to be linearly separable. However, the computational cost of optimizing an objective function in a higher-dimensional space increases drastically with the increased dimensions. Often the kernel space is infinite-dimensional and thus, direct computations in the kernel space are impossible. The kernel trick allows a representation of the data in higher-dimensional space without explicitly mapping it to

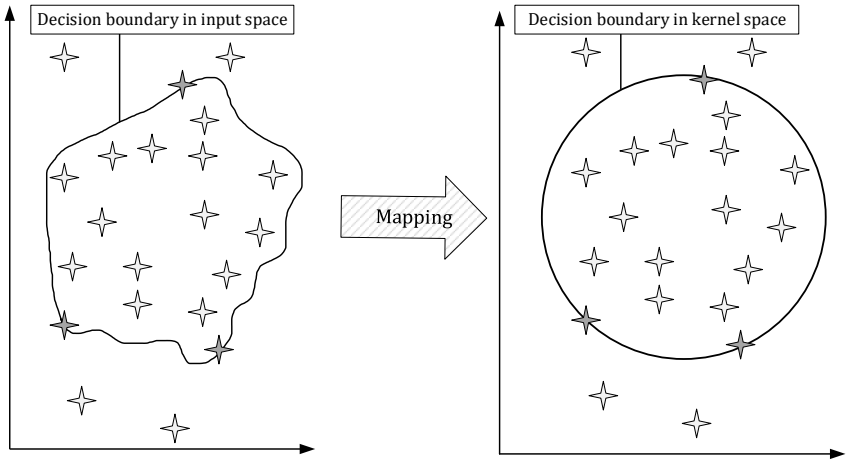

Figure 2.2 Depiction of non-linear transformation in case of SVDD.

the higher-dimensions [65]. It was introduced for obtaining a non-linear decision boundary for Support Vector Machines (SVMs) and has been used successfully for many different algorithms. It can be applied to any linear method that involves the dot product in the optimization process. Hence, a function corresponding to an inner product in higher-dimensional feature space can be used as a kernel function. One of the popular kernel functions is the Radial Basis Function (RBF) kernel:

$$[\mathbf{K}_{ij}] = \exp\left(\frac{-||\mathbf{x}_i - \mathbf{x}_j||_2^2}{2\sigma^2}\right), \tag{2.31}$$

where $[\mathbf{K}_{ij}] = \phi(\mathbf{x}_i)^\intercal \phi(\mathbf{x}_j)$ and σ is a hyperparameter for determining the width of the RBF kernel. We employed the kernel-trick to derive non-linear versions of some of the proposed methods in this thesis. One of the limitations of the kernel trick is that it requires the problem formulation to contain only dot products. For example, it cannot be applied to Ellipsoidal Support Vector Data Description (ESVDD) [66] formulation as ESVDD includes outer products rather than inner products. An alternative to the kernel trick is the Non-Linear Projection Trick (NPT) [67] discussed in the next section.

2.4.2 Non-linear Projection Trick

NPT is an alternative to the kernel trick developed to extend the applicability of the kernel methods to algorithms that do not use dot products in the optimization process. It is applied to transform the data at the beginning of the process, while the subsequent optimization follows the linear method. In order to apply NPT over data points, the kernel matrix is first computed using a kernel function such as 2.31 and then centralized as

$$\hat{\mathbf{K}} = (\mathbf{I} - \mathbf{J})\mathbf{K}(\mathbf{I} - \mathbf{J}), \tag{2.32}$$

where $\mathbf{I} \in \mathbb{R}^{N \times N}$ is an identity matrix, and $\mathbf{J} \in \mathbb{R}^{N \times N}$ is computed as

$$\mathbf{J} = \frac{1}{N}\mathbf{1}\mathbf{1}^{\intercal}. \tag{2.33}$$

$\mathbf{1}$ represents a vector containing all values as ones. The centered kernel matrix is then decomposed by using eigendecomposition:

$$\hat{\mathbf{K}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\intercal}, \tag{2.34}$$

where $\mathbf{\Lambda}$ is a diagonal matrix containing the non-negative eigenvalues in its diagonal. The columns of $\mathbf{U}$ contain the corresponding eigenvectors. The data in the reduced dimensional kernel space is obtained as

$$\mathbf{\Phi} = \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}^{\intercal}. \tag{2.35}$$

$\mathbf{\Phi}$ can be used instead of the training data matrix $\mathbf{X}$ in the original linear method to obtain a non-linear transformation. We exploited the NPT to derive non-linear versions of the proposed methods in this thesis.

2.4.3 Test Data Transformation

In order to apply a non-linear transformation for a test instance $\mathbf{x}_*$ in the standard kernel trick case, the kernel vector is first computed as

$$\mathbf{k}_* = \mathbf{\Phi}^\intercal \phi(\mathbf{x}_*). \tag{2.36}$$

and then the method is applied. In the case of applying NPT over test instances, the kernel vector $\mathbf{k}_*$ is first computed and then centered as

$$\hat{\mathbf{k}}_* = (\mathbf{I} - \mathbf{J})[\mathbf{k}_* - \frac{1}{N}\mathbf{K}\mathbf{1}]. \tag{2.37}$$

The centered kernel vector is then mapped to

$$\phi_* = (\mathbf{\Phi}^\intercal)^+ \hat{\mathbf{k}}_*, \tag{2.38}$$

which is used in testing following the original linear approach. We apply the test data transformation in the experiments during the testing phase of the trained models for all the proposed methods in this thesis.

3 CONTRIBUTIONS

This thesis proposes several methods that share the main idea of mapping data from a high-dimensional space to a low-dimensional space optimized for SVDD. Figure 3.1 depicts the idea with the corresponding α_i values for the class description.

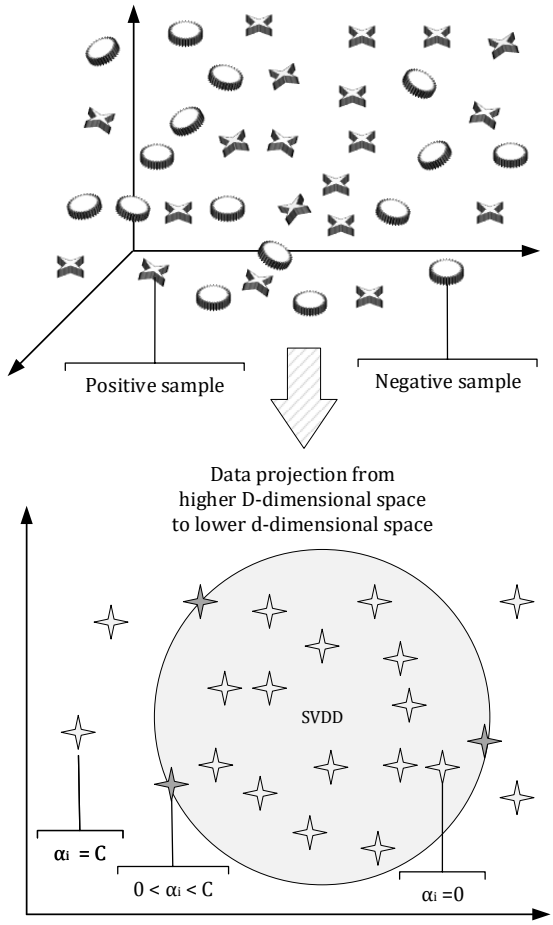

Figure 3.1 Depiction of data mapping from a higher-dimensional space to a lower-dimensional space optimized for One-Class Classification [P4].

This chapter gives a detailed description of the contributions of this thesis. It begins with Subspace Support Vector Data Description (SSVDD), which is the basis of this thesis. The chapter continues by introducing ESSVDD, which is a more generalized extension of SSVDD. Afterward, this chapter describes Multimodal Subspace Support Vector Data Description (MSSVDD) as an extension of SSVDD from a unimodal data description to a multimodal data description in a common shared subspace for OCC. Next, this chapter presents Graph-Embedded Subspace Support Vector Data Description (GESSVDD), which is a framework that encodes different geometric data relationships in the subspace and generalizes the subspace learning techniques for SVDD. Additionally, this chapter discusses the proposed framework for rare benthic macroinvertebrates identification as an example of using SSVDD for complementing CNN-based taxa identification.

3.1 Subspace Support Vector Data Description

We propose a novel method SSVDD for defining a data description along with data mapping to low-dimensional feature space optimized for OCC. Let $\mathbf{x}_i$, $i = 1,\ldots,N$, be the data points living in a D-dimensional feature space (i.e., $\mathbf{x}_i \in \mathbb{R}^D$). The aim is to find a projection $\mathbf{Q} \in \mathbb{R}^{d\times D}$ for transforming the data to a lower ($d \leq D$) dimensional space to better encapsulate it inside a hypersphere. The data in the lower d-dimensional space is represented as

$$\mathbf{y}_i = \mathbf{Q}\mathbf{x}_i,\ i = 1,\ldots,N. \tag{3.1}$$

In order to obtain a data description, a hypersphere enclosing the data needs to be determined. The objective is to minimize the radius of the hypersphere so that all data is inside the hypersphere:

$$\begin{aligned} &\min \quad F(R,\mathbf{a}) = R^2 \\ \text{s.t.} \quad &||\mathbf{Q}\mathbf{x}_i - \mathbf{a}||_2^2 \leq R^2,\ i = 1,\ldots,N, \end{aligned} \tag{3.2}$$

where R is the radius of the hypersphere and $\mathbf{a} \in \mathbb{R}^d$ is the center of the hypersphere. In order to address the possibility of outliers in the training set, a set of slack variables

ξ_i is introduced. The relaxed version of the above optimization problem becomes

$$\begin{aligned} \min \quad & F(R,\mathbf{a}) = R^2 + C\sum_{i=1}^{N}\xi_i \\ \text{s.t.} \quad & ||\mathbf{Q}\mathbf{x}_i - \mathbf{a}||_2^2 \leq R^2 + \xi_i, \\ & \xi_i \geq 0,\ \forall i \in \{1,\ldots,N\}, \end{aligned} \tag{3.3}$$

where $C > 0$ is a hyperparameter that controls the number of outliers in the data description. The optimization problem (3.3) corresponds to the optimization problem of SVDD with an additional variable $\mathbf{Q}$. The constraints are incorporated into the objective function by using Lagrange multipliers

$$L = R^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\gamma_i\xi_i - \sum_{i=1}^{N}\alpha_i\Big(R^2 + \xi_i - \mathbf{x}_i^\intercal\mathbf{Q}^\intercal\mathbf{Q}\mathbf{x}_i + 2\mathbf{a}^\intercal\mathbf{Q}\mathbf{x}_i - \mathbf{a}^\intercal\mathbf{a}\Big), \tag{3.4}$$

where $\alpha_i \geq 0$ and $\gamma_i \geq 0$ are Lagrange multipliers. The saddle points are calculated as follows

$$\frac{\partial L}{\partial R} = 0 \quad \Rightarrow \quad \sum_{i=1}^{N}\alpha_i = 1, \tag{3.5}$$

$$\frac{\partial L}{\partial \mathbf{a}} = 0 \quad \Rightarrow \quad \mathbf{a} = \sum_{i=1}^{N}\alpha_i\mathbf{Q}\mathbf{x}_i, \tag{3.6}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \quad \Rightarrow \quad C - \alpha_i - \gamma_i = 0, \tag{3.7}$$

$$\frac{\partial L}{\partial \mathbf{Q}} = 0 \quad \Rightarrow \quad \mathbf{Q} = \left(\sum_{i=1}^{N}\alpha_i\mathbf{x}_i\mathbf{x}_i^\intercal\right)^{-1}\left(\sum_{i=1}^{N}\alpha_i\mathbf{x}_i\mathbf{a}^\intercal\right). \tag{3.8}$$

We notice that α values and $\mathbf{Q}$ are interconnected and cannot be optimized simultaneously. Therefore, we follow an iterative process where we fix one variable at a time and optimize the other. The two steps of the iterative process, class description optimization and subspace optimization, are introduced in the following subsections.

3.1.1 Class Description Optimization

Class description optimization, i.e., finding optimal α_i values in the subspace, can be solved by fixing the projection matrix $\mathbf{Q}$ and following the standard SVDD ap-

proach. After substituting (3.1), (3.5), (3.6) and (3.7) in (3.4) we obtain

$$L = \sum_{i=1}^{N} \alpha_i \mathbf{y}_i^\intercal \mathbf{y}_i - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \mathbf{y}_i^\intercal \mathbf{y}_j \alpha_j. \tag{3.9}$$

The expression in (3.9) is equivalent to (2.8) for the data in the subspace. Hence, maximizing (3.9) will give us α values defining the position of training instances in the subspace.

3.1.2 Subspace Optimization

The second step of the iterative process is to optimize the projection matrix $\mathbf{Q}$ for the determined set of α. To this end, we optimize the Lagrangian with an additional regularization term ψ:

$$L = \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^\intercal \mathbf{Q}^\intercal \mathbf{Q} \mathbf{x}_i - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \mathbf{x}_i^\intercal \mathbf{Q}^\intercal \mathbf{Q} \mathbf{x}_j \alpha_j + \beta \psi, \tag{3.10}$$

where β is used to control the importance of the regularization term. ψ expresses the class variance in the d-dimensional space as

$$\psi = \mathrm{Tr}(\mathbf{Q}\mathbf{X}\lambda\lambda^\intercal \mathbf{X}^\intercal \mathbf{Q}^\intercal). \tag{3.11}$$

We propose four different forms for the regularization term by changing the values in vector λ as follows:

- In the first strategy, we set $\lambda_i = 0, i = 1, \ldots, N$, which causes the regularization term to become obsolete. In other words, the regularization term is not used at all in this setting.
- In the second strategy, the λ vector is replaced by a vector of all ones, i.e., λ=**1**. This leads to selecting all the samples in the regularization term.
- In the third strategy, the λ vector is replaced by a vector containing the α values of the corresponding instances. Thus, this strategy discards all the data points which are inside the boundary from the regularization term.
- In the fourth strategy, the λ_i values are replaced by the α_i values for the samples belonging to the boundary and by zeros for all other samples. Hence, we select only the support vectors in the regularization term.

The above mentioned regularization strategies are denoted in [P1] by Ψ_1, Ψ_2, Ψ_3, and Ψ_4, respectively.

We update the projection matrix $\mathbf{Q}$ by using the gradient of (3.10):

$$\mathbf{Q} \leftarrow \mathbf{Q} - \eta \Delta L, \tag{3.12}$$

where η is a learning rate parameter used as a step of the gradient. The gradient ΔL is

$$\Delta L = 2\sum_{i=1}^{N} \alpha_i \mathbf{Q}\mathbf{x}_i\mathbf{x}_i^\intercal - 2\sum_{i=1}^{N}\sum_{j=1}^{N} \mathbf{Q}\mathbf{x}_i\mathbf{x}_j^\intercal \alpha_i\alpha_j + \beta\Delta\psi \tag{3.13}$$

and the derivative of regularization term ψ is calculated with respect to $\mathbf{Q}$ as follows

$$\Delta\psi = 2\mathbf{Q}\mathbf{X}\lambda\lambda^\intercal\mathbf{X}^\intercal. \tag{3.14}$$

The gradient ΔL can be further simplified and written as

$$\Delta L = 2\mathbf{Q}\mathbf{X}(\mathbb{A} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal)\mathbf{X} + \beta\Delta\psi, \tag{3.15}$$

where the matrix $\mathbb{A} \in \mathbb{R}^{N\times N}$ is a diagonal matrix containing α_i values in its diagonal, $\boldsymbol{\alpha}$ is a vector of α_i values.

3.1.3 Testing Phase

During testing, an instance $\mathbf{x}_* \in \mathbb{R}^D$ is first projected to d-dimensional space using the optimized $\mathbf{Q}$ matrix as $\mathbf{y}_* = \mathbf{Q}\mathbf{x}_* \in \mathbb{R}^d$ and its distance from the center of the hypersphere is calculated:

$$||\mathbf{y}_* - \mathbf{a}||_2^2 = \mathbf{y}_*^\intercal\mathbf{y}_* - 2\sum_{i=1}^{N}\alpha_i\mathbf{y}_*^\intercal\mathbf{y}_i + \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\mathbf{y}_i^\intercal\mathbf{y}_j. \tag{3.16}$$

The instance is accepted as a positive instance when $||\mathbf{y}_* - \mathbf{a}||_2^2 \leq R^2$ and as negative otherwise. The threshold R^2 is calculated as follows

$$R^2 = (\mathbf{Q}\mathbf{s})^\intercal\mathbf{Q}\mathbf{s} - 2(\mathbf{Q}\mathbf{s})^\intercal\mathbf{a} + \mathbf{a}^\intercal\mathbf{a}, \tag{3.17}$$

where $\mathbf{s}$ is any support vector.

3.1.4 Non-linear Subspace Support Vector Data Description

In [P1], we also proposed a kernel-based non-linear data description of SSVDD. We followed a standard kernel-based learning approach where the data is mapped to the kernel space $\mathscr{F}$ using a non-linear function $\phi(\cdot)$, such that $\mathbf{x}_i \in \mathbb{R}^D \rightarrow \phi(\mathbf{x}_i) \in \mathscr{F}$. In the kernel space, a linear projection to a subspace is given by:

$$\mathbf{y}_i = \mathbf{Q}\phi(\mathbf{x}_i),\ i = 1,\ldots,N. \tag{3.18}$$

We employed the kernel trick for calculating the data representations $\mathbf{y}_i$, $i = 1,\ldots,N$. $\mathbf{Q}$ can be expressed as a linear combination of the data in the kernel space $\mathscr{F}$ as $\mathbf{W}\mathbf{\Phi}^\intercal$ and, thus,

$$\mathbf{y}_i = \mathbf{W}\mathbf{\Phi}^\intercal\phi(\mathbf{x}_i) = \mathbf{W}\mathbf{k}_i,\ i = 1,\ldots,N, \tag{3.19}$$

where $\mathbf{\Phi} \in \mathbb{R}^{|\mathscr{F}|\times N}$ represents training data in the kernel space, $\mathbf{W} \in \mathbb{R}^{d\times N}$ denotes a matrix that contain the weights of $\mathbf{\Phi}$ which are required for constructing the projection matrix, and $\mathbf{k}_i$ represents the i-th column of the kernel matrix $\mathbf{K} \in \mathbb{R}^{N\times N}$ having elements equal to $\mathbf{K}_{ij} = \phi(\mathbf{x}_i)^\intercal\phi(\mathbf{x}_j)$. We used the RBF kernel in our experiments (See Section 2.4.1).

In the kernel SSVDD, the iterative process becomes as follows: For a given $\mathbf{W}$, the data $\mathbf{x}_i$, $i = 1,\ldots,N$ is transformed to subspace $\mathbf{y}_i$, $i = 1,\ldots,N$ using (3.19) and α_i, $i = 1,\ldots,N$ are calculated by optimizing (3.9). Afterwards, $\mathbf{W}$ is updated using the gradient-based approach with

$$\Delta L = 2\sum_{i=1}^{N}\alpha_i\mathbf{W}\mathbf{k}_i\mathbf{k}_i^\intercal - 2\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{W}\mathbf{k}_i\mathbf{k}_j^\intercal\alpha_i\alpha_j + \beta\Delta\Psi, \tag{3.20}$$

where

$$\Delta\Psi = 2\mathbf{W}\mathbf{K}\lambda\lambda^\intercal\mathbf{K}^\intercal. \tag{3.21}$$

An alternative non-linear data description of SSVDD can be obtained by using NPT. In NPT-based non-linear data description, a non-linear mapping is applied at the beginning, while the optimization follows the linear SSVDD process. The steps required for obtaining a non-linear mapping in the case of NPT are described in Section 2.4.2.

3.1.5 Overall SSVDD Algorithm

The overall SSVDD method is summarized in Algorithm 1.

Algorithm 1: SSVDD optimization

Input : $\mathbf{X}$, // Input data
β, // Regularization parameter for controlling significance of ψ
η, // Learning rate parameter
d, // Dimensionality of subspace
C, // Regularization parameter in SVDD
λ // Used for selecting different forms of ψ

Output: $\mathbf{Q}$ // Projection matrix
R, // Radius of hypersphere
α // Defines the data description

Random initialization of $\mathbf{Q}$;
Orthogonalize $\mathbf{Q}$ using QR decomposition;
Row normalize $\mathbf{Q}$ using l_2 norm;

for $iter = 1 : max\_iter$ **do**

Project data to subspace $\mathbf{y}_i = \mathbf{Q}\mathbf{x}_i$;
Calculate α_i, $i = 1,\ldots,N$ by maximizing
$L = \sum_{i=1}^{N} \alpha_i \mathbf{y}_i^\intercal \mathbf{y}_i - \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \mathbf{y}_i^\intercal \mathbf{y}_j \alpha_j$;
Calculate $\Delta L = 2\mathbf{Q}\mathbf{X}(\mathbb{A} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal)\mathbf{X} + 2\beta\mathbf{Q}\mathbf{X}\boldsymbol{\lambda}\boldsymbol{\lambda}^\intercal\mathbf{X}^\intercal$;
Update $\mathbf{Q} \leftarrow \mathbf{Q} - \eta\Delta L$;
Orthogonalize $\mathbf{Q}$ using QR decomposition;
Row normalize $\mathbf{Q}$ using l_2 norm;

end

Project data to subspace $\mathbf{y}_i = \mathbf{Q}\mathbf{x}_i$;
Calculate α_i, $i = 1,\ldots,N$ by maximizing
$L = \sum_{i=1}^{N} \alpha_i \mathbf{y}_i^\intercal \mathbf{y}_i - \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \mathbf{y}_i^\intercal \mathbf{y}_j \alpha_j$;
Identify any support vector $\mathbf{s}$ having $0 < \alpha_s < C$;
Compute radius $R = \sqrt{(\mathbf{Q}\mathbf{s})^\intercal\mathbf{Q}\mathbf{s} - 2(\mathbf{Q}\mathbf{s})^\intercal\mathbf{a} + \mathbf{a}^\intercal\mathbf{a}}$;

We impose an additional constraint $\mathbf{QQ}^\intercal = \mathbf{I}$ to obtain an orthogonal projection; hence we orthogonalized and normalized the projection matrix $\mathbf{Q}$ at each iteration. To obtain an optimized projection matrix along with the data description, we iteratively apply the two-step process described in Section 3.1.1 and Section 3.1.2. Algorithm 1 depicts the process for linear SSVDD. However, similar steps are followed for the kernel-based and NPT-based non-linear SSVDD.

3.1.6 Complexity Analysis of SSVDD

The main steps of linear SSVDD are: 1) Projecting data to a lower d-dimensional space, 2) Solving SVDD to obtain α values, 3) Computing the gradient ΔL, 4) Updating the projection matrix $\mathbf{Q}$ and 5) QR decomposition. All of these steps are analyzed and the overall complexity of a single iteration of SSVDD is computed as follows:

1. The complexity of projecting data to d-dimensional space is the complexity of multiplying two $d \times D$ and $D \times N$ matrices. This has a complexity of $\mathcal{O}(dDN)$.
2. SVDD has the complexity of $\mathcal{O}(N^3)$ for N data points [68].
3. The gradient ΔL is computed using (3.15). The first term in (3.15) has the complexity of $\mathcal{O}(dDN + dDN^2 + DN^2)$. The second term in (3.15) is the regularization term calculated by (3.14), which has a complexity of $\mathcal{O}(2dDN + 2dN)$. By adding these terms the complexity of computing the gradient ΔL becomes $\mathcal{O}(dDN + dDN^2 + DN^2 + 2dDN + 2dN)$.
4. The complexity of updating $\mathbf{Q}$ is $\mathcal{O}(dD)$.
5. The complexity of QR decomposition is $\mathcal{O}(dD^2)$ [69].

Now adding the complexities of the above steps and dropping the relatively lower computational steps, the complexity becomes $\mathcal{O}(N^3 + (d+1)DN^2)$. Assuming that the number of samples N is greater than the dimensionality D, i.e., $D << N$, the complexity of the linear SSVDD becomes $\mathcal{O}(N^3)$. For the non-linear version with NPT, the matrix $\mathbf{K}$ is computed, centralized, and decomposed via eigendecomposition. These three steps have the combined complexity of $\mathcal{O}(DN^2 + 2N^3) = \mathcal{O}(N^3)$. In the non-linear case, the dimensionality changes from D to N for all correspond-

ing steps. Hence, the total complexity of each iteration becomes $\mathcal{O}(3N^3) = \mathcal{O}(N^3)$. The overall complexity also stays the same for the non-linear version with NPT.

3.1.7 Results and Discussion

We experimented with 14 different standard datasets downloaded from UCI [70] and Delft repository [71]. A summary of the datasets is provided in Table 3.1. The datasets were divided into train and test sets by considering 70% of the data as training data and the remaining 30% as test data. The 70-30% splits were selected randomly by keeping the representation of each class similar to the original dataset. The 70-30% splits selection was carried out five times. Hence, we performed five experiments for each dataset.

Table 3.1 List of datasets used in experiments of SSVDD [P1].

No.	Dataset name	N	D	Target class
1	Balance scale	625	4	Left
2	Iris	150	4	Iris-virginica
3	Lenses	24	4	No contact lenses
4	Seeds	210	7	Kama
5	Haberman's survival	306	3	Survived
6	Qualitative bankruptcy	250	7	Bankrupt
7	User knowledge modeling	403	5	Low
8	Pima Indians diabetes	768	8	No diabetes
9	Banknote authentication	1372	5	No
10	TA evaluation	151	5	High
11	PDelft pump	1500	64	Normal
12	Vehicle Opel	864	18	Opel
13	Sonar	208	60	Mines
14	Breast Wisconsin	699	9	Malignant

We used the *F1 measure* for evaluating the performance of SSVDD. It takes into account both Precision (Pre) and True Positive Rate (TPR) and provides a single metric that weights the two crucial metrics in a balanced way. The F1 measure is

defined as the harmonic mean of TPR and Pre: i.e.,

$$F1 = 2 * \frac{\text{Pre} * \text{TPR}}{\text{Pre} + \text{TPR}}, \tag{3.22}$$

where Pre is defined as the proportion of True Positives (TPs) out of all positive predictions (TPs+False Positives (FPs)):

$$\text{Pre} = \frac{\text{TP}s}{\text{TP}s + \text{FP}s}, \tag{3.23}$$

TPR is true positive rate which is also known as sensitivity, recall or hit rate. It is the proportion of correctly classified positive samples. We selected the hyperparameters using 5-fold cross-validation according to the best F1 measure.

The results of SSVDD are compared with the original SVDD (linear and kernel), OCSVM (linear and kernel), Graph-Embedded Support Vector Machine (GESVM) and GESVDD. The results are provided in Table 3.2.

Table 3.2 F1 results for linear and non-linear data description over 14 datasets [P1].

Dataset	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Linear														
SVDD	0.703	0.762	0.609	0.774	0.834	0.686	0.634	0.791	0.764	0.485	0.846	0.853	0.625	0.958
OCSVM	0.688	0.612	0.394	0.619	0.644	0.562	0.532	0.529	0.657	**0.532**	0.632	0.590	0.535	0.660
SSVDD Ψ_1	**0.907**	**0.899**	0.620	0.756	0.836	0.692	**0.960**	0.786	**0.908**	0.482	0.856	**0.855**	0.618	0.957
SSVDD Ψ_2	0.898	0.897	**0.724**	**0.827**	0.839	0.720	0.957	**0.793**	0.889	0.502	0.857	**0.855**	0.599	**0.960**
SSVDD Ψ_3	0.896	0.881	0.649	0.798	**0.841**	**0.722**	0.946	0.787	0.886	0.507	0.856	**0.855**	0.633	**0.960**
SSVDD Ψ_4	0.896	0.868	0.694	0.778	0.821	0.715	0.954	0.784	0.852	0.458	**0.857**	0.854	**0.638**	0.953
Non-linear														
SVDD	0.734	0.827	0.413	**0.858**	0.835	0.605	0.651	0.785	0.804	0.396	0.836	0.852	0.609	0.962
OCSVM	0.544	0.673	0.523	0.444	0.743	0.550	0.409	0.786	0.700	0.274	0.661	0.679	0.530	0.630
GESVDD	0.757	0.857	0.314	0.799	0.811	0.554	0.654	**0.790**	0.797	0.484	0.830	0.847	0.550	**0.966**
GESVM	**0.815**	**0.869**	0.398	0.800	0.816	0.594	**0.658**	0.667	**0.930**	**0.498**	0.613	0.788	0.593	0.962
SSVDD Ψ_1	0.635	0.725	**0.736**	0.727	0.842	0.700	0.518	0.786	0.728	0.472	0.836	**0.858**	0.504	0.961
SSVDD Ψ_2	0.662	0.573	0.603	0.540	**0.845**	**0.762**	0.523	**0.790**	0.717	0.473	**0.856**	**0.858**	**0.637**	0.783
SSVDD Ψ_3	0.734	0.694	0.624	0.719	0.838	0.620	0.578	0.785	0.720	0.417	**0.856**	**0.858**	**0.637**	0.902
SSVDD Ψ_4	0.495	0.700	**0.736**	0.774	0.841	0.632	0.562	0.572	0.703	0.474	0.832	**0.858**	**0.637**	0.951

The F1 measure for all experiments is reported for the linear and the kernel versions of the proposed method along with the competing methods. The test results for SSVDD are reported after 100 training iterations. It is observed that SSVDD

outperforms all other linear methods in the majority of cases. In non-linear comparisons, SSVDD outperforms other competing methods on 6 out of 14 datasets. It is evident from the obtained results that a data description for OCC in an optimized subspace improves the classification performance during inference compared to the data description in the original feature space. Moreover, it was noticed that in both linear and non-linear SSVDD, using all samples for describing the data variance in the regularization term achieves the best performance in terms of F1 measure most often. The code relevant to SSVDD is available online[1].

The proposed SSVDD forms the basis of this thesis. The obtained results validate the idea that incorporating the model information in the optimization process of subspace learning for OCC indeed improves the OCC accuracy and the initial objective is achieved. However, the hyperspherical nature of the closed boundary can be too rigid to characterize the data in the subspace or may require many iterations to obtain an optimized description. We investigate the use of a more flexible boundary in the next section. Moreover, the proposed SSVDD is suited for one-class unimodal data description only; we extend the method to multimodal data description in Section 3.3.

3.2 Ellipsoidal Subspace Support Vector Data Description

The previous section introduced SSVDD, which maps data to a subspace optimized for encapsulating the data in a hypersphere. However, the shape of the hypersphere can result in superfluous regions that do not contain any instances from the target class. The generalization of the hyperspherical boundary is needed for better encapsulation of the target class data in OCC. In this section, we generalize SSVDD by using a hyperellipsoid instead of a hypersphere for encapsulating the target class data. The shape of an ellipsoidal boundary is less conservative than a sphere. The proposed ESSVDD method takes into account the covariance of the data in the subspace and yields a more flexible boundary which is a better fit for OCC.

[1] https://github.com/fahadsohrab/ssvdd

The optimization objective for ESSVDD is formulated as

$$\begin{aligned} \min \quad & F(R,\mathbf{a}) = R^2 + C\sum_{i=1}^{N}\xi_i \\ \text{s.t.} \quad & (\mathbf{Q}\mathbf{x}_i - \mathbf{a})^\intercal \mathbf{E}^{-1}(\mathbf{Q}\mathbf{x}_i - \mathbf{a}) \le R^2 + \xi_i, \\ & \xi_i \ge 0, \forall i \in \{1,\ldots,N\}, \end{aligned} \tag{3.24}$$

where $\mathbf{a}$ denotes the center of a hyperellipsoid and the inverse of the covariance matrix $\mathbf{E}$, also referred to as the concentration matrix, is defined as

$$\mathbf{E}^{-1} = (\mathbf{Q}\mathbf{X}\mathbf{X}^\intercal\mathbf{Q}^\intercal)^{-1}. \tag{3.25}$$

Defining a new vector $\mathbf{u} = \mathbf{E}^{-\frac{1}{2}}\mathbf{a}$, (3.24) can be written as

$$\begin{aligned} \min \quad & F(R,\mathbf{u}) = R^2 + C\sum_{i=1}^{N}\xi_i \\ \text{s.t.} \quad & ||\mathbf{E}^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i - \mathbf{u}||_2^2 \le R^2 + \xi_i, \\ & \xi_i \ge 0, \forall i \in \{1,\ldots,N\}. \end{aligned} \tag{3.26}$$

By using Lagrange multipliers, the constraints are incorporated into the corresponding objective function as

$$\begin{aligned} L = R^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\big(R^2 + \xi_i \\ -(\mathbf{E}^{-\frac{1}{2}}\mathbf{y}_i)^\intercal\mathbf{E}^{-\frac{1}{2}}\mathbf{y}_i + 2\mathbf{u}^\intercal\mathbf{E}^{-\frac{1}{2}}\mathbf{y}_i - \mathbf{u}^\intercal\mathbf{u}\big) - \sum_{i=1}^{N}\gamma_i\xi_i, \end{aligned} \tag{3.27}$$

where $\alpha_i \ge 0$ and $\gamma_i \ge 0$ are Lagrange multipliers.

Using partial derivatives with respect to R, $\mathbf{u}$ and ξ_i, and setting them to zero,

we get

$$\frac{\partial L}{\partial R} = 0 \quad \Rightarrow \quad \sum_{i=1}^{N} \alpha_i = 1 \tag{3.28}$$

$$\frac{\partial L}{\partial \mathbf{u}} = 0 \quad \Rightarrow \quad \mathbf{u} = \sum_{i=1}^{N} \alpha_i \mathbf{E}^{-\frac{1}{2}} \mathbf{Q} \mathbf{x}_i \tag{3.29}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \quad \Rightarrow \quad C - \alpha_i - \gamma_i = 0. \tag{3.30}$$

By substituting (3.28)-(3.30) into (3.27), we get

$$L = \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^{\intercal} \mathbf{Q}^{\intercal} \mathbf{E}^{-1} \mathbf{Q} \mathbf{x}_i - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \mathbf{x}_i^{\intercal} \mathbf{Q}^{\intercal} \mathbf{E}^{-1} \mathbf{Q} \mathbf{x}_j \alpha_j. \tag{3.31}$$

This corresponds to the standard SVDD over $\mathbf{y}_i$ after mapping the data to the subspace as

$$\mathbf{y}_i = \mathbf{E}^{-\frac{1}{2}} \mathbf{Q} \mathbf{x}_i, \ , i = 1, \ldots, N. \tag{3.32}$$

Following the same iterative process as for SSVDD, ESSVDD is solved by alternating between solving the class description from (3.31) and optimizing the projection matrix from a regularized version of (3.31) given as

$$L = \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^{\intercal} \mathbf{Q}^{\intercal} (\mathbf{Q} \mathbf{X} \mathbf{X}^{\intercal} \mathbf{Q}^{\intercal})^{-1} \mathbf{Q} \mathbf{x}_i -$$
$$\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \mathbf{x}_i^{\intercal} \mathbf{Q}^{\intercal} (\mathbf{Q} \mathbf{X} \mathbf{X}^{\intercal} \mathbf{Q}^{\intercal})^{-1} \mathbf{Q} \mathbf{x}_j \alpha_j + \beta \Upsilon. \tag{3.33}$$

The regularization term Υ expresses the class variance in the lower d-dimensional space also taking into account the concentration matrix:

$$\Upsilon = \mathrm{Tr}(\mathbf{E}^{-\frac{1}{2}} \mathbf{Q} \mathbf{X} \lambda \lambda^{\intercal} \mathbf{X}^{\intercal} \mathbf{Q}^{\intercal} \mathbf{E}^{-\frac{\intercal}{2}}). \tag{3.34}$$

Analogous to the regularization strategies proposed for ψ in SSVDD, we propose different forms of λ in the regularization term Υ as well. In the first form, λ is replaced by a vector of all ones, i.e., λ=**1**. Hence, all samples are used to describe the covariance of the class. In the second form, λ vector is replaced by a vector containing the α values of the corresponding instances. Thus, this strategy selects the

data samples belonging to the boundary and outside the boundary to describe class covariance. In the third form, λ_i values for the support vectors are replaced by the corresponding α_i values and by zeros for all other instances. These regularization terms are denoted as Υ_1, Υ_2, and Υ_3, respectively. Moreover, it is noticed that by replacing the covariance matrix $\mathbf{E}$ in (3.34) with the $d \times d$ identity matrix $\mathbf{I}$, the regularization term Υ becomes equivalent to the regularization term ψ defined in (3.11). In our experiments, we also consider the first, second, and third forms of λ for the regularization term ψ as well, which are denoted by ψ_1, ψ_2 and ψ_3, respectively. We denote the case where no regularization term is used in ESSVDD as $\psi_0\Upsilon_0$.

The Lagrangian of ESSVDD can be simplified further and written as

$$L = \mathrm{Tr}((\mathbf{Q}\mathbf{X}\mathbf{X}^\intercal\mathbf{Q}^\intercal)^{-1}\mathbf{Q}\mathbf{X}(\mathbb{A}-\boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal)\mathbf{X}^\intercal\mathbf{Q}^\intercal) + \beta\Upsilon, \tag{3.35}$$

where the matrix $\mathbb{A} \in \mathbb{R}^{N\times N}$ contains α_i values in its diagonal. The projection matrix is updated by using the gradient of (3.35), which can be computed as

$$\Delta L = 2\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}(\mathbb{A}-\boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal)\mathbf{X}^\intercal - 2\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}(\mathbb{A}-\boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal)\mathbf{X}^\intercal\mathbf{Q}^\intercal\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}\mathbf{X}^\intercal + \beta\Delta\Upsilon, \tag{3.36}$$

where

$$\Delta\Upsilon = 2\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}\boldsymbol{\lambda}\boldsymbol{\lambda}^\intercal\mathbf{X}^\intercal - 2\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}\boldsymbol{\lambda}\boldsymbol{\lambda}^\intercal\mathbf{X}^\intercal\mathbf{Q}^\intercal\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}\mathbf{X}^\intercal. \tag{3.37}$$

We optimize the projection matrix using the gradient-based approach. We also orthogonalize and normalize the projection matrix during the iterative process. Algorithm 2 summarizes the optimization steps for ESSVDD.

3.2.1 Test Phase

In the test phase of the linear ESSVDD, a test instance denoted as $\mathbf{x}_*$ is first projected to the optimized d-dimensional subspace as

$$\mathbf{y}_* = \mathbf{Q}\mathbf{x}_*. \tag{3.38}$$

The test instance is assigned a target or outlier label based on its distance from the center of the data description. The distance of the test instance from the center is

Algorithm 2: ESSVDD optimization

Input : $\mathbf{X}$, // Input data
β, // Regularization parameter for controling significance of ψ
η, // Learning rate parameter
d, // Dimensionality of subspace
C, // Regularization parameter in SVDD
λ // Used for selecting different forms of ψ
Output: $\mathbf{Q}$ // Projection matrix
R, // Radius of hypersphere
α // Defines the data description

Random initialization of $\mathbf{Q}$;

for $iter = 1 : max\_iter$ **do**
  Compute concentration matrix $\mathbf{E}^{-1} = (\mathbf{Q}\mathbf{X}\mathbf{X}^\intercal\mathbf{Q}^\intercal)^{-1}$;
  Calculate α_i, $i = 1,\ldots,N$ by maximizing
  $L = \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^\intercal \mathbf{Q}^\intercal \mathbf{E}^{-1} \mathbf{Q}\mathbf{x}_i - \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \mathbf{x}_i^\intercal \mathbf{Q}^\intercal \mathbf{E}^{-1} \mathbf{Q}\mathbf{x}_j \alpha_j$;
  Calculate $\Delta\Upsilon = 2\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}\lambda\lambda^\intercal\mathbf{X}^\intercal - 2\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}\lambda\lambda^\intercal\mathbf{X}^\intercal\mathbf{Q}^\intercal\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}\mathbf{X}^\intercal$;
  Calculate $\Delta L =$
  $2\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}(\mathbb{A} - \alpha\alpha^\intercal)\mathbf{X}^\intercal - 2\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}(\mathbb{A} - \alpha\alpha^\intercal)\mathbf{X}^\intercal\mathbf{Q}^\intercal\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}\mathbf{X}^\intercal + \beta\Delta\Upsilon$;
  Update $\mathbf{Q} \leftarrow \mathbf{Q} - \eta\Delta L$;
  Orthogonalize $\mathbf{Q}$ using QR decomposition;
  Row normalize $\mathbf{Q}$ using l_2 norm;
end
Compute concentration matrix $\mathbf{E}^{-1} = (\mathbf{Q}\mathbf{X}\mathbf{X}^\intercal\mathbf{Q}^\intercal)^{-1}$;
Calculate α_i, $i = 1,\ldots,N$ by maximizing
$L = \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^\intercal \mathbf{Q}^\intercal \mathbf{E}^{-1} \mathbf{Q}\mathbf{x}_i - \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \mathbf{x}_i^\intercal \mathbf{Q}^\intercal \mathbf{E}^{-1} \mathbf{Q}\mathbf{x}_j \alpha_j$;
Compute center of data description in the subspace as $\mathbf{u} = \sum_{i=1}^{N} \alpha_i \mathbf{E}^{-\frac{1}{2}} \mathbf{Q}\mathbf{x}_i$;
Identify any support vector $\mathbf{s}$ having $0 < \alpha_s < C$;
Compute radius $R = \sqrt{(\mathbf{E}^{-\frac{1}{2}}\mathbf{Q}\mathbf{s})^\intercal \mathbf{E}^{-\frac{1}{2}}\mathbf{Q}\mathbf{s} - 2(\mathbf{E}^{-\frac{1}{2}}\mathbf{Q}\mathbf{s})^\intercal\mathbf{u} + \mathbf{u}^\intercal\mathbf{u}}$;

calculated as follows:

$$||\mathbf{E}^{-\frac{1}{2}}\mathbf{y}_* - \mathbf{u}||_2^2 = (\mathbf{E}^{-\frac{1}{2}}\mathbf{y}_*)^\intercal \mathbf{E}^{-\frac{1}{2}}\mathbf{y}_* - 2(\mathbf{E}^{-\frac{1}{2}}\mathbf{y}_*)^\intercal\mathbf{u} + \mathbf{u}^\intercal\mathbf{u}, \tag{3.39}$$

where $\mathbf{u}$ is calculated as in (3.29). The test instance is assigned to the positive class if $||\mathbf{E}^{-\frac{1}{2}}\mathbf{y}_* - \mathbf{u}||_2^2 \leq R^2$. If $||\mathbf{E}^{-\frac{1}{2}}\mathbf{y}_* - \mathbf{u}||_2^2 > R^2$, the test instance is classified as an outlier.

The threshold R^2 is calculated as

$$R^2 = (\mathbf{E}^{-\frac{1}{2}}\mathbf{Qs})^\intercal \mathbf{E}^{-\frac{1}{2}}\mathbf{Qs} - 2(\mathbf{E}^{-\frac{1}{2}}\mathbf{Qs})^\intercal \mathbf{u} + \mathbf{u}^\intercal \mathbf{u}, \tag{3.40}$$

where $\mathbf{s}$ is any support vector.

3.2.2 Complexity Analysis of ESSVDD

The complexity of ESSVDD is similar to that of SSVDD with the additional computation of the concentration matrix and its square root. The complexity of determining the concentration matrix involves the multiplication of $\mathbf{QXX}^\intercal\mathbf{Q}^\intercal$ and then finding its inverse. These two steps together have a complexity of $\mathcal{O}(dDN + N^3)$. Computing the square root of the concentration matrix has a complexity of $\mathcal{O}(N^3)$ as well. Adding these terms to the total complexity of SSVDD is equivalent to the complexity of ESSVDD. In terms of big $\mathcal{O}$ notation, the complexity of ESSVDD stays at $\mathcal{O}(N^3)$. In the case of a non-linear data description based on NPT, the complexity of ESSVDD remains at $\mathcal{O}(N^3)$.

3.2.3 Results and Discussion

We evaluated the performance of ESSVDD and competing methods over 6 different datasets downloaded from UCI repository [70]. All the datasets were converted to one-class datasets by considering a single class as the target class and the rest of the data as outliers. For example, in the Seeds(Kama) dataset, we selected the Kama category as the target class while selecting the Rosa and Canadian categories together as the outlier class. Similarly, in Seeds(Rosa), the Rosa category was selected as the target class and the Kama and Canadian categories as the outlier class. A similar conversion of the datasets into one-class form was carried out for all datasets used in the experiments. The one-class datasets were divided into training and test sets by selecting 70% of the data for training and the remaining 30% for testing. The train-test splits were selected randomly by keeping the proportions of each class similar to the original dataset. All experiments were repeated five times using different random train/test splits. A summary of the datasets is provided in Table 3.3. The original dimensionality of each dataset in the original feature space is given in the the last column (**D**) of Table 3.3.

Table 3.3 Datasets used in the experiments [P2].

Abbreviation	Dataset Name (Target Class)	Total Samples	Target Samples	D
S-K	Seeds (Kama)	210	70	7
S-R	Seeds (Rosa)	210	70	7
S-C	Seeds (Canadian)	210	70	7
QB-B	Qualitative bankruptcy (bankruptcy)	250	107	6
QB-N	Qualitative bankruptcy (non-bankruptcy)	250	143	6
SH-H	Somerville happiness (happy)	143	77	6
SH-U	Somerville happiness (un-happy)	143	66	6
I-S	Iris (Setosa)	150	50	4
I-VC	Iris (Versicolor)	150	50	4
I-V	Iris (Virginica)	150	50	4
IS-B	Ionosphere (bad)	351	126	34
IS-G	Ionosphere (good)	351	225	34
SR-R	Sonar (rock)	208	97	60
SR-M	Sonar (mines)	208	111	60

We used Geometric Mean (GM) as the evaluation metric because it takes into account both TPR and True Negative Rate (TNR). GM is defined as the square root of the product of TPR and TNR:

$$\mathrm{GM} = \sqrt{\mathrm{TPR} \times \mathrm{TNR}}. \tag{3.41}$$

The GM results for the linear and non-linear methods are reported in tables 3.4 and 3.5, respectively. The test results for all iterative methods are reported after 10 training iterations. We used 5-fold cross-validation over the training set for selecting the appropriate hyperparameters for the final training.

ESSVDD generalizes SSVDD by using ellipsoidal data description instead of hyperspherical data description. It is observed that when compared to other border-based methods, the proposed ESSVDD performed better in terms of GM than other competing methods over all the datasets, except the iris dataset in the linear data description. In the non-linear cases, ESSVDD performs best on average over half of the datasets used in the experiments. The average *Gmean* value is also calculated over 5 test splits for the different datasets on each iteration. Figure 3.2 shows the performance of ESSVDD and SSVDD on the test set after every training iteration. It can be seen that for both the linear and non-linear methods, ESSVDD converges faster than SSVDD to an optimal solution.

The main benefit of ESSVDD compared to SVDD is the more flexible data description. In SVDD, the data must always be transformed to a hyperspherical form

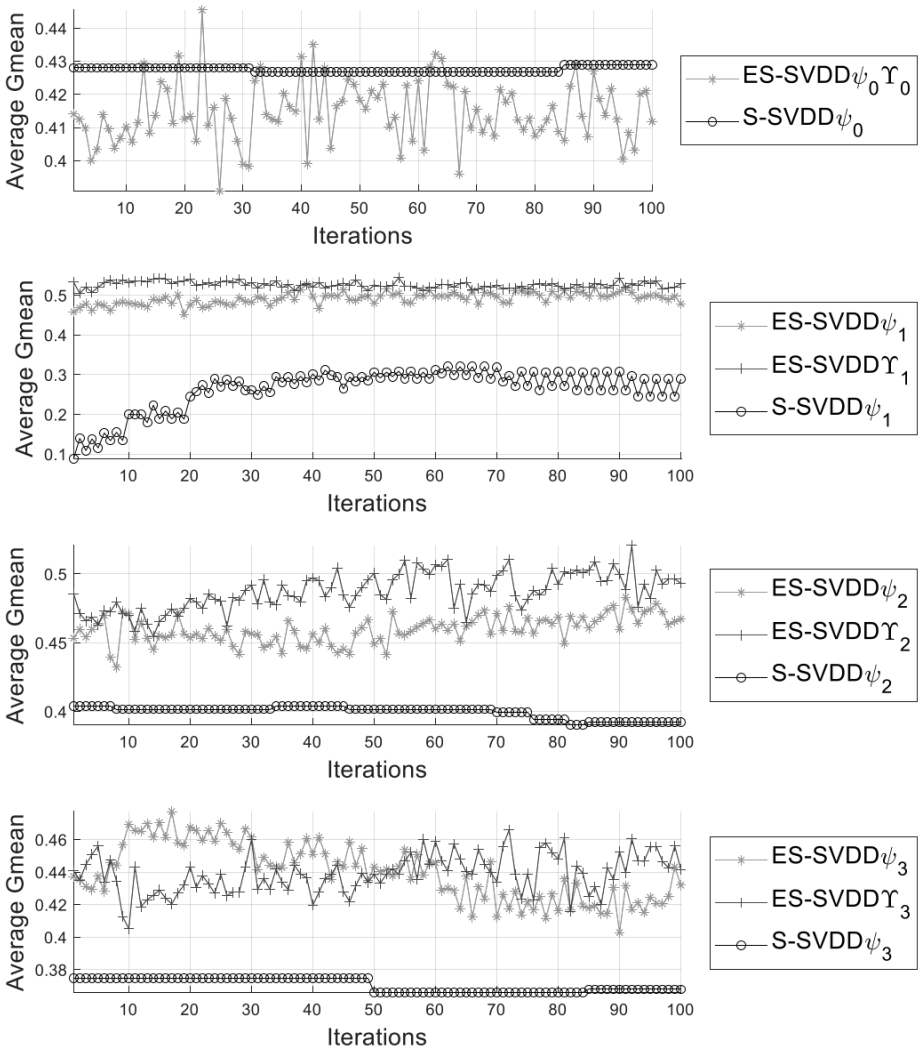

Figure 3.2 Comparison of different regularization terms for non-linear ESSVDD and SSVDD on dataset Sonar (rock) [P2].

for optimal solutions, but ESSVDD can produce good results with less regular data distributions, which can lead to faster convergence of the iterative optimization process. In fact, we observed that if we initialize the projection vector $\mathbf{Q}$ with PCA, there is no significant difference between SVDD and ESSVDD, but ESSVDD converges much faster with randomly initialized $\mathbf{Q}$. This illustrates both the benefit of the hyperellipsoidal data description and motivates the use of PCA initialization if

a hyperspherical data description is used. Similar to SSVDD, ESSVDD operates on unimodal data. In the next section, we turn our attention to multimodal data.

More detailed analysis and observations are provided in the corresponding publication [P2]. The ESSVDD code is publicly available online[2].

Table 3.4 GM results for **linear** methods over different datasets [P2].

Dataset	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.
ESSVDD $\psi_0\Upsilon_0$	0.83	0.91	0.77	0.84	0.84	0.26	0.55	0.41	0.51	0.46
ESSVDD ψ_1	0.83	0.89	0.89	0.87	0.76	0.46	**0.61**	0.47	0.47	0.47
ESSVDD ψ_2	0.82	0.89	0.87	0.86	0.85	0.18	0.51	0.53	0.51	0.52
ESSVDD ψ_3	0.79	0.90	0.87	0.86	0.90	0.23	0.57	0.46	0.35	0.40
ESSVDD Υ_1	0.82	0.92	0.90	0.88	0.85	0.32	0.58	0.46	0.52	0.49
ESSVDD Υ_2	0.84	0.91	**0.91**	**0.89**	0.81	0.30	0.56	**0.55**	**0.53**	**0.54**
ESSVDD Υ_3	**0.85**	0.88	0.88	0.87	0.87	0.33	0.60	0.49	0.47	0.48
SSVDD ψ_0	0.79	0.86	0.81	0.82	0.72	0.50	**0.61**	0.49	0.48	0.48
SSVDD ψ_1	0.72	0.76	0.77	0.75	0.85	0.34	0.59	0.46	0.46	0.46
SSVDD ψ_2	0.81	0.82	0.77	0.80	0.75	0.40	0.58	0.47	0.48	0.48
SSVDD ψ_3	0.80	**0.93**	0.75	0.83	0.72	0.41	0.57	0.49	0.46	0.48
SVDD	0.82	0.92	0.86	0.87	0.83	0.04	0.43	0.54	0.48	0.51
ESVDD	0.80	0.87	0.86	0.84	**0.96**	0.20	0.58	0.54	0.41	0.48
OCSVM	0.43	0.46	0.58	0.49	0.46	**0.55**	0.51	0.45	0.42	0.44
Non-support-vector-based methods										
K-means	**0.86**	**0.94**	**0.91**	**0.90**	0.71	0.41	0.56	0.56	0.39	0.47
Parzen	0.49	0.33	0.58	0.47	**0.98**	**0.60**	**0.79**	**0.58**	**0.43**	**0.50**

Dataset	I-S	I-VC	I-V	Av.	IS-B	IS-G	Av.	SR-R	SR-M	Av.
ESSVDD $\psi_0\Upsilon_0$	0.64	0.75	0.70	0.70	0.16	0.89	0.52	**0.50**	0.64	0.57
ESSVDD ψ_1	0.92	0.86	0.77	0.85	**0.52**	0.85	**0.69**	**0.50**	0.56	0.53
ESSVDD ψ_2	0.87	0.82	0.79	0.83	0.31	0.87	0.59	0.48	**0.67**	**0.58**
ESSVDD ψ_3	0.93	0.87	0.71	0.84	0.35	0.89	0.62	0.48	0.65	0.57
ESSVDD Υ_1	0.85	0.84	0.86	0.85	0.26	0.87	0.57	0.47	**0.67**	0.57
ESSVDD Υ_2	**0.96**	0.83	0.74	0.84	0.31	0.89	0.60	0.47	0.65	0.56
ESSVDD Υ_3	0.80	0.85	0.79	0.81	0.35	**0.90**	0.62	0.49	0.65	0.57
SSVDD ψ_0	0.87	0.75	0.64	0.76	0.16	0.75	0.46	0.37	0.37	0.37
SSVDD ψ_1	0.88	0.81	0.75	0.81	0.50	0.71	0.61	0.44	0.36	0.40
SSVDD ψ_2	0.87	0.84	0.58	0.76	0.43	0.72	0.58	0.46	0.40	0.43
SSVDD ψ_3	0.81	0.68	0.63	0.71	0.27	0.66	0.46	0.46	0.41	0.43
SVDD	0.94	**0.90**	**0.89**	**0.91**	0.04	0.73	0.39	**0.50**	0.52	0.51
ESVDD	0.89	0.88	0.86	0.88	0.33	0.00	0.17	0.00	0.00	0.00
OCSVM	0.50	0.52	0.39	0.47	0.47	0.45	0.46	0.44	0.52	0.48
Non-support-vector-based methods										
K-means	**0.94**	**0.92**	**0.89**	**0.91**	**0.37**	**0.88**	**0.63**	**0.49**	**0.68**	**0.58**
Parzen	0.85	0.68	0.79	0.77	0.32	0.25	0.28	0.00	0.00	0.00

[2]https://github.com/fahadsohrab/essvdd

Table 3.5 GM results for **non-linear** methods over different datasets [P2].

Dataset	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.
ESSVDD $\psi_0\Upsilon_0$	0.78	0.88	**0.93**	0.87	0.83	0.61	0.72	0.52	0.42	0.47
ESSVDD ψ_1	0.80	0.88	0.88	0.85	0.80	0.34	0.57	0.51	0.42	0.47
ESSVDD ψ_2	0.80	0.90	**0.93**	0.87	0.90	0.35	0.62	0.52	0.33	0.42
ESSVDD ψ_3	0.82	0.86	0.72	0.80	0.89	0.64	**0.76**	0.52	0.45	0.49
ESSVDD Υ_1	0.85	**0.92**	0.91	**0.89**	0.87	0.47	0.67	0.47	0.38	0.43
ESSVDD Υ_2	0.82	0.88	0.89	0.86	0.84	**0.68**	**0.76**	0.52	0.34	0.43
ESSVDD Υ_3	0.85	0.88	0.91	0.88	0.87	0.54	0.71	0.52	0.45	0.49
SSVDD ψ_0	0.74	0.74	0.83	0.77	0.23	0.49	0.36	0.45	0.29	0.37
SSVDD ψ_1	0.71	0.78	0.81	0.77	0.11	0.08	0.09	0.39	0.32	0.35
SSVDD ψ_2	0.72	0.85	0.83	0.80	0.36	0.37	0.37	0.47	0.32	0.40
SSVDD ψ_3	0.60	0.76	0.76	0.71	0.36	0.40	0.38	0.46	0.29	0.37
SVDD	**0.86**	0.91	0.88	0.88	0.88	0.55	0.71	0.54	**0.48**	0.51
ESVDD	0.84	0.85	0.85	0.85	**0.96**	0.51	0.74	0.55	0.42	0.48
GESVDD	0.84	0.90	0.76	0.83	0.94	0.17	0.55	0.54	0.47	0.50
OCSVM	0.79	0.60	0.63	0.67	0.67	0.52	0.59	**0.57**	**0.48**	**0.52**
GESVM	0.83	0.88	0.89	0.87	0.88	0.58	0.73	**0.57**	0.42	0.50
Non-support-vector-based methods										
SOM	0.80	0.90	0.89	0.86	0.79	0.37	0.58	0.28	0.26	0.27
Gaussian	**0.85**	**0.95**	**0.94**	**0.91**	0.63	0.46	0.54	**0.52**	**0.42**	**0.47**
GE-OC-ELM	**0.85**	0.93	0.89	0.89	**1.00**	**0.80**	**0.90**	0.31	0.31	0.31

Dataset	I-S	I-VC	I-V	Av.	IS-B	IS-G	Av.	SR-R	SR-M	Av.
ESSVDD $\psi_0\Upsilon_0$	0.93	0.84	**0.86**	0.88	0.44	0.89	0.67	0.41	0.67	0.54
ESSVDD ψ_1	**0.94**	0.81	0.74	0.83	**0.71**	0.90	**0.80**	0.48	0.55	0.51
ESSVDD ψ_2	0.91	0.87	0.83	0.87	0.31	0.87	0.59	0.47	0.66	0.56
ESSVDD ψ_3	0.89	0.84	0.74	0.82	0.32	0.88	0.60	0.47	0.66	0.57
ESSVDD Υ_1	0.81	0.89	0.70	0.80	0.47	0.86	0.67	0.53	0.65	0.59
ESSVDD Υ_2	0.91	0.83	0.81	0.85	0.68	0.86	0.77	0.47	**0.70**	0.58
ESSVDD Υ_3	**0.94**	0.88	0.83	0.88	0.45	0.85	0.65	0.41	**0.70**	0.55
SSVDD ψ_0	0.92	0.85	0.78	0.85	0.24	0.53	0.38	0.43	0.41	0.42
SSVDD ψ_1	0.89	0.89	0.63	0.80	0.68	0.64	0.66	0.20	0.48	0.34
SSVDD ψ_2	0.91	0.84	0.77	0.84	0.21	0.61	0.41	0.40	0.52	0.46
SSVDD ψ_3	0.92	0.85	0.73	0.83	0.35	0.62	0.49	0.37	0.16	0.27
SVDD	**0.94**	**0.91**	0.84	**0.89**	0.31	0.80	0.55	0.53	0.66	0.59
ESVDD	0.89	0.84	**0.86**	0.86	0.30	0.00	0.15	0.00	0.00	0.00
GESVDD	0.91	0.88	0.85	0.88	0.26	0.81	0.54	**0.56**	0.66	**0.61**
OCSVM	0.45	0.65	0.66	0.59	0.27	0.63	0.45	0.51	0.58	0.54
GESVM	0.92	0.90	**0.86**	**0.89**	0.39	**0.91**	0.65	0.54	0.67	**0.61**
Non-support-vector-based methods										
SOM	0.91	0.84	**0.88**	**0.88**	0.06	**0.87**	0.47	0.46	0.32	0.39
Gaussian	0.97	0.86	0.80	**0.88**	0.33	0.50	0.42	**0.47**	**0.59**	**0.53**
GE-OC-ELM	**0.99**	**0.89**	0.78	**0.88**	**0.48**	0.81	**0.65**	0.38	0.55	0.47

3.3 Multimodal Subspace Support Vector Data Description

MSSVDD is a novel technique for transforming multimodal data into a shared subspace. In MSSVDD, we optimize the subspace shared by data coming from different modalities for OCC. Different regularization and decision strategies along with linear and non-linear formulations are also proposed.

Let the number of total modalities be M and the data in each modality m be represented by $\mathbf{X}_m = [\mathbf{x}_{m,1}, \mathbf{x}_{m,2}, \dots \mathbf{x}_{m,N}]$, $\mathbf{x}_{m,i} \in \mathbb{R}^{D_m}$, where D_m is the data dimension in modality m and N is the total number of instances in each modality. The proposed MSSVDD tries to find a projection matrix $\mathbf{Q}_m \in \mathbb{R}^{d \times D_m}$ for all modalities to transform the data from the corresponding modalities to a shared lower d-dimensional subspace optimized for SVDD. Figure 3.3 depicts the basic idea of MSSVDD.

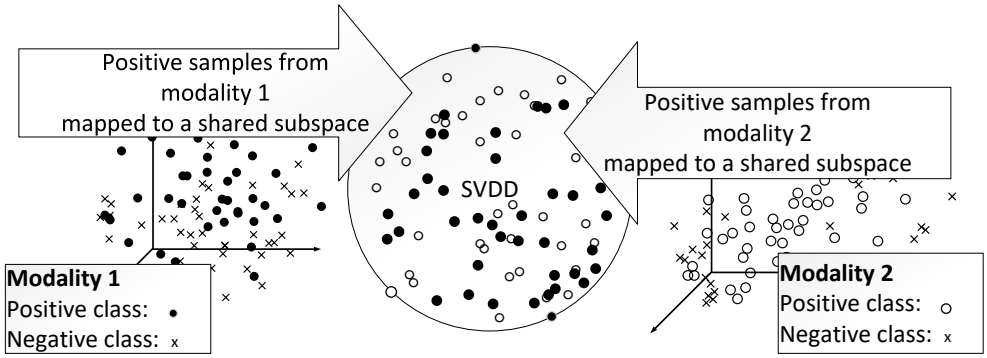

Figure 3.3 Depiction of Multimodal Subspace Support Vector Data Description.

The data is projected to a common shared subspace via the corresponding projection matrix as

$$\mathbf{y}_{m,i} = \mathbf{Q}_m \mathbf{x}_{m,i}, \forall m \in \{1, \dots, M\} \ , \forall i \in \{1, \dots, N\}. \tag{3.42}$$

The criterion to minimize is as follows:

$$\begin{aligned} \min \quad & F(R, \mathbf{a}) = R^2 + C \sum_{m=1}^{M} \sum_{i=1}^{N} \xi_{m,i} \\ \text{s.t.} \quad & ||\mathbf{Q}_m \mathbf{x}_{m,i} - \mathbf{a}||_2^2 \leq R^2 + \xi_{m,i}, \\ & \xi_{m,i} \geq 0, \forall m \in \{1, \dots, M\}, \forall i \in \{1, \dots, N\}. \end{aligned} \tag{3.43}$$

The Lagrange function can be given as

$$L = R^2 + C\sum_{m=1}^{M}\sum_{i=1}^{N}\xi_{m,i} - \sum_{m=1}^{M}\sum_{i=1}^{N}\gamma_{m,i}\xi_{m,i} - \sum_{m=1}^{M}\sum_{i=1}^{N}\alpha_{m,i}\Big(R^2 + \xi_{m,i} - \mathbf{x}_{m,i}^{\intercal}\mathbf{Q}_m^{\intercal}\mathbf{Q}_m\mathbf{x}_{m,i} + 2\mathbf{a}^{\intercal}\mathbf{Q}_m\mathbf{x}_{m,i} - \mathbf{a}^{\intercal}\mathbf{a}\Big). \quad (3.44)$$

In order to obtain a tight boundary around the target class data in a shared subspace, we use SVDD. Analogous to the SVDD optimization process, the Lagrangian function should be maximized with respect to the Lagrange multipliers $\alpha_{m,i} \geq 0$ and $\gamma_{m,i} \geq 0$, and minimized with respect to the other variables R, $\mathbf{a}$, $\xi_{m,i}$, and $\mathbf{Q}_m$. By setting the partial derivatives to zero, we get

$$\frac{\partial L}{\partial R} = 0 \quad \Rightarrow \quad \sum_{m=1}^{M}\sum_{i=1}^{N}\alpha_{m,i} = 1, \quad (3.45)$$

$$\frac{\partial L}{\partial \mathbf{a}} = 0 \quad \Rightarrow \quad \mathbf{a} = \sum_{m=1}^{M}\sum_{i=1}^{N}\alpha_{m,i}\mathbf{Q}_m\mathbf{x}_{m,i}, \quad (3.46)$$

$$\frac{\partial L}{\partial \xi_{m,i}} = 0 \quad \Rightarrow \quad C - \alpha_{m,i} - \gamma_{m,i} = 0. \quad (3.47)$$

Substituting (3.32), (3.45), (3.46) and (3.47) into (3.44), we get

$$L = \sum_{m=1}^{M}\sum_{i=1}^{N}\alpha_{m,i}\mathbf{y}_{m,i}^{\intercal}\mathbf{y}_{m,i} - \sum_{m=1}^{M}\sum_{i=1}^{N}\sum_{n=1}^{M}\sum_{j=1}^{N}\alpha_{m,i}\mathbf{y}_{m,i}^{\intercal}\mathbf{y}_{n,j}\alpha_{n,j}. \quad (3.48)$$

Optimizing (3.48) corresponds to SVDD applied in the subspace where $\alpha_{m,i}$ defines the position of the corresponding sample with respect to the hypersphere:

- Samples with $0 < \alpha_{m,i} < C$ are the support vectors lying on the boundary of the hypersphere.
- Samples with $\alpha_{m,i} = C$ are outside the hypersphere.
- Samples with $\alpha_{m,i} = 0$ lie inside the hypersphere.

We apply a two-step iterative optimization process. In the first step, the α values for all data points are obtained. In the second step, we update the corresponding projection matrices using gradient descent.

For updating the projection matrix $\mathbf{Q}$, we add a regularization term ω to (3.48):

$$L = \sum_{m=1}^{M}\sum_{i=1}^{N}\alpha_{m,i}\mathbf{x}_{m,i}^{\intercal}\mathbf{Q}_m^{\intercal}\mathbf{Q}_m\mathbf{x}_{m,i} - \sum_{m=1}^{M}\sum_{i=1}^{N}\sum_{n=1}^{M}\sum_{j=1}^{N}\alpha_{m,i}\mathbf{x}_{m,i}^{\intercal}\mathbf{Q}_m^{\intercal}\mathbf{Q}_n\mathbf{x}_{n,j}\alpha_{n,j} + \beta\omega. \tag{3.49}$$

where β controls the significance of ω. The different variants of ω are defined as

$$\omega_0 = 0, \tag{3.50}$$

$$\omega_1 = \sum_{m=1}^{M}\mathrm{Tr}(\mathbf{Q}_m\mathbf{X}_m\mathbf{X}_m^{\intercal}\mathbf{Q}_m^{\intercal}), \tag{3.51}$$

$$\omega_2 = \sum_{m=1}^{M}\mathrm{Tr}(\mathbf{Q}_m\mathbf{X}_m\boldsymbol{\alpha}_m\boldsymbol{\alpha}_m^{\intercal}\mathbf{X}_m^{\intercal}\mathbf{Q}_m^{\intercal}), \tag{3.52}$$

$$\omega_3 = \sum_{m=1}^{M}\mathrm{Tr}(\mathbf{Q}_m\mathbf{X}_m\boldsymbol{\lambda}_m\boldsymbol{\lambda}_m^{\intercal}\mathbf{X}_m^{\intercal}\mathbf{Q}_m^{\intercal}), \tag{3.53}$$

$$\omega_4 = \sum_{m=1}^{M}\sum_{n=1}^{M}\mathrm{Tr}(\mathbf{Q}_m\mathbf{X}_m\mathbf{X}_n^{\intercal}\mathbf{Q}_n^{\intercal}), \tag{3.54}$$

$$\omega_5 = \sum_{m=1}^{M}\sum_{n=1}^{M}\mathrm{Tr}(\mathbf{Q}_m\mathbf{X}_m\boldsymbol{\alpha}_m\boldsymbol{\alpha}_n^{\intercal}\mathbf{X}_n^{\intercal}\mathbf{Q}_n^{\intercal}), \tag{3.55}$$

$$\omega_6 = \sum_{m=1}^{M}\sum_{n=1}^{M}\mathrm{Tr}(\mathbf{Q}_m\mathbf{X}_m\boldsymbol{\lambda}_m\boldsymbol{\lambda}_n^{\intercal}\mathbf{X}_n^{\intercal}\mathbf{Q}_n^{\intercal}), \tag{3.56}$$

where $\boldsymbol{\alpha}_m \in \mathbb{R}^N$ in (3.52) and (3.55) contain the elements $\alpha_{m,1},...,\alpha_{m,N}$. Thus, $\boldsymbol{\alpha}_m$ have zero values for samples inside the hypersphere and non-zero values otherwise. $\boldsymbol{\lambda}_m \in \mathbb{R}^N$ in (3.53) and (3.56) contain the elements of $\boldsymbol{\alpha}_m$ that are smaller than C. Hence, $\boldsymbol{\lambda}_m$ contains non-zero values only for the samples lying on the boundary. For ω_0, the regularization term is not used in the optimization process. It was also noticed that for a single modality, the regularization terms become equivalent to the regularization terms proposed for SSVDD.

The projection matrix $\mathbf{Q}_m$ is updated by using the gradient of L with respect to $\mathbf{Q}_m$,

$$\mathbf{Q}_m \leftarrow \mathbf{Q}_m - \eta\Delta L, \tag{3.57}$$

where η is the learning rate parameter. The gradient of L is calculated as

$$\frac{\partial L}{\partial \mathbf{Q}_m} = 2\sum_{i=1}^{N} \alpha_{m,i} \mathbf{Q}_m \mathbf{x}_{m,i} \mathbf{x}_{m,i}^{\intercal} - 2\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{n=1}^{M} \mathbf{Q}_n \mathbf{x}_{n,j} \mathbf{x}_{m,i}^{\intercal} \alpha_{m,i} \alpha_{n,j} + \beta \Delta\omega, \quad (3.58)$$

where $\Delta\omega$ denotes the derivative of ω with respect to $\mathbf{Q}_m$:

$$\Delta\omega_0 = 0, \quad (3.59)$$

$$\Delta\omega_1 = 2\mathbf{Q}_m \mathbf{X}_m \mathbf{X}_m^{\intercal}, \quad (3.60)$$

$$\Delta\omega_2 = 2\mathbf{Q}_m \mathbf{X}_m \boldsymbol{\alpha}_m \boldsymbol{\alpha}_m^{\intercal} \mathbf{X}_m^{\intercal}, \quad (3.61)$$

$$\Delta\omega_3 = 2\mathbf{Q}_m \mathbf{X}_m \boldsymbol{\lambda}_m \boldsymbol{\lambda}_m^{\intercal} \mathbf{X}_m^{\intercal}, \quad (3.62)$$

$$\Delta\omega_4 = 2\sum_{n=1}^{M} (\mathbf{Q}_n \mathbf{X}_n \mathbf{X}_m^{\intercal}), \quad (3.63)$$

$$\Delta\omega_5 = 2\sum_{n=1}^{M} (\mathbf{Q}_n \mathbf{X}_n \boldsymbol{\alpha}_n \boldsymbol{\alpha}_m^{\intercal} \mathbf{X}_m^{\intercal}), \quad (3.64)$$

$$\Delta\omega_6 = 2\sum_{n=1}^{M} (\mathbf{Q}_n \mathbf{X}_n \boldsymbol{\lambda}_n \boldsymbol{\lambda}_m^{\intercal} \mathbf{X}_m^{\intercal}). \quad (3.65)$$

We use PCA for initializing the projection matrix $\mathbf{Q}_m$ corresponding to each modality. Furthermore, the projection matrix is orthogonalized and normalized using QR decomposition at each iteration so that

$$\mathbf{Q}_m \mathbf{Q}_m^{T} = \mathbf{I}, \quad (3.66)$$

where $\mathbf{I}$ is the $d \times d$ identity matrix. Algorithm 3 describes the overall MSSVDD algorithm.

Algorithm 3: MSSVDD optimization

Inputs : $\mathbf{X}_m$ for each $m = 1,...,M$, // Input data from all modalities
β, // Regularization parameter for controlling significance of ω
η, // Learning rate parameter
d, // Dimensionality of joint subspace
C, // Regularization parameter in SVDD
M // Total number of modalities

Outputs: $\mathbf{Q}_m$ for each $m = 1,...,M$, // Projection matrices for different modalities
R, // Radius of hypersphere
α // Defines the data description

for *m=1:M* **do**
  Initialize $\mathbf{Q}_m$ via PCA;
end

for $iter = 1 : max\_iter$ **do**
  For each m, project data to subspace as $\mathbf{y}_{m,i} = \mathbf{Q}_m \mathbf{x}_{m,i}$;
  Calculate α values by maximizing $L = \sum_{m=1}^{M}\sum_{i=1}^{N} \alpha_{m,i}\mathbf{y}_{m,i}^T\mathbf{y}_{m,i} - \sum_{m=1}^{M}\sum_{i=1}^{N}\sum_{n=1}^{M}\sum_{j=1}^{N} \alpha_{m,i}\mathbf{y}_{m,i}^T\mathbf{y}_{n,j}\alpha_{n,j}$;
  for *m=1:M* **do**
    Calculate $\Delta\omega$ as in (3.59)-(3.65);
    Calculate $\Delta L = 2\sum_{i=1}^{N} \alpha_{m,i}\mathbf{Q}_m\mathbf{x}_{m,i}\mathbf{x}_{m,i}^T - 2\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{n=1}^{M} \mathbf{Q}_n\mathbf{x}_{n,j}\mathbf{x}_{m,i}^T\alpha_{m,i}\alpha_{n,j} + \beta\Delta\omega$;
    Update $\mathbf{Q}_m \leftarrow \mathbf{Q}_m - \eta\Delta L$;
    Orthogonalize and normalize $\mathbf{Q}_m$ using QR decomposition (eigendecomposition);
  end
end
For each m, project data to subspace as $\mathbf{y}_{m,i} = \mathbf{Q}_m \mathbf{x}_{m,i}$;
Calculate α values by maximizing $L = \sum_{m=1}^{M}\sum_{i=1}^{N} \alpha_{m,i}\mathbf{y}_{m,i}^T\mathbf{y}_{m,i} - \sum_{m=1}^{M}\sum_{i=1}^{N}\sum_{n=1}^{M}\sum_{j=1}^{N} \alpha_{m,i}\mathbf{y}_{m,i}^T\mathbf{y}_{n,j}\alpha_{n,j}$;
Compute center of data description as $\mathbf{a} = \sum_{m=1}^{M}\sum_{i=1}^{N} \alpha_{m,i}\mathbf{Q}_m\mathbf{x}_{m,i}$;
Identify any support vector $\mathbf{s}_m$ having $0 \leq \alpha_s \leq C$ and compute $\mathbf{v} = \mathbf{Q}_m\mathbf{s}_m$;
Compute radius $R = \sqrt{\mathbf{v}^T\mathbf{v} - 2\sum_{m=1}^{M}\sum_{i=1}^{N} \alpha_{m,i}\mathbf{y}_{m,i}^T\mathbf{v} + \sum_{m=1}^{M}\sum_{i=1}^{N}\sum_{n=1}^{M}\sum_{j=1}^{N} \alpha_{m,i}\alpha_{n,j}\mathbf{y}_{m,i}^T\mathbf{y}_{n,j}}$;

3.3.1 Complexity Analysis of MSSVDD

The linear version of MSSVDD has the following six main steps:

1. Initializing of $\mathbf{Q}_m$ via PCA comprises two steps, i.e., computing the covariance matrix and then the eigenvalue decomposition. The complexity of these steps for a single modality is $\mathcal{O}\big(ND_m \times min(N, D_m)\big)$ and $\mathcal{O}\big(D_m^3\big)$, respectively [72]. The complexity of initializing $\mathbf{Q}_m$ for all modalities is $\mathcal{O}\big(min(N^2D_1, D_1^2 N) + D_1^3) + (min(N^2D_2, D_2^2 N) + D_2^3) + \cdots + (min(N^2D_M, D_M^2 N) + D_M^3\big)$. We denote the sum of dimensions as $\Sigma_{\mathscr{D}} = D_1 + D_2 + \cdots + D_M$, and in a similar manner, the sum of squared dimensions as $\Sigma_{\mathscr{D}^2} = D_1^2 + D_2^2 + \cdots + D_M^2$ (note that $\Sigma_{\mathscr{D}^2} \neq (\Sigma_{\mathscr{D}})^2$) and so on. The complexity of initializing all projection matrices becomes $\mathcal{O}\big(min(N^2\Sigma_{\mathscr{D}}, \Sigma_{\mathscr{D}^2} N) + \Sigma_{\mathscr{D}^3}\big)$.
2. The complexity of projecting the data to a d-dimensional space is the complexity of multiplying two $d \times D_m$ and $D_m \times N$ matrices. This has the complexity of $\mathcal{O}\big(dD_m N\big)$. Projecting the data from all modalities to a lower d-dimensional space has the complexity of $\mathcal{O}\big(d\Sigma_{\mathscr{D}} N\big)$.
3. SVDD has the complexity of $\mathcal{O}\big(N^3\big)$ for N data points [68]. In the case of MSSVDD, we have data points represented in M different modalities. Hence the complexity of SVDD in this case becomes $\mathcal{O}\big(M^3 N^3\big)$.
4. The gradient ΔL is computed using (3.58). The second term in (3.58) has the highest complexity of $\mathcal{O}(2dN^2 D_m \Sigma_{\mathscr{D}})$. Since this step is repeated for all modalities as well, the complexity of this step becomes $\mathcal{O}(2dN^2 {\Sigma_{\mathscr{D}}}^2)$.
5. The complexity of updating the $\mathbf{Q}_m$ for all modalities is $\mathcal{O}\big(d\Sigma_{\mathscr{D}}\big)$.
6. For a single modality, the complexity of QR decomposition is $\mathcal{O}(dD_m{}^2)$ [69]. Thus, the complexity of QR decomposition for data from all modalities becomes $\mathcal{O}(d\Sigma_{\mathscr{D}^2})$.

The relatively low computational complexity steps are not considered in the complexity analysis. By adding the remaining terms, the complete complexity of MSSVDD reduces to $\mathcal{O}\big(min(N^2\Sigma_{\mathscr{D}}, \Sigma_{\mathscr{D}^2} N) + \Sigma_{\mathscr{D}^3} + M^3 N^3\big)$. With the assumption that the total number of samples $M * N$ is always greater than $\Sigma_{\mathscr{D}}$ and the number of modalities is smaller than the number of samples, i.e., $M << N$, the time complexity of MSSVDD

in terms of the big $\mathcal{O}$ notation is $\mathcal{O}(N^3)$ for a single iteration. During testing, an instance is first projected to the lower d-dimensional subspace, and then its distance is compared to R. The total complexity in the testing phase is $\mathcal{O}(d\Sigma_{\mathscr{D}} + Md)$.

In the case of non-linear version with NPT, the kernel matrix $\mathbf{K}_m$ is formed, centralized, and decomposed via eigendecomposition. These steps have the complexity of $\mathcal{O}(N^3)$. The dimensionality of each modality changes from D_m to N for all corresponding steps. Hence the total complexity of the subsequent steps becomes $\mathcal{O}\big(MN^3 + M^3N^3\big)$. Thus, the overall complexity remains at $\mathcal{O}\big(N^3\big)$ for $M << N$. Furthermore, the overall complexity stays the same in the case of a non-linear version with the standard kernel trick. The complexity of the non-linear variants during the testing phase increases to $\mathcal{O}(N\Sigma_{\mathscr{D}} + dMN + Md)$.

3.3.2 Decision Strategies

The decision to assign the test instance in multimodal data description can be taken based on different strategies:

- **Decision strategy 1 ($ds1$):** Decision strategy one is called AND gate strategy. The test instance is assigned the target label if the representations from all modalities for the test instance are classified as the target class. The test instance is assigned a non-target label otherwise.
- **Decision strategy 2 ($ds2$):** This decision strategy is called the OR gate strategy. Here, the overall decision for a particular instance is taken in favor of the target class if any of the modalities is classified as the target class.
- **Decision strategy 3 ($ds3$):** In this strategy, the decision is taken solely based on the predicted labels of the first modality.
- **Decision strategy 4 ($ds4$):** In this strategy, the decision is taken solely based on the predicted labels of the second modality.

3.3.3 Results and Discussion

We evaluated the performance of MSSVDD over five different multimodal datasets. The Robot Execution Failures dataset, Single Proton Emission Computed Tomography (SPECTF) heart dataset, and Ionosphere dataset were downloaded from UCI

repository [70]. The Handwritten and the Caltech-7 dataset were downloaded from the multi-view learning repository [73]. All of these datasets are multimodal in nature, and each modality has its own statistical properties. Thus, to combine the information from all modalities in a subspace, we created and optimized a joint compact description of data coming from all the modalities.

In the Robot Execution Failures dataset, force and torque measurements are used as different modalities of the collected instances. In the SPECTF dataset, features corresponding to rest and stress conditions of SPECTF images depict the multimodal nature of data. In the Ionosphere dataset, the categories are described by two attributes per pulse number. The two attributes in the Ionosphere dataset are obtained from the complex electromagnetic signal, processed by an autocorrelation function, and are considered as different modalities. In the Handwritten dataset, we used the Zernike moment (ZER) and morphological (MOR) features as our two different modalities. We used the Gabor feature and Wavelet moments as data from two different modalities in the Caltech-7 dataset.

We provide the results over the Handwritten dataset in Table 3.6. We only show the best-performing versions of the proposed method in the table. The table contains the results of the proposed method with the regularization term ω_4 based on decision strategies 4 (*ds4*) and 1 (*ds1*). The experiments over the Handwritten dataset were performed on 70-30% training and testing sets. We selected the samples of numeral 0 as the target class. The 70-30% splits were selected randomly five times while preserving the distribution of classes similar to the original data. We performed 5-fold cross-validation on the training set to determine the hyperparameters for final training. The hyperparameters were selected based on the GM score over the validation data. We report the average Accuracy (Acc), TPR, TNR, Pre, F1, and GM on the test sets. Acc is defined as the ratio of the number of correctly classified samples (TPs+True Negatives (TNs)) to the total number of samples N:

$$\text{Acc} = \frac{\text{TP}s + \text{TN}s}{N}. \tag{3.67}$$

We compare our results with OCSVM, SVDD and SSVDD. We observed that the linear and NPT versions of the proposed method outperformed all competing methods in most cases. Among the non-linear variants of the MSSVDD, NPT-based non-linear multimodal data description performed better in terms of GM than kernel trick-based data description. The decision strategies were also evaluated in de-

Table 3.6 Test results for Handwritten dataset [P3].

	Linear						Non-linear					
	Acc	TPR	TNR	Pre	F1	GM	Acc	TPR	TNR	Pre	F1	GM
Proposed method												
MSSVDD $\omega_4 ds4$	0.98	0.99	0.98	0.90	0.93	**0.98**	0.99	0.99	1.00	0.98	0.98	**0.99**
MSSVDD $\omega_4 ds1$	0.98	0.90	0.99	0.89	0.89	0.94	0.98	0.95	0.99	0.91	0.93	0.97
Concatenated features												
SSVDDψ_1	0.78	0.92	0.76	0.34	0.49	0.83	0.53	0.40	0.54	0.05	0.09	0.14
SSVDDψ_2	0.82	0.88	0.81	0.40	0.54	0.84	0.62	0.66	0.61	0.18	0.25	0.44
SSVDDψ_3	0.82	0.97	0.81	0.39	0.55	0.88	0.63	0.58	0.64	0.20	0.25	0.30
SSVDD ψ_4	0.84	0.92	0.83	0.42	0.56	0.87	0.71	0.39	0.75	0.08	0.13	0.17
OCSVM	0.50	0.51	0.50	0.12	0.19	0.49	0.95	0.51	1.00	1.00	0.68	0.71
SVDD	0.95	0.93	0.95	0.69	0.79	0.94	0.95	0.92	0.96	0.74	0.81	0.94
ZER												
SSVDD ψ_1	0.55	0.92	0.51	0.18	0.30	0.68	0.59	0.41	0.61	0.06	0.10	0.24
SSVDD ψ_2	0.52	0.88	0.48	0.17	0.28	0.64	0.62	0.78	0.60	0.17	0.27	0.48
SSVDD ψ_3	0.50	0.96	0.45	0.19	0.31	0.63	0.57	0.61	0.57	0.31	0.20	0.37
SSVDD ψ_4	0.64	0.90	0.61	0.21	0.34	0.74	0.55	0.60	0.54	0.09	0.15	0.24
OCSVM	0.43	0.42	0.43	0.09	0.14	0.41	0.95	0.52	1.00	0.93	0.67	0.72
SVDD	0.88	0.90	0.88	0.47	0.61	0.89	0.92	0.88	0.92	0.56	0.68	0.90
MOR												
SSVDD ψ_1	0.84	0.99	0.82	0.48	0.61	0.90	0.84	0.01	0.93	0.00	0.00	0.03
SSVDD ψ_2	0.92	0.99	0.91	0.66	0.76	0.95	0.58	0.44	0.60	0.43	0.22	0.20
SSVDD ψ_3	0.86	0.99	0.84	0.52	0.64	0.91	0.61	0.70	0.60	0.44	0.42	0.36
SSVDD ψ_4	0.84	0.99	0.82	0.48	0.61	0.90	0.25	0.67	0.20	0.27	0.14	0.04
OCSVM	0.54	0.45	0.55	0.13	0.18	0.39	0.99	0.87	1.00	1.00	0.93	0.93
SVDD	0.93	0.91	0.93	0.75	0.78	0.92	0.99	0.96	1.00	1.00	0.98	0.98

tail for each dataset. We observed that if information from a certain modality is more substantial than other(s), taking the final decision solely based on that particular modality in the testing phase is useful. Nevertheless, the proposed MSSVDD improves the classification accuracy compared to models using information from the single modality only. When the information depicted by different modalities together is vital for data description and taking the final decision, the AND gate

strategy performs better than other decision strategies. We provide a detailed analysis in publication [P3]. Moreover, we provide extensive results comparing different variants of the MSSVDD, train and test times of the experiments, and the figures for sensitivity analysis of different hyperparameters in the supplementary material appended with the corresponding publication. The code of MSSVDD is publicly available online[3].

The proposed MSSVDD addressed the research question about finding an optimized subspace and training a common model for multimodal data in the case of OCC. The developed method achieved the objective of inferring a shared latent representation and exploiting the data relationship between the modalities in OCC. The proposed method uses a hypersphere to encapsulate target data which can be further improved by using hyperellipsoids in the subspace. Moreover, the proposed decision strategies can be further enhanced by using model-based decision strategies.

3.4 Graph-Embedded Subspace Support Vector Data Description

In this section, we formulate the subspace learning for OCC in a graph embedded framework. We proposed GESSVDD which allows incorporating different data relationships in the form of graph structures in the optimization process. The framework also helps reveal spectral and spectral regression-based solutions as alternatives to the previously used gradient-based technique for solving the subspace learning for OCC problem and leads to various novel insights. Moreover, SSVDD and ESSVDD are deduced as special cases of the proposed framework.

In GESSVDD, the following optimization criterion is minimized:

$$\begin{aligned} \min \quad & R^2 + C\sum_{i=1}^{N}\xi_i \\ \text{s.t.} \quad & (\mathbf{Q}\mathbf{x}_i - \mathbf{a})^\intercal \mathbf{S}_Q^{-1}(\mathbf{Q}\mathbf{x}_i - \mathbf{a}) \leq R^2 + \xi_i, \\ & \xi_i \geq 0, \forall i \in \{1,\ldots,N\}, \end{aligned} \tag{3.68}$$

where the matrix $\mathbf{S}_Q$ is used to encode different geometric data relationships in the

[3] https://github.com/fahadsohrab/mssvdd

subspace as

$$\mathbf{S}_Q = \mathbf{Q}\mathbf{X}\mathbf{L}_x\mathbf{X}^\intercal\mathbf{Q}^\intercal = \mathbf{Q}\mathbf{S}_x\mathbf{Q}^\intercal. \tag{3.69}$$

$\mathbf{L}_x$ is a graph Laplacian and can take different forms. We define a new vector $\mathbf{u} = \mathbf{S}_Q^{-\frac{1}{2}}\mathbf{a}$ as the center of graph-embedded data description relative to the center of the hypersphere $\mathbf{a}$ and rewrite (3.68) as

$$\begin{aligned} \min \quad & R^2 + C\sum_{i=1}^{N}\xi_i \\ \text{s.t.} \quad & ||\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i - \mathbf{u}||_2^2 \leq R^2 + \xi_i, \\ & \xi_i \geq 0, \forall i \in \{1,\ldots,N\}. \end{aligned} \tag{3.70}$$

Note that $\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}$ can be considered as a projection matrix to the graph-embedded subspace. The Lagrangian function of GESVDD can be written as

$$\begin{aligned} L = R^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\big(R^2 + \xi_i - \\ (\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i)^\intercal\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i + 2\mathbf{u}^\intercal\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i - \mathbf{u}^\intercal\mathbf{u}\big) - \sum_{i=1}^{N}\gamma_i\xi_i. \end{aligned} \tag{3.71}$$

By setting partial derivatives to zero, we get

$$\frac{\partial L}{\partial R} = 0 \quad \Rightarrow \quad \sum_{i=1}^{N}\alpha_i = 1, \tag{3.72}$$

$$\frac{\partial L}{\partial \mathbf{u}} = 0 \quad \Rightarrow \quad \mathbf{u} = \sum_{i=1}^{N}\alpha_i\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i, \tag{3.73}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \quad \Rightarrow \quad C - \alpha_i - \gamma_i = 0. \tag{3.74}$$

By substituting (3.72)-(3.74) into (3.71), we get

$$L = \sum_{i=1}^{N}\alpha_i\mathbf{x}_i^\intercal\mathbf{Q}^\intercal\mathbf{S}_Q^{-1}\mathbf{Q}\mathbf{x}_i - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\mathbf{x}_i^\intercal\mathbf{Q}^\intercal\mathbf{S}_Q^{-1}\mathbf{Q}\mathbf{x}_j\alpha_j. \tag{3.75}$$

We denote the vectors in the graph-embedded d-dimensional space as $\mathbf{z}_i = \mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i$ and write (3.75) as

$$L = \sum_{i=1}^{N} \alpha_i \mathbf{z}_i^\intercal \mathbf{z}_i - \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j \mathbf{z}_i^\intercal \mathbf{z}_j. \tag{3.76}$$

Maximizing (3.76) corresponds to solving SVDD for data in the graph-embedded subspace.

We can further simplify (3.76) by writing it in a trace form as

$$\begin{aligned} L &= \mathrm{Tr}(\mathbf{S}_Q^{-1}\mathbf{Q}\mathbf{X}\mathbb{A}\mathbf{X}^\intercal\mathbf{Q}^\intercal) - \mathrm{Tr}(\mathbf{S}_Q^{-1}\mathbf{Q}\mathbf{X}\boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal\mathbf{X}^\intercal\mathbf{Q}^\intercal) \\ &= \mathrm{Tr}((\mathbf{Q}\mathbf{X}\mathbf{L}_x\mathbf{X}^\intercal\mathbf{Q}^\intercal)^{-1}\mathbf{Q}\mathbf{X}(\mathbb{A} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal)\mathbf{X}^\intercal\mathbf{Q}^\intercal), \end{aligned} \tag{3.77}$$

where the matrix $\mathbb{A} \in \mathbb{R}^{N\times N}$ is a diagonal matrix containing the α_i values in its diagonal, $\boldsymbol{\alpha}$ is a vector of the α_i values. Now by defining the matrices

$$\mathbf{L}_\alpha = \mathbb{A} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal \tag{3.78}$$

$$\mathbf{S}_\alpha = \mathbf{X}\mathbf{L}_\alpha\mathbf{X}^\intercal, \tag{3.79}$$

we can further simplify (3.77) as

$$L = \mathrm{Tr}\big((\mathbf{Q}\mathbf{S}_x\mathbf{Q}^\intercal)^{-1}\mathbf{Q}\mathbf{S}_\alpha\mathbf{Q}^\intercal\big). \tag{3.80}$$

We see that (3.80) is in the form of the ratio trace problem of (2.17), which can be solved by finding generalized eigenvalues and eigenvectors, and is commonly used as an approximation for the Fisher-Rao's criterion typically encountered in subspace learning as discussed in Section 2.2.2. Hence, (3.80) reveals an alternative generalized eigenvalue-based solution to SSVDD. While in the well-known subspace learning techniques, such as LDA, the generalized eigenvalue-based solution is an approximation, it provides the exact solution in our case.

Furthermore, the similarity of (3.80) and (2.22) reveals that we can describe sub-

space learning for SVDD in a novel graph-embedded framework as

$$\begin{aligned}\mathbf{Q}^* &= \underset{\mathrm{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L}_x\mathbf{X}^\intercal\mathbf{Q}^\intercal)=m}{\arg\min} \sum_{i\neq j}(\mathbf{Q}\mathbf{x}_i - \mathbf{Q}\mathbf{x}_j)^2\alpha_i\alpha_j \\ &= \arg\min \frac{\mathrm{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L}_\alpha\mathbf{X}^\intercal\mathbf{Q}^\intercal)}{\mathrm{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L}_x\mathbf{X}^\intercal\mathbf{Q}^\intercal)},\end{aligned} \tag{3.81}$$

where $\mathbf{L}_\alpha$ is the graph Laplacian of a special SVDD graph having the corresponding weight matrix (also known as the similarity matrix) $\mathbf{A}$ defined as $[\mathbf{A}]_{ij} = \alpha_i\alpha_j$ and the Laplacian $\mathbf{L}_x$ can be used for encoding different graph information in the optimization process. As described in Section 2.3.1, an alternative solution to (3.81) can be obtained via spectral regression

Previously, in SSVDD and ESSVDD, the criterion has been minimized via gradient descent. However, the new framework suggests that instead of minimizing, the objective can also be maximized depending on the associated graph. For example, a common intuition is that the within-cluster scatter should be minimized while the between-cluster scatter should be maximized. The approach to maximize the objective has previously been followed in [74], where kernel PCA was applied for novelty detection.

We provide the pseudo-code of the main GESSVDD in Algorithm 4 with the options to select gradient, spectral and spectral regression-based updates described in Sub-algorithms 1, 2, and 3, respectively.

Algorithm 4: GESSVDD optimization

Input : $\mathbf{X}$, // Input data
$\mathbf{L}_x$ // Input Laplacian
η, // Learning rate parameter
d, // Dimensionality of subspace
C, // Regularization parameter in SVDD
min or max // Either minimize or maximize the criterion

Output: $\mathbf{Q}$ // Projection matrix
R, // Radius of hypersphere
$\boldsymbol{\alpha}$ // Defines the data description

Initialize $\mathbf{Q}$ via PCA; // Select d-vectors corresponding to d largest eigenvalues.
Compute $\mathbf{S}_x = \mathbf{X}\mathbf{L}_x\mathbf{X}^\intercal$;

for $iter = 1 : max\_iter$ **do**
  Calculate $\mathbf{S}_{inv} = \mathbf{S}_Q^{-1} = (\mathbf{Q}\mathbf{S}_x\mathbf{Q}^\intercal)^{-1}$;
  Project data to subspace $\mathbf{z}_i = \mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i = (\mathbf{S}_{inv})^{\frac{1}{2}}\mathbf{Q}\mathbf{x}_i$;
  Calculate α values by maximizing
  $L = \sum_{i=1}^{N}\alpha_i\mathbf{z}_i^\intercal\mathbf{z}_i - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\mathbf{z}_i^\intercal\mathbf{z}_j$;
  Compute $\mathbf{L}_\alpha = \mathbb{A} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal$;
  if *gradient-based update*
    Call Sub-algorithm 1 to obtain $\mathbf{Q}$;
  elseif *spectral update*
    Call Sub-algorithm 2 to obtain $\mathbf{Q}$;
  elseif *spectral regression-based update*:
    Call Sub-algorithm 3 to obtain $\mathbf{Q}$;
  endif
  Orthogonalize $\mathbf{Q}$ using QR decomposition;
end

Project data to subspace $\mathbf{z}_i = \mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i$;
Calculate α values by maximizing $L = \sum_{i=1}^{N}\alpha_i\mathbf{z}_i^\intercal\mathbf{z}_i - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\mathbf{z}_i^\intercal\mathbf{z}_j$;
Compute center of data description in the subspace as $\mathbf{u} = \sum_{i=1}^{N}\alpha_i\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i$;
Identify any support vector $\mathbf{s}$ having $0 < \alpha_s < C$;
Compute radius $R = \sqrt{(\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{s})^\intercal\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{s} - 2(\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{s})^\intercal\mathbf{u} + \mathbf{u}^\intercal\mathbf{u}}$;

Sub-algorithm 1: Gradient-based update

Input : $\mathbf{Q}$, $\mathbf{X}$, $\mathbf{S}_x$, $\mathbf{S}_{inv}$ $\mathbf{L}_\alpha$, η, *min/max* //Input from Algorithm 4
Output: $\mathbf{Q}$ //Return output to Algorithm 4

Compute $\mathbf{S}_\alpha = \mathbf{X}\mathbf{L}_\alpha\mathbf{X}^\intercal$;
Compute $\Delta L = 2\mathbf{S}_{inv}\mathbf{Q}\mathbf{S}_\alpha - 2\mathbf{S}_{inv}\mathbf{Q}\mathbf{S}_\alpha\mathbf{Q}^\intercal\mathbf{S}_{inv}\mathbf{Q}\mathbf{S}_x^\intercal$;
if *minimization*
Update $\mathbf{Q} \leftarrow \mathbf{Q} - \eta\Delta L$;
elseif *maximization*
Update $\mathbf{Q} \leftarrow \mathbf{Q} + \eta\Delta L$;

Sub-algorithm 2: Spectral update

Input : $\mathbf{X}$, $\mathbf{S}_x$, $\mathbf{L}_\alpha$, *min/max* //Input from Algorithm 4
Output: $\mathbf{Q}$ //Return output to Algorithm 4

Compute $\mathbf{S}_\alpha = \mathbf{X}\mathbf{L}_\alpha\mathbf{X}^\intercal$;
Solve generalized eigenvalue problem $\mathbf{S}_\alpha\mathbf{q} = \upsilon\mathbf{S}_x\mathbf{q}$;
if *minimization*
Select the eigenvectors corresponding to d smallest positive eigenvalues as rows of $\mathbf{Q}$;
elseif *maximization*
Select the eigenvectors corresponding to d largest eigenvalues as rows of $\mathbf{Q}$;

Sub-algorithm 3: Spectral regression-based update

Input : $\mathbf{X}$, $\mathbf{L}$, $\mathbf{L}_\alpha$, *min/max* //Input from Algorithm 4
Output: $\mathbf{Q}$ //Return output to Algorithm 4

Solve generalized eigenvalue problem: $\mathbf{L}_\alpha\mathbf{t} = \upsilon\mathbf{L}_x\mathbf{t}$;
if *minimization* then
Select the eigenvectors corresponding to d smallest positive eigenvalues as columns of $\mathbf{T}$;
elseif *maximization* then
Select the eigenvectors corresponding to d largest eigenvalues as columns of $\mathbf{T}$;
Obtain $\mathbf{Q} = \mathbf{T}^\intercal\mathbf{X}^\intercal(\mathbf{X}\mathbf{X}^\intercal + \eta\mathbf{I})^{-1}$;

3.4.1 Graph Laplacians in GESSVDD

In GESSVDD, $\mathbf{L}_x$ can be replaced by any suitable graph. We provide a list of options that we also used in our experiments. The first option has no data-dependent constraint, here $\mathbf{S}_x$ in (3.80) is replaced by the $D \times D$ identity matrix $\mathbf{I}$. This corresponds to the orthogonality constraint. We denote the first option by GESSVDD-0. The second options, denoted by GESSVDD-I, uses $\mathbf{L}_x = \mathbf{I}$. The third option uses the PCA graph: $\mathbf{S}_x = \frac{1}{N}\mathbf{S}_t = \frac{1}{N}\mathbf{X}\mathbf{L}_t\mathbf{X}^\intercal = \frac{1}{N}\mathbf{X}(\mathbf{I} - \frac{1}{N})\mathbf{X}^\intercal$, which is denoted by GESSVDD-PCA. In our experiments, we also cluster the positive training samples using k-means and then set $\mathbf{S}_x = \mathbf{S}_w$ or $\mathbf{S}_x = \mathbf{S}_b$. The Laplacians for the within-class scatter matrix $\mathbf{S}_w = \mathbf{X}\mathbf{L}_w\mathbf{X}^\intercal$ and the between-class scatter matrix $\mathbf{S}_b = \mathbf{X}\mathbf{L}_b\mathbf{X}^\intercal$ are solved as in (2.25) and (2.26), respectively. The variable c, in this case, refers to a cluster, not a class. Moreover, local geometric information can be exploited by employing k-Nearest Neighbors (kNN) and setting

$$\mathbf{S}_x = \mathbf{S}_{kNN} = \mathbf{X}\Big(\mathbf{D}_{kNN} - \mathbf{A}_{kNN}\Big)\mathbf{X}^\intercal = \mathbf{X}\mathbf{L}_{kNN}\mathbf{X}^\intercal, \tag{3.82}$$

where $[\mathbf{A}]_{ij} = 1$, if $\mathbf{x}_i \in \mathcal{N}_j$ or $\mathbf{x}_j \in \mathcal{N}_i$ and 0, otherwise. $\mathcal{N}_i$ represents the nearest neighbors of $\mathbf{x}_i$. We denote this option by GESSVDD-kNN. Each of these options in the framework can be solved by gradient-based (GR), spectral ($\mathscr{S}$), or spectral regression-based (SR) updates and by either maximizing or minimizing the criterion in (3.80). We denote all these variants by GESSVDD-0-GR-min, GESSVDD-0-GR-max, GESSVDD-0-$\mathscr{S}$-min and so on. Note that GESSVDD-0-GR-min and GESSVDD-I-GR-min corresponds to SSVDD and ESSVDD, respectively.

We now turn our attention to the fixed SVDD graph represented by $\mathbf{L}_\alpha$ in (3.80). We notice a strong resemblance of $\mathbf{L}_\alpha$ to the Laplacian of the PCA graph. PCA finds the projection matrix by maximizing the variance of the samples to their center

$\mu = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i$, i.e.,

$$\begin{aligned}
\mathbf{S}_{pca} &= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i-\mu)(\mathbf{x}_i-\mu)^\intercal \\
&= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i\mathbf{x}_i^\intercal - 2\mathbf{x}_i\mu^\intercal + \mu\mu^\intercal) \\
&= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i\mathbf{x}_i^\intercal) - 2\mu\mu^\intercal + \mu\mu^\intercal = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i\mathbf{x}_i^\intercal) - \mu\mu^\intercal \\
&= \frac{1}{N}\mathbf{X}\mathbf{X}^\intercal - \frac{1}{N^2}\mathbf{X}\mathbf{1}\mathbf{1}^\intercal\mathbf{X}^\intercal = \frac{1}{N}\mathbf{X}(\mathbf{I}-\frac{1}{N}\mathbf{1}\mathbf{1}^\intercal)\mathbf{X}^\intercal \\
&= \mathbf{X}\mathbf{L}_{pca}\mathbf{X}^\intercal,
\end{aligned} \tag{3.83}$$

where $\mathbf{1} \in \mathbb{R}^N$ symbolizes a vector of ones, $\mathbf{L}_{pca} = \mathbf{D}_{pca} - \mathbf{A}_{pca}$ and $[\mathbf{A}_{pca}]_{ij} = 1/N^2 \;\; \forall i \neq j$ and $[\mathbf{A}_{pca}]_{ii} = 0$. We use a similar derivation with the constraint $\sum_{i=1}^{N}\alpha_i = 1$ for GESSVDD. The center of SVDD is defined as $\mathbf{a} = \sum_{i=1}^{N}\alpha_i\mathbf{x}_i$:

$$\begin{aligned}
\mathbf{S}_\alpha &= \sum_{i=1}^{N}(\mathbf{x}_i-\mathbf{a})(\mathbf{x}_i-\mathbf{a})^\intercal\alpha_i \\
&= \sum_{i=1}^{N}(\alpha_i\mathbf{x}_i\mathbf{x}_i^\intercal - 2\alpha_i\mathbf{x}_i\mathbf{a}^\intercal + \alpha_i\mathbf{a}\mathbf{a}^\intercal) \\
&= \sum_{i=1}^{N}(\alpha_i\mathbf{x}_i\mathbf{x}_i^\intercal) - 2\mathbf{a}\mathbf{a}^\intercal + \mathbf{a}\mathbf{a}^\intercal = \sum_{i=1}^{N}(\alpha_i\mathbf{x}_i\mathbf{x}_i^\intercal) - \mathbf{a}\mathbf{a}^\intercal \\
&= \mathbf{X}\mathrm{diag}(\boldsymbol{\alpha})\mathbf{X}^\intercal - \mathbf{X}\boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal\mathbf{X}^\intercal = \mathbf{X}(\mathbb{A}-\boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal)\mathbf{X}^\intercal \\
&= \mathbf{X}\mathbf{L}_\alpha\mathbf{X}^\intercal.
\end{aligned} \tag{3.84}$$

We see that the graph Laplacian $\mathbf{L}_\alpha$ represents the weighted variance of data points with positive α values. The resemblance is depicted in Figure 3.4. In the PCA graph, all the data points are connected with equal weights; however, in the SVDD graph $\mathbf{L}_\alpha$, the data points i and j are connected with the weight of $\alpha_i\alpha_j$. Thus, the samples with zero α values are not connected in the SVDD graph, i.e., only the samples lying on the boundary and outside the boundary are connected.

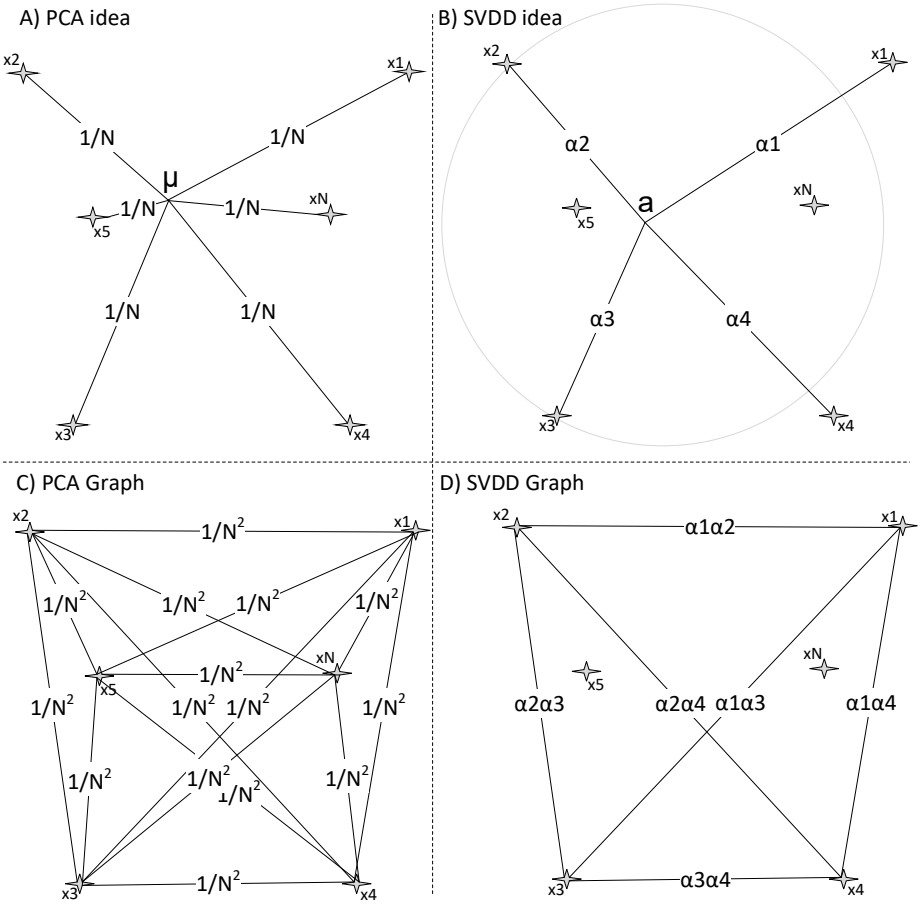

Figure 3.4 **A**) PCA takes into account the (unweighted) variance of all samples from the center. **B**) SVDD takes into account the weighted variance of samples with positive α values to the SVDD center. **C**) In the PCA graph, all samples are connected with equal weights. **D**) In the SVDD graph, the samples inside the hypersphere are not connected, and the weights vary according to the α values [P4].

3.4.2 Complexity Analysis of GESSVDD

GESSVDD can be solved via spectral, spectral regression-based, and gradient-based techniques. We carry out the complexity analysis of the shared steps (main algorithm) and then proceed to the steps different in each solution update. The following steps are in the main algorithm and contribute to the overall complexity of the algorithm:

1. Initializing of $\mathbf{Q}$ via PCA comprises two steps, i.e., computing the covariance matrix and then the eigenvalue decomposition. The complexity of these steps is $\mathcal{O}\big(ND \times min(N,D)\big)$ and $\mathcal{O}\big(D^3\big)$, respectively [72].
2. Computing $\mathbf{S}_x = \mathbf{X}\mathbf{L}_x\mathbf{X}^\intercal$ for a given $\mathbf{L}_x$ is simply the multiplication of three matrices. This step has the complexity of $\mathcal{O}(DN^2 + ND^2)$.
3. For a given encoding matrix $\mathbf{S}_x$, the matrix $\mathbf{S}_Q$ is computed as the multiplication of three matrices i.e., $\mathbf{S}_Q = \mathbf{Q}\mathbf{S}_x\mathbf{Q}^\intercal$. This has the complexity of $\mathcal{O}(dD^2 + d^2D)$. Since, $D > d$, the complexity becomes $\mathcal{O}(dD^2)$.
4. Computing the inverse $\mathbf{S}_{inv}$, and square-root of the matrix $\mathbf{S}_Q$ has the complexity of $\mathcal{O}(N^3)$.
5. SVDD has the complexity of $\mathcal{O}\big(N^3\big)$ for N data points [68].
6. The complexity of QR decomposition is $\mathcal{O}(dD^2)$ [69].

Adding all these steps and dropping relatively lower computational costs, the complexity becomes $\mathcal{O}\big(N^3 + D^3\big)$. The total number of samples is assumed to be always greater than the dimensionality. The overall complexity of the steps in the main algorithm (without the projection matrix update steps) in terms of big $\mathcal{O}$ analysis becomes $\mathcal{O}\big(N^3\big)$. We now analyze the complexity of each update solution separately in the following subsections.

3.4.2.1 Complexity Analysis of Gradient-based Update

The gradient-based update has the following steps involved in the sub-algorithm.

1. Computing $\mathbf{S}_\alpha = \mathbf{X}\mathbf{L}_\alpha\mathbf{X}^\intercal$ for a given $\mathbf{L}_\alpha$ has a complexity of multiplying three matrices as well. Thus, computing $\mathbf{S}_\alpha$ has the complexity of $\mathcal{O}(DN^2 + ND^2)$.
2. Computing $\Delta L = 2\mathbf{S}_{inv}\mathbf{Q}\mathbf{S}_\alpha - 2\mathbf{S}_{inv}\mathbf{Q}\mathbf{S}_\alpha\mathbf{Q}^\intercal\mathbf{S}_{inv}\mathbf{Q}\mathbf{S}_x^\intercal$ requires several matrix multiplications with the highest complexity being $\mathcal{O}(dD^2)$ and, thus, the overall complexity of this step also becomes $\mathcal{O}(dD^2)$.
3. The complexity of updating the $\mathbf{Q}$ is $\mathcal{O}\big(dD\big)$.

After adding all complexities of the gradient-based updates along with the complexity of the main algorithm, the gradient-based solution has a complexity of $\mathcal{O}(N^3)$. In a non-linear case, the kernel matrix $\mathbf{K}$ is formed, centralized, and decomposed via

eigendecomposition. These steps have the complexity of $\mathcal{O}(N^3)$. The dimensionality of data in the non-linear case changes from D to N for all corresponding steps. The total complexity of the non-linear version stays at $\mathcal{O}(N^3)$.

3.4.2.2 Complexity Analysis of Spectral-based Update

The main relatively intensive computational steps in the spectral-based update are computing $\mathbf{S}_\alpha$ and the generalized eigenvalue problem.

1. Computing $\mathbf{S}_\alpha$ has the complexity of $\mathcal{O}(DN^2 + ND^2)$.
2. Solving the eigenvalue problem $\mathbf{S}_\alpha \mathbf{q} = \upsilon \mathbf{S}_x \mathbf{q}$ has the complexity of $\mathcal{O}(D^3)$.

By adding the above complexities, the complexity of the spectral-based update becomes $\mathcal{O}(DN^2 + ND^2 + D^3)$. By adding this to the complexity of the main algorithm, the complexity becomes $\mathcal{O}\big(N^3 + DN^2 + ND^2 + D^3\big)$. Hence for the linear case, the complexity is $\mathcal{O}\big(N^3\big)$. In the non-linear case, the dimensionality of the data changes from D to N. In this case, the complexity becomes $\mathcal{O}\big(4N^3\big)$. In terms of $\mathcal{O}$ notation, the complexity in the non-linear case stays at $\mathcal{O}\big(N^3\big)$.

3.4.2.3 Complexity Analysis of Spectral Regression-based Update

The main computational steps in the spectral regression-based update are computing the generalized eigenvalue problem and obtaining the projection matrix $\mathbf{Q}$ in a least-square sense.

1. Solving the generalized eigenvalue problem $\mathbf{L}_\alpha \mathbf{t} = \upsilon \mathbf{L}_x \mathbf{t}$ has the complexity of $\mathcal{O}(N^3)$.
2. Solving $\mathbf{Q} = \mathbf{T}^\intercal \mathbf{X}^\intercal (\mathbf{X}\mathbf{X}^\intercal + \eta \mathbf{I})^{-1}$ involves the following steps: Computing $\mathbf{X}\mathbf{X}^\intercal$ has the complexity of $\mathcal{O}(D^2 N)$. The complexity of multiplying η with each element of $\mathbf{I}$ is $\mathcal{O}(D^2)$. Adding $\mathbf{X}\mathbf{X}^\intercal$ with $\eta\mathbf{I}$ has the complexity of $\mathcal{O}(D^2)$. Taking inverse of $(\mathbf{X}\mathbf{X}^\intercal + \eta\mathbf{I})$ has the complexity of D^3 and multiplying the rest of matrices has the complexity of $dND + D^3$. Adding the complexities of all these steps, the complexity of $\mathbf{Q} = \mathbf{T}^\intercal \mathbf{X}^\intercal (\mathbf{X}\mathbf{X}^\intercal + \eta\mathbf{I})^{-1}$ becomes $\mathcal{O}(D^2 N + 2D^2 + D^3 + dND + D^3)$.

Adding the above two complexities and the complexity of the main algorithm, and assuming that the dimensionality D of the data is always lower than the number of

samples N, the total complexity in terms of big $\mathcal{O}$ notation becomes $\mathcal{O}(N^3)$. In the non-linear case, the steps involved in NPT have the complexity of $\mathcal{O}(N^3)$. Moreover the dimensionality changes from D to N. Thus, the complexity increases, but in terms of the big $\mathcal{O}$ notation, still stays as $\mathcal{O}(N^3)$.

3.4.3 Results and Discussion

We evaluated different variants derived from the proposed framework over nine different datasets. The datasets used in the experiments are Seeds, Qualitative bankruptcy, Somerville happiness, Liver, Iris, Ionosphere, Sonar, Heart (from the UCI machine learning repository [70]) and MNIST [75] with original dimensionalities of 7, 6, 6, 6, 4, 34, 60, 13, and 784, respectively. MNIST has 10 classes, Seeds and Iris datasets are ternary, while the rest of the datasets are binary. The summary of Seeds, Qualitative bankruptcy, Somerville happiness, Iris, Ionosphere, and Sonar is given in Table 3.3. The Liver dataset contains 145 samples from the Disorder Present (DP) category and 200 samples from the Disorder Absent (DA) category. The Heart dataset contains 139 samples from the DP category and 164 samples from the DA category. MNIST dataset contains 5923, 6742, 5958, 6131, 5842, 5421, 5918, 6265, 5851, 5949 samples in the training set for classes 0-9, respectively. The MNIST test set contains 980, 1135, 1032, 1010, 982, 892, 958, 1028, 974, and 1009 samples from the corresponding classes (0-9). In our experiments, we select 10% of the data from MNIST while keeping the representation of each class in the train and test sets similar to the original train and test split in the dataset.

All datasets were converted into OCC datasets by considering a single class as the positive class at a given time. The splits of the datasets were selected as described in Section 3.1.7. We used a 5-fold cross-validation technique over the training set to select the hyperparameters during the training phase. More details about the experimental setup, hyperparameter tuning and details of extensive results are provided in the corresponding publication [P4]. The best performing linear and non-linear variants are compared against SSVDD and ESSVDD, and the competing methods GESVDD, GESVM, SVDD, and OCSVM in Table 3.7 and 3.8. For all iterative methods, the total number of iterations was set to 5.

It is noticed that in both linear and non-linear methods, the gradient-based solution performs better in terms of GM than the spectral and spectral regression-

Table 3.7 GM results for linear and non-linear data description over different datasets, selected variants from the proposed framework vs. other one-class classification methods [P4].

Dataset	Seeds				Qualitative bankruptcy			Somerville happiness			Liver		
Target class	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.	DP	DA	Av.
Linear													
GESSVDD-kNN-GR-min	0.83	**0.94**	**0.95**	**0.91**	0.80	**0.46**	**0.63**	0.44	0.44	0.44	0.45	0.38	0.42
ESSVDD	**0.87**	0.92	0.90	0.90	0.90	0.12	0.51	**0.51**	0.39	0.45	0.35	0.38	0.37
SSVDD	0.85	0.93	**0.95**	**0.91**	0.90	0.17	0.53	0.49	0.43	0.46	0.32	0.34	0.33
ESVDD	0.79	0.87	0.87	0.84	**0.96**	0.19	0.58	0.42	0.41	0.41	0.35	**0.40**	0.38
SVDD	0.85	0.92	0.94	0.90	0.94	0.00	0.47	0.41	0.36	0.39	**0.50**	0.39	**0.45**
OCSVM	0.48	0.69	0.45	0.54	0.37	0.41	0.39	0.45	**0.53**	**0.49**	0.40	0.36	0.38
Non-Linear													
GESSVDD-kNN-SR-max	0.86	0.92	**0.96**	0.91	0.81	**0.71**	**0.76**	0.47	0.47	0.47	0.41	0.42	0.41
ESSVDD	0.83	0.91	0.90	0.88	0.92	0.28	0.60	**0.59**	0.39	0.49	0.40	0.49	0.45
SSVDD	**0.87**	**0.94**	0.94	**0.92**	0.94	0.46	0.70	0.47	0.35	0.41	0.37	0.39	0.38
ESVDD	0.81	0.88	0.87	0.85	0.00	0.00	0.00	0.00	0.31	0.16	0.43	0.54	0.49
SVDD	0.85	0.91	0.95	0.90	0.33	0.28	0.31	0.40	0.32	0.36	0.49	0.40	0.45
OCSVM	0.47	0.60	0.45	0.51	0.36	0.58	0.47	0.47	0.49	0.48	0.27	0.08	0.17
GESVDD-PCA	0.85	0.93	0.93	0.90	0.94	0.28	0.61	0.50	0.48	0.49	**0.51**	0.49	0.50
GESVDD-Sw	0.82	0.93	0.93	0.89	0.94	0.28	0.61	0.49	**0.50**	0.49	**0.51**	0.52	0.51
GESVDD-kNN	0.84	0.92	0.94	0.90	0.84	0.31	0.57	0.50	0.45	0.47	**0.51**	0.52	0.52
GESVM-pca	0.85	0.90	0.93	0.89	**0.95**	0.26	0.60	0.52	0.48	**0.50**	0.50	**0.55**	0.52
GESVM-Sw	0.85	0.90	0.91	0.89	0.93	0.20	0.57	0.55	0.41	0.48	0.50	0.51	0.51
GESVM-kNN	0.84	0.90	0.90	0.88	0.92	0.20	0.56	0.55	0.51	0.53	**0.51**	0.55	**0.53**

Dataset	Iris				Ionosphere			Sonar			Heart		
Target class	I-S	I-VC	S-V	Av.	I-B	I-G	Av.	S-R	S-M	Av.	DP	DA	Av.
Linear													
GESSVDD-kNN-GR-min	**0.97**	0.89	**0.91**	**0.92**	0.42	**0.92**	**0.67**	**0.54**	0.57	**0.56**	0.54	0.61	0.58
ESSVDD	0.93	0.82	0.89	0.88	0.36	0.90	0.63	0.52	**0.58**	0.55	0.53	**0.69**	**0.61**
SSVDD	0.96	**0.91**	0.90	**0.92**	0.12	0.78	0.45	0.51	0.55	0.53	**0.59**	0.62	**0.61**
ESVDD	0.89	0.85	0.86	0.87	0.33	0.88	0.61	0.00	0.03	0.02	0.56	0.62	0.59
SVDD	0.92	0.90	0.89	0.91	0.02	0.86	0.44	0.52	0.56	0.54	0.46	0.35	0.41
OCSVM	0.58	0.50	0.46	0.51	**0.49**	0.51	0.50	0.48	0.45	0.46	0.57	0.63	0.60
Non-Linear													
GESSVDD-kNN-SR-max	**0.94**	0.87	0.83	0.88	**0.67**	0.86	0.76	0.52	0.47	0.49	0.42	0.43	0.42
ESSVDD	**0.94**	0.88	0.89	0.90	0.64	0.89	**0.77**	0.54	0.55	0.54	0.38	0.37	0.37
SSVDD	**0.94**	0.92	**0.90**	**0.92**	0.40	0.89	0.65	0.48	0.47	0.47	0.53	0.49	0.51
ESVDD	0.68	0.84	0.83	0.78	0.37	0.88	0.63	0.55	0.52	0.53	0.34	0.27	0.31
SVDD	0.92	0.92	0.88	0.90	0.21	0.85	0.53	0.53	0.59	0.56	0.53	0.55	0.54
OCSVM	0.56	0.26	0.55	0.46	0.52	0.47	0.49	0.47	0.55	0.51	0.20	0.23	0.21
GESVDD-PCA	0.83	0.92	0.89	0.88	0.38	0.88	0.63	**0.55**	**0.60**	0.57	0.68	**0.74**	**0.71**
GESVDD-Sw	0.89	0.87	0.90	0.89	0.36	0.90	0.63	0.53	0.54	0.54	0.68	0.73	0.70
GESVDD-kNN	0.83	0.91	0.89	0.88	0.34	0.89	0.62	0.54	**0.60**	**0.57**	**0.70**	0.72	0.71
GESVM-pca	0.90	0.90	**0.90**	0.90	0.38	**0.91**	0.64	0.52	0.61	0.57	0.66	0.71	0.68
GESVM-Sw	0.89	**0.93**	0.88	0.90	0.45	0.90	0.67	0.54	0.59	0.57	0.67	0.70	0.68
GESVM-kNN	0.89	0.89	0.89	0.89	0.41	0.88	0.65	0.54	0.58	0.56	0.67	0.72	0.70

Table 3.8 GM results for linear and non-linear data description over MNIST dataset, selected variants from the proposed framework vs. other one-class classification methods [P4].

Dataset	Mnist										
Target class	0	1	2	3	4	5	6	7	8	9	Av.
Linear											
GESSVDD-kNN-GR-min	0.40	0.84	0.33	0.47	0.60	0.38	0.69	0.53	0.51	0.58	0.53
GESSVDD-I-GR-min (ESSVDD)	0.38	0.83	0.31	0.46	0.47	0.34	0.62	0.65	0.40	0.50	0.50
GESSVDD-0-GR-min (SSVDD)	0.41	0.81	0.29	0.39	0.45	0.31	0.57	0.52	0.40	0.44	0.46
ESVDD	0.00	0.81	0.00	0.00	0.00	0.00	0.06	0.22	0.00	0.16	0.13
SVDD	0.47	0.55	**0.51**	0.50	0.51	**0.52**	0.45	0.57	0.49	0.51	0.51
OCSVM	**0.57**	**0.92**	0.47	**0.52**	**0.64**	0.41	**0.73**	**0.74**	**0.53**	**0.63**	0.62
Non-Linear											
GESSVDD-kNN-SR-max	0.38	0.53	0.16	0.34	0.49	0.46	0.48	0.43	0.31	0.50	0.41
GESSVDD-I-GR-min (ESSVDD)	0.36	0.34	0.18	0.09	0.19	0.52	0.46	0.43	0.36	0.21	0.31
GESSVDD-0-GR-min (SSVDD)	0.60	0.34	0.48	0.39	0.43	0.49	0.43	0.35	0.44	0.17	0.41
ESVDD	0.54	0.19	0.34	0.14	0.39	0.52	0.42	0.32	0.17	0.36	0.34
SVDD	0.15	0.05	0.63	0.14	0.12	0.17	0.11	0.13	0.13	0.13	0.18
OCSVM	0.59	0.69	0.56	0.46	0.61	0.64	0.66	0.56	0.53	0.66	0.60
GESVDD-PCA	**0.92**	**0.96**	**0.75**	0.74	0.84	**0.73**	0.86	0.86	0.73	0.85	0.82
GESVDD-Sw	**0.92**	**0.96**	**0.75**	0.74	0.84	0.72	0.00	0.85	0.71	0.85	0.74
GESVDD-kNN	0.91	**0.96**	**0.75**	0.74	0.84	0.72	0.86	0.86	0.73	0.85	0.82
GESVM-pca	0.90	0.95	**0.75**	0.74	**0.87**	0.71	0.89	**0.86**	**0.76**	**0.86**	**0.83**
GESVM-Sw	0.90	0.95	**0.75**	0.74	0.85	0.66	0.87	0.84	0.75	0.85	0.82
GESVM-kNN	0.90	0.95	0.74	**0.76**	**0.87**	0.71	**0.89**	0.85	0.73	0.85	0.82

based solutions in most cases. Overall in linear methods, employing the kNN graph in the subspace produces better results than other competing graphs. GESSVDD-kNN-GR-min performs best over 5 and second-best over 2 out of 9 datasets. In non-linear methods, a more varying performance is noticed. However, we recommend GESSVDD-kNN-SR-max in the non-linear case because we did not observe any case where the variant would under-perform with a significant margin. Comparing the minimization/maximization, it can be observed that maximizing $\mathbf{L}_b$ and minimizing $\mathbf{L}_w$ typically provide better results. This is expected because an intuitive assumption is that between-cluster scatter should be maximized and within-cluster scatter minimized.

3.5 Macroinvertebrate Taxa Identification With One-Class Classification

In this section, we demonstrate the capability of OCC methods to improve the performance of deep CNNs over a specific domain application. We selected the task of macroinvertebrate taxa identification both due to its practical importance and because the selected large-scale dataset with a very imbalanced class distribution is well suited for the intended demonstration of OCC capabilities. The dataset contains specimens belonging to rare species which are challenging to model by traditional machine learning techniques.

Benthic macroinvertebrate monitoring is a crucial task for understanding ecological changes. The presence of benthic macroinvertebrates is widely used for environmental decision making, and ecological status assessment in aquatic ecosystems [76]. Traditionally, taxa identification is carried out manually and requires a lot of tedious work. Alternative automated solutions based on training predictive machine learning models using datasets of images are being developed [77]. However, solutions based on machine learning require correctly labeled and ample training data from all taxa to be identified. Due to the unavailability of well-balanced big datasets, more work is still needed to enable the automatic identification of benthic macroinvertebrates in real-life biomonitoring. Recently, larger publicly available datasets have been introduced [78, 79] along with developing more practical and accurate imaging devices [78, 80]. However, one of the main challenges in adopting automated or semi-automated techniques in real-life biomonitoring stems from the imbalanced nature of the dataset. The number of specimens from the rarest taxa in the datasets remains too low to efficiently learn with CNNs, while the taxa may be important from the biodiversity monitoring perspective. At the same time, specimens from the same taxa can exhibit significantly different features, while different taxa can be very similar, leading to a fine-grained classification problem.

In order to better identify some rare species, we proposed a two-step solution. In step 1, the features are extracted via a state-of-the-art deep neural network. In step 2, we create a model based on OCC for specific taxa. The proposed framework is shown in Fig. 3.5. In the proposed framework, the samples of macroinvertebrates are first collected, and then the preprocessing is carried out. The preprocessed images

are used as input to a deep CNN for identification in the first phase. The features are also extracted from the network, and they are transformed into a lower-dimensional space so that the OCC method can focus on the key features. The OCC models are trained for specific taxa, and each model identifies a rare species of interest. The instances predicted belonging to the target class are sent for manual identification to then be biologically assessed.

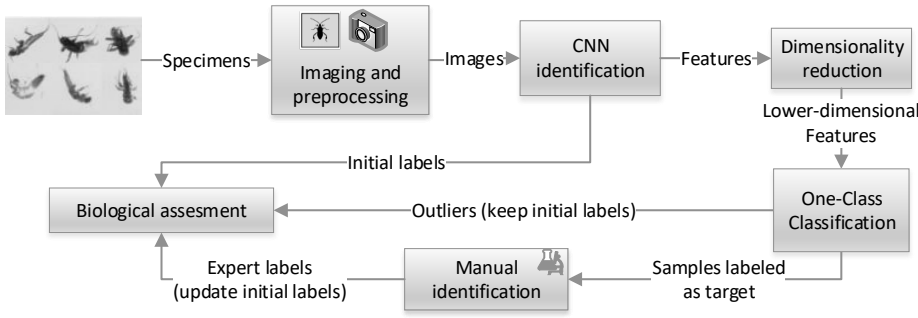

Figure 3.5 Proposed pipeline for boosting rare benthic macroinvertebrates taxa identification with one-class classification [P5].

3.5.1 Dataset and Experimental Setup

In our experiments, we used the *FIN-Benthic2* dataset[4] [79]. The dataset consists of 460004 images of 9631 benthic macroinvertebrate specimens belonging to 39 different taxa. The dataset is very imbalanced, and the images per taxon range from 490 to 44240. Ten different splits of the images are provided for training, validation, and testing in the dataset. In this work, we used Split 1 for our experiments. For proof of concept, we selected three different taxa, namely, *Capnopsis schilleri*, *Nemoura cinerea*, and *Leuctra nigra*, as our rare species for the target classes. The selected taxa are rare in the dataset, and they are picked as a proof-of-concept, not based on their environmental significance. Table 3.9 summarizes the selected classes.

Our experiments used a VGG16 network as the base model, pre-trained on ImageNet and fine-tuned on the FIN-Benthic2 dataset. On top of the VGG16 convolutional output, we added two dense layers with 4069 and 39 neurons, respectively.

[4]https://etsin.fairdata.fi/dataset/a11cdc26-b9d0-4af1-9285-803d65a696a3

Table 3.9 Image numbers in Split 1 of FIN-Benthic2 dataset [P5].

	Train	Validation	Test
Capnopsis schilleri	600	100	350
Nemoura cinerea	650	100	50
Leuctra nigra	1100	50	200
Whole dataset	321407	45912	92685

The first layer used ReLU activation, while the second (output) layer used soft-max activation. To avoid overfitting, we also applied dropout over the dense layers. For the dimensionality reduction block, we used PCA and kept the first 100 principal components as an input to the OCC module. We used OCSVM, SVDD, SSVDD, SSVDDr1, and SSVDDr2 in our experiments. Where SSVDDr1 is the variant of SSVDD, which uses all the instances in the regularization terms, and SSVDDr2 refers to using only support vectors in the regularization term ψ (See 3.1.2).

3.5.2 Results and Discussion

The obtained results are summarized in Table 3.10. We report the results in terms of four different evaluating metrics. The TPR refers to the sensitivity of the model and measures the fraction of correctly identified target samples. The GM considers the TPR and TNR; it is the square root of the product of the TPR and TNR. The TNR refers to the specificity of the model and measures the fraction correctly identified as outliers. We report the TP, which corresponds to the total number of correctly identified target samples. The total number of samples predicted as positives for manual identification is reported as Positive Detected (PD). PD is calculated as the sum of the TP's and FP's.

It is clear from the experimental results that one-class classifiers can indeed complement a deep neural network. The best performing classification model in terms of GM was regularized linear SSVDD, where only the support vectors are selected in the regularization term. The proposed framework allows dividing the tasks between human experts and machines, thus improving the overall efficiency. In this work, we met the objective of assessing the performance of OCC methods over a specific domain application.

The proposed taxa identification framework is an example of demonstrating the

capability of OCC methods for improving the performance of CNN-based algorithm over imbalanced datasets. The proposed framework can be used for similar applications of dividing the identification task between machines and human experts. The framework can be further enhanced by considering the classification confidences of both the CNN and one-class classifiers. The proposed framework considers unimodal data, while there are multiple images of a single specimen in the given dataset. In the future, the framework can be adapted to using multimodal OCC techniques.

Table 3.10 One-class classifier results for different target species [P5].

	Capnopsis schilleri				*Nemoura cinerea*				*Leuctra nigra*			
	TPR	GM	TP	PD	TPR	GM	TP	PD	TPR	GM	TP	PD
CNN classification												
VGG16	0.046	0.214	16	101	0.020	0.141	1	39	0.170	0.412	34	174
Linear one-class classification												
OCSVM	0.906	0.613	317	54367	0.660	0.357	33	74739	0.625	0.437	125	64304
SVDD	0.346	0.586	121	701	0.280	0.525	14	1422	0.730	0.832	146	4860
SSVDD	0.557	0.740	195	1893	0.480	0.676	24	4385	0.805	0.838	161	11910
SSVDDr1	0.609	0.773	213	1977	0.340	0.567	17	5209	0.805	0.837	161	12103
SSVDDr2	0.706	**0.825**	247	3573	0.560	**0.702**	28	11178	0.855	**0.876**	171	9625
EOCC	0.620	0.779	217	2077	0.380	0.601	19	4487	0.820	0.848	164	11497
Non-linear one-class classification												
OCSVM	0.034	0.185	12	87	0.000	0.000	0	51	0.220	0.469	44	102
SVDD	0.331	0.574	116	658	0.300	0.543	15	1441	0.730	0.832	146	4904
SSVDD	0.503	0.705	176	1169	0.440	0.649	22	3890	0.815	0.853	163	10085
SSVDDr1	0.540	0.730	189	1404	0.400	0.622	20	3138	0.780	0.854	156	6221
SSVDDr2	1.000	0.000	350	92685	0.220	0.465	11	1762	0.995	0.003	199	92683
EOCC	0.503	0.705	176	1160	0.300	0.542	15	1883	0.775	0.851	155	6245

4 CONCLUSIONS

Subspace learning and OCC are two different active research areas for many targeted applications. Subspace learning aims to map high-dimensional data to a lower-dimensional space by preserving the crucial information and finding an optimized subspace for the task at hand. PCA is a classic example of subspace learning and has been used in many real-world problems to obtain critical features in the latent space. On the other hand, traditional OCC methods aim at obtaining a data description in a given feature space by using data only from one category. SVDD is a typical example of a method obtaining a data description by finding a minimum hyperspherical boundary around a target class data.

In this thesis, we focused on developing new techniques of subspace learning for OCC. In [P1], we proposed a gradient-based subspace learning method SSVDD for OCC. SSVDD defines a data description in a lower-dimensional space optimized for OCC. It forms the basis of this thesis and the initial research question of creating a methodology where the subspace learning and learning an OCC model complement each other for improved OCC performance was answered, and hence the initial objective was met. However, SSVDD is limited by the rigid hyperspherical boundary. The hyperspherical boundary can result in superfluous regions that do not contain any instances from the target class. The proposed method in [P1] was further enhanced by using a more flexible boundary in [P2] by taking into account the covariance of the data in the subspace. ESSVDD developed in [P2], converges faster than [P1] and generalizes SSVDD by using a hyperellipsoid instead of a hypersphere. In [P2], the objective of further optimizing the data description by considering alternatives for the traditional hyperspherical description to improve the classification performance of OCC was met.

The third objective of the thesis was to infer a shared latent representation for multimodal data and exploit the relationship between the modalities in OCC. We met this objective in [P3], where we proposed MSSVDD, which is a novel technique

for transforming multimodal data into a shared sub-space optimized for OCC. In [P4], the objective of integrating different solutions into a unified framework was met. In [P4], GESSVDD is proposed, which allows incorporating different data relationships in the form of graph structures in the optimization process. The proposed framework also highlighted the similarities and differences between different methods. Moreover, as an alternative to the initial proposed gradient-based solution, spectral and spectral regression-based solutions were proposed. The novel GESSVDD led to improved performance against the baselines and other competing methods. SSVDD and ESSVDD were shown to be special cases of the proposed framework.

In [P5], we demonstrated the capability of OCC methods to improve the performance of a deep CNN. We developed a framework to identify rare benthic macroinvertebrates. The proposed framework is useful for similar applications where one of the categories to be classified is rare in nature. We met the objective of accessing the performance of OCC methods over different domain applications in all the publications; however, in [P5], we emphasized the usage of OCC methods over a particular dataset in addition to developing the specific method. Based on the contributions of this thesis, we answered the research questions and met the objectives set at the beginning of the thesis.

To summarize, we introduced novel subspace learning methods for OCC. We guided the traditional SVDD towards a new unexplored era of jointly optimizing a subspace and data description. We unified different subspace learning techniques for OCC in a single framework and introduced different solutions. We provided linear and non-linear formulations for all the proposed methods and achieved better performance over standard datasets. We also used the proposed framework to identify rare benthic macroinvertebrates and showed the capability of OCC methods in general and the proposed SSVDD method in particular to complement deep neural networks.

While the proposed methods have shown promising results over datasets from various domains, there are some limitations, and the methods still fall short of meeting all the challenges in the field of OCC. OCC methods are employed to identify a novel pattern or behavior, and failing to identify them may have a high risk associated with it. We assumed an equal cost for all errors and used traditional evaluation metrics for all datasets from different domains. However, the cost of failing to detect a specific condition in terms of risk is very different for health diagnostic tests than

intrusion detection systems. There is a need to develop methods that incorporate the risk information in the optimization process of the OCC models. Moreover, the proposed methods cannot deal with the changing nature of the input data. In some real-world applications, such as environmental monitoring, the data changes its nature over time, and an outlier may represent the target class concept in the future. Thus adaptive techniques of dealing with non-stationary data for the proposed methods are also needed.

In the future, the proposed methods should be further enhanced to reduce their computational complexity. The proposed methods are iterative, and further research can help to increase the convergence speed of these methods. For example, using Newton's method rather than gradient descent in the proposed methods should be considered. We used an RBF kernel in all our experiments for non-linear data description, and there is no mechanism in place for identifying the best kernel function beforehand. We recommend using and analyzing other kernel types such as polynomial and sigmoid kernels as well. Using other kernel types instead of RBF may provide better computational complexity or classification accuracy. GESSVDD proposed in [P4] can be further enhanced through exploiting multiple graphs by combining the geometric data relationships using a weight parameter. The proposed framework in [P5] can be extended to an end-to-end deep learning-based solution by embedding the OCC methods inside the deep learning methodology.

The developed methods in this thesis use support vector-based techniques for encapsulating the training data. We suggest investigating similar methodologies for other OCC approaches such as density estimation and reconstruction-based methods. Moreover, the multimodal approach developed in this thesis uses hyperspherical encapsulation of the data, which can be further improved by using hyperellipsoid or graph-based structures in the subspace optimized for multimodal data.

The proposed methods can be used for target (or anomaly) detection in cases where it is often difficult or impossible to obtain data from one of the categories, such as fraud detection, abnormal event detection in a multi-sensory environment, and anomaly detection in medical diagnosis. Most possible situations representing fraud, abnormal events, and anomalies are difficult to collect for training the model. In medical diagnosis, the data from non-healthy subjects are either hard or simply impossible to obtain. For example, in mammography for cancer detection, the specific target class recognition of cognitive brain functions, in the interstitial lung diseases

categorization or detecting nosocomial infections through clinical data, a representative training set representing non-healthy data is challenging to obtain. In such cases, the proposed methods can be used for training a OCC model.

REFERENCES

[1] M. Erfani, F. Shoeleh and A. A. Ghorbani. Financial Fraud Detection using Deep Support Vector Data Description. *IEEE International Conference on Big Data*. IEEE. 2020, 2274–2282.

[2] X. Zong, L. Zhang, J. Du, L. Wei and Q. Huang. Abnormal Event Detection in Video Based on SVDD. *IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*. Vol. 1. IEEE. 2019, 368–371.

[3] G. Cohen, M. Hilario, H. Sax, S. Hugonnet, C. Pellegrini and A. Geissbuhler. An application of one-class support vector machines to nosocomial infection detection. *MEDINFO*. IOS Press. 2004, 716–720.

[4] T. Petsche, A. Marcantonio, C. Darken, S. Hanson, G. Kuhn and N. Santoso. A neural network autoassociator for induction motor failure prediction. *Advances in Neural Information Processing Systems* 8 (1995), 924–930.

[5] M. Dellomo. Helicopter gearbox fault detection: a neural network based approach. *Journal of Vibration and Acoustics* 121 (1999), 265–272.

[6] A. Anaissi, N. L. D. Khoa, T. Rakotoarivelo, M. M. Alamdari and Y. Wang. Adaptive online one-class support vector machines with applications in structural health monitoring. *ACM Transactions on Intelligent Systems and Technology* 9.6 (2018), 1–20.

[7] J. Rodriguez-Ruiz, J. I. Mata-Sanchez, R. Monroy, O. Loyola-Gonzalez and A. Lopez-Cuevas. A one-class classification approach for bot detection on Twitter. *Computers and Security* 91 (2020), 101715.

[8] Z. Shi, M. Huang, C. Zhao, L. Huang, X. Du and Y. Zhao. Detection of LSSUAV using hash fingerprint based SVDD. *IEEE International Conference on Communications*. IEEE. 2017, 1–5.

[9] F. Sohrab. Railway vehicle detection from audio recordings using one-class classification. MA thesis. Sabancı University, 2016.

[10] Q. Wei, Y. Ren, R. Hou, B. Shi, J. Y. Lo and L. Carin. Anomaly detection for medical images based on a one-class classification. *Medical Imaging: Computer-Aided Diagnosis*. Vol. 10575. International Society for Optics and Photonics. 2018, 105751M.

[11] L. M. Manevitz and M. Yousef. One-class SVMs for document classification. *Journal of Machine Learning Research* 2.Dec (2001), 139–154.

[12] G. Cohen, H. Sax and A. Geissbühler. Novelty Detection using One-class Parzen Density Estimator. An Application to Surveillance of Nosocomial Infections. *International Congress of the European Federation for Medical Informatics*. IOS Press, 2008, 21–26.

[13] W. Khreich, B. Khosravifar, A. Hamou-Lhadj and C. Talhi. An anomaly detection system based on variable N-gram features and one-class SVM. *Information and Software Technology* 91 (2017), 186–197.

[14] L. Yin, H. Wang and W. Fan. Active learning based support vector data description method for robust novelty detection. *Knowledge-Based Systems* 153 (2018), 40–52.

[15] D. M. J. Tax. One-class classification: Concept learning in the absence of counter examples. PhD thesis. Delft University of Technology, 2002.

[16] K. Hempstalk, E. Frank and I. H. Witten. One-class classification by combining density and class probability estimation. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer. 2008, 505–519.

[17] S. M. Erfani, S. Rajasegarar, S. Karunasekera and C. Leckie. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognition* 58 (2016), 121–134.

[18] L. Ruff et al. Deep one-class classification. *International Conference on Machine Learning*. PMLR. 2018, 4393–4402.

[19] D. M. Tax and R. P. Duin. Support vector data description. *Machine Learning* 54.1 (2004), 45–66.

[20] V. Mygdalis, A. Iosifidis, A. Tefas and I. Pitas. Graph embedded one-class classifiers for media data classification. *Pattern Recognition* 60 (2016), 585–595.

[21] S. S. Khan and M. G. Madden. One-class classification: taxonomy of study and review of techniques. *The Knowledge Engineering Review* 29.3 (2014), 345–374.

[22] K. Lee, D.-W. Kim, K. H. Lee and D. Lee. Density-induced support vector data description. *IEEE Transactions on Neural Networks* 18.1 (2007), 284–289.

[23] T. Kohonen and P. Somervuo. Self-organizing maps of symbol strings. *Neurocomputing* 21.1-3 (1998), 19–30.

[24] C. Fraley and A. E. Raftery. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* 97.458 (2002), 611–631.

[25] B. Krawczyk, M. Woźniak and B. Cyganek. Clustering-based ensembles for one-class classification. *Information Sciences* 264 (2014), 182–195.

[26] B. Scholkopf, R. C. Williamson, A. Smola and J. Shawe-Taylor. SV Estimation of a Distribution's Support. *Neural Information Processing Systems*. MIT Press, 2000, 582–588.

[27] P. Perera and V. M. Patel. Learning deep features for one-class classification. *IEEE Transactions on Image Processing* 28.11 (2019), 5450–5463.

[28] S. Goyal, A. Raghunathan, M. Jain, H. V. Simhadri and P. Jain. DROCC: Deep robust one-class classification. *International Conference on Machine Learning*. PMLR. 2020, 3711–3721.

[29] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth and G. Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. *International Conference on Information Processing in Medical Imaging*. Springer. 2017, 146–157.

[30] J. Platt. *Sequential minimal optimization: A fast algorithm for training support vector machines*. Tech. rep. MSR-TR-98-14. Microsoft Research, 1998.

[31] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2.3 (2011), 1–27.

[32] J. Wang, W. Liu, K. Qiu, H. Xiong and L. Zhao. Dynamic hypersphere SVDD without describing boundary for one-class classification. *Neural Computing and Applications* 31.8 (2019), 3295–3305.

[33] T. Kenaza, K. Bennaceur and A. Labed. An efficient hybrid SVDD/clustering approach for anomaly-based intrusion detection. *Symposium on Applied Computing*. ACM. 2018, 435–443.

[34] V. Mygdalis, A. Iosifidis, A. Tefas and I. Pitas. Semi-supervised subclass support vector data description for image and video classification. *Neurocomputing* 278 (2018), 51–61.

[35] M. Rahmanimanesh, J. A. Nasiri, S. Jalili and N. M. Charkari. Adaptive three-phase support vector data description. *Pattern Analysis and Applications* 22.2 (2019), 491–504.

[36] B. Liu, Y. Xiao, L. Cao, Z. Hao and F. Deng. SVDD-based outlier detection on uncertain data. *Knowledge and Information Systems* 34.3 (2013), 597–618.

[37] Y. Forghani, S. Effati, H. S. Yazdi and R. S. Tabrizi. Support Vector Data Description by using hyper-ellipse instead of hyper-sphere. *International eConference on Computer and Knowledge Engineering*. 2011, 22–27.

[38] Y. Xiao et al. Multi-sphere Support Vector Data Description for Outliers Detection on Multi-distribution Data. *International Conference on Data Mining Workshops*. 2009, 82–87.

[39] G. Huang, H. Chen, Z. Zhou, F. Yin and K. Guo. Two-class support vector data description. *Pattern Recognition* 44.2 (2011), 320–329.

[40] R. E. Bellman. *Adaptive control processes*. Princeton University Press, 2015.

[41] M. U. Chaudhry and J.-H. Lee. Feature selection for high dimensional data using Monte Carlo tree search. *IEEE Access* 6 (2018), 76036–76048.

[42] Z. Zhao, L. Wang, H. Liu and J. Ye. On similarity preserving feature selection. *IEEE Transactions on Knowledge and Data Engineering* 25.3 (2011), 619–632.

[43] N. Zhou, Y. Xu, H. Cheng, J. Fang and W. Pedrycz. Global and local structure preserving sparse subspace learning: An iterative approach to unsupervised feature selection. *Pattern Recognition* 53 (2016), 87–101.

[44] F. Z. Kermani, E. Eslami and F. Sadeghi. Global Filter–Wrapper method based on class-dependent correlation for text classification. *Engineering Applications of Artificial Intelligence* 85 (2019), 619–633.

[45] R. Vijayanand and D. Devaraj. A Novel Feature Selection Method Using Whale Optimization Algorithm and Genetic Operators for Intrusion Detection System in Wireless Mesh Network. *IEEE Access* 8 (2020), 56847–56854.

[46] K. K. Bharti and P. K. Singh. Hybrid dimension reduction by integrating feature selection with feature extraction method for text clustering. *Expert Systems with Applications* 42.6 (2015), 3105–3114.

[47] F. Moslehi and A. Haeri. A novel hybrid wrapper–filter approach based on genetic algorithm, particle swarm optimization for feature subset selection. *Journal of Ambient Intelligence and Humanized Computing* 11.3 (2020), 1105–1127.

[48] H. Liu, M. Zhou and Q. Liu. An embedded feature selection method for imbalanced data classification. *IEEE/CAA Journal of Automatica Sinica* 6.3 (2019), 703–715.

[49] Q. Gu, Z. Li and J. Han. Joint Feature Selection and Subspace Learning. *International Joint Conference on Artificial Intelligence*. IJCAI/AAAI, 2011, 1294–1299.

[50] I. T. Jolliffe and J. Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2065 (2016), 20150202.

[51] H. Wang, S. Yan, D. Xu, X. Tang and T. Huang. Trace Ratio vs. Ratio Trace for Dimensionality Reduction. *IEEE Conference on Computer Vision and Pattern Recognition*. 2007, 1–8.

[52] A. M. Martinez and A. C. Kak. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23.2 (2001), 228–233.

[53] H. Xie, J. Li, Q. Zhang and Y. Wang. Comparison among dimensionality reduction techniques based on Random Projection for cancer classification. *Computational Biology and Chemistry* 65 (2016), 165–172.

[54] H. Liu and W.-S. Chen. A novel random projection model for Linear Discriminant Analysis based face recognition. *2009 International Conference on Wavelet Analysis and Pattern Recognition*. 2009, 112–117. DOI: 10.1109/ICWAPR.2009.5207431.

[55] H. Sahbi. Kernel PCA for similarity invariant shape recognition. *Neurocomputing* 70.16-18 (2007), 3034–3045.

[56] X.-L. Xu, C.-X. Ren, R.-C. Wu and H. Yan. Sliced Inverse Regression With Adaptive Spectral Sparsity for Dimension Reduction. *IEEE Transactions on Cybernetics* 47.3 (2017), 759–771.

[57] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19.7 (1997), 711–720.

[58] H. Cai, V. W. Zheng and K. C.-C. Chang. A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE Transactions on Knowledge and Data Engineering* 30.9 (2018), 1616–1637.

[59] N. Satish et al. Navigating the maze of graph analytics frameworks using massive graph datasets. *SIGMOD International Conference on Management of Data*. ACM, 2014, 979–990.

[60] Y. Bengio, A. Courville and P. Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.8 (2013), 1798–1828.

[61] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang and S. Lin. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.1 (2006), 40–51.

[62] D. Cai, X. He and J. Han. *Spectral regression for dimensionality reduction*. Tech. rep. UIUCDCS-R-2007-2856. Computer Science Department, University of Illinois, 2007.

[63] D. Cai, X. He and J. Han. SRDA: An efficient algorithm for large-scale discriminant analysis. *IEEE Transactions on Knowledge and Data Engineering* 20.1 (2007), 1–12.

[64] K. Chumachenko, J. Raitoharju, A. Iosifidis and M. Gabbouj. Speed-up and multi-view extensions to subclass discriminant analysis. *Pattern Recognition* 111 (2021), 107660.

[65] B. Scholkopf and A. J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT Press, 2001.

[66] K. Wang, H. Xiao and Y. Fu. Ellipsoidal support vector data description in kernel PCA subspace. *International Conference on Digital Information Processing, Data Mining, and Wireless Communications*. IEEE. 2016, 13–18.

[67] N. Kwak. Nonlinear projection trick in kernel methods: An alternative to the kernel trick. *IEEE Transactions on Neural Networks and Learning Systems* 24.12 (2013), 2113–2119.

[68] S. Zheng. Smoothly approximated support vector domain description. *Pattern Recognition* 49 (2016), 55–64.

[69] A. Sharma, K. K. Paliwal, S. Imoto and S. Miyano. Principal component analysis using QR decomposition. *International Journal of Machine Learning and Cybernetics* 4 (2013).

[70] D. Dua and C. Graff. *UCI Machine Learning Repository*. 2017. URL: `http://archive.ics.uci.edu/ml`.

[71] *One-class datasets, Pattern Recognition Lab*. URL: `http://homepage.tudelft.nl/n9d04/occ/index.html`.

[72] T. Elgamal and M. Hefeeda. Analysis of PCA algorithms in distributed environments. *ArXiv Preprint arXiv:1503.05214* (2015).

[73] Y. Li, F. Nie, H. Huang and J. Huang. Large-Scale Multi-View Spectral Clustering via Bipartite Graph. *AAAI Conference on Artificial Intelligence*. 2015, 2750–2756.

[74] H. Hoffmann. Kernel PCA for novelty detection. *Pattern Recognition* 40.3 (2007), 863–874. ISSN: 0031-3203.

[75] L. Deng. The mnist database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine* 29.6 (2012), 141–142.

[76] D. Buchner et al. Analysis of 13,312 benthic invertebrate samples from German streams reveals minor deviations in ecological status class between abundance and presence/absence data. *Plos One* 14.12 (2019), e0226547.

[77] T. T. Høye et al. Deep learning and computer vision will transform entomology. *Proceedings of the National Academy of Sciences* 118.2 (2021). ISSN: 0027-8424.

[78] J. Raitoharju et al. Benchmark database for fine-grained image classification of benthic macroinvertebrates. *Image and Vision Computing* 78 (2018).

[79] J. Ärje et al. Human experts vs. machines in taxa recognition. *Signal Processing: Image Communication* (2020), 115917.

[80] J. Ärje et al. Automatic image-based identification and biomass estimation of invertebrates. *Methods in Ecology and Evolution* 11.8 (2020), 922–931.

ERRATA FOR PUBLICATIONS

Publication II:
Eq. (17) should be as follows

$$\frac{\partial L}{\partial \xi_i} = 0 \Rightarrow \qquad C - \alpha_i - \gamma_i = 0. \tag{17}$$

Eq. (34) should be as follows

$$R^2 = (\mathbf{E}^{-\frac{1}{2}}\mathbf{Q}\mathbf{s})^\intercal \mathbf{E}^{-\frac{1}{2}}\mathbf{Q}\mathbf{s} - 2(\mathbf{E}^{-\frac{1}{2}}\mathbf{Q}\mathbf{s})^\intercal \mathbf{u} + \mathbf{u}^\intercal \mathbf{u}, \tag{34}$$

PUBLICATIONS

Publication I

Subspace Support Vector Data Description

F. Sohrab, J. Raitoharju, M. Gabbouj and A. Iosifidis
International Conference on Pattern Recognition, 2018, pp. 722-727
DOI: 10.1109/ICPR.2018.8545819

Subspace Support Vector Data Description

Fahad Sohrab[†], Jenni Raitoharju[†], Moncef Gabbouj[†] and Alexandros Iosifidis[‡]
[†]Laboratory of Signal Processing, Tampere University of Technology, Finland
[‡]Department of Engineering, Electrical and Computer Engineering, Aarhus University, Denmark
{fahad.sohrab, jenni.raitoharju, moncef.gabbouj}@tut.fi, alexandros.iosifidis@eng.au.dk

***Abstract*—This paper proposes a novel method for solving one-class classification problems. The proposed approach, namely Subspace Support Vector Data Description, maps the data to a subspace that is optimized for one-class classification. In that feature space, the optimal hypersphere enclosing the target class is then determined. The method iteratively optimizes the data mapping along with data description in order to define a compact class representation in a low-dimensional feature space. We provide both linear and non-linear mappings for the proposed method. Experiments on 14 publicly available datasets indicate that the proposed Subspace Support Vector Data Description provides better performance compared to baselines and other recently proposed one-class classification methods.**

***Index Terms*—One-class Classification, Support Vector Data Description, Subspace Learning**

I. INTRODUCTION

In data classification, the overall goal is to define a model that can classify data into a predefined set of classes. During the training phase, the parameters of the classification model are estimated using samples belonging to the classes of interest. When the classification task involves two or more classes, model training requires a sufficient number of samples from each class and the corresponding class labels. However, in cases where we are interested in the distinction of a class from all other classes, the application of multi-class classification methods is usually not appropriate.

In case where both the class of interest (here-after called positive class) and all other classes (used to form the negative class) are sufficiently represented in the training set, class-specific models can be employed [1] [2] [3] [4], whereas, if the negative class is not sufficiently represented, one-class models should be applied. The main conceptual difference between class-specific and one-class models is that the former ones try to discriminate the positive class from every other class, while the latter ones try to describe the positive class without exploiting information related to negative samples. This is why one-class models can be applied in problems where only the positive class can be sufficiently sampled, while the negative one is either too rare or expensive to sample [5].

One-class classification problem has been tackled mainly by three approaches: density estimation, reconstruction and class boundary description [6]. For the density estimation approach, the Gaussian model, the mixture of Gaussians [7] and the Parzen density [8] are the most popular ones [9]. In reconstruction methods, the class is modelled by making some assumptions about the process which generates the target data. Some examples of reconstruction methods are based on K-means clustering, learning vector quantization and self-organizing maps [10]. In boundary description, a closed boundary around the target data is optimally formed. Support Vector Data Description (SVDD) [11] is one of the popular boundary methods used for solving one-class classification problems, by defining a hypersphere enclosing the target class. The hypersphere of SVDD can be made more flexible using kernel methods [11].

Other boundary methods have also been proposed for one-class classification. In [12], One-Class Support Vector Machine (OC-SVM) is proposed, in which the objective is to define the hyperplane that discriminates the data from the origin with maximum margin. It has also been proven that the solutions of SVDD and OC-SVM are equivalent for normalized data representations in the kernel space [13] [14]. In [15], Graph Embedded OC-SVM (GE-OC-SVM) and Graph Embedded SVDD (GE-SVDD) methods are introduced as extensions of [12] and [11], respectively. These methods incorporate geometric class information expressed by generic graph structures in OC-SVM and SVDD optimization that acts as a regularizer to their solution.

One-class classification has been used for many different applications. In [16], one-class classification is used for detecting faults in induction motors. In [17], one-class classification, particularly SVDD, is used in remote sensing for mapping a specific land-cover class, illustrated with an example of classification of a local government district in Cambridgeshire, England. In [18], an SVDD-based algorithm for target detection in hyperspectral images is developed. In [19], three different one-class classifiers, i.e., one-class Gaussian mixture, one-class SVM and one-class Nearest Neighbor are employed to label sound events as fall or part of the daily routine for elderly people based on sound signatures. In [20] one-class classification is used for video summarization based on human activities.

In this paper, we propose a novel method for generic one-class classification, namely Subspace Support Vector Data Description (S-SVDD). S-SVDD defines a model for the positive class in a low-dimensional feature space optimized for one-class classification. By allowing nonlinear data mappings, simple class models can be defined in the low-dimensional feature space that correspond to complex models in the original feature space. Such an approach allows us to simplify the information required for describing the class of interest,

while at the same time it can provide a good performance in nonlinear problems.

The rest of paper is organized as follows. The proposed S-SVDD method is described in detail in Section II. Experiments conducted in order to evaluate its performance on generic one-class classification problems are provided in Section III. Finally, conclusions are drawn in Section IV.

II. SUBSPACE SUPPORT VECTOR DATA DESCRIPTION

Let us assume that the class to be modeled is represented by a set of vectors $\mathbf{x}_i,\ i = 1, \ldots, N$, living in a D-dimensional feature space (i.e. $\mathbf{x}_i \in \mathbb{R}^D$). Subspace Support Vector Data Description (S-SVDD) tries to determine a d-dimensional feature space ($d \leq D$), in which the class can be optimally modeled. When linear projection is considered, the objective is to determine a matrix $\mathbf{Q} \in \mathbb{R}^{d \times D}$, such that:

$$\mathbf{y}_i = \mathbf{Q}\mathbf{x}_i,\ \ , i = 1, \ldots, N, \tag{1}$$

can be used in order to better model the class using a one-class classification model. We will describe how non-linear mappings can be exploited to this end using kernels in Subsection II-D.

The one-class classifier employed in this work is SVDD [11], which models the class by defining the hypesphere tightly enclosing the class. That is, given the data representation in the low-dimensional feature space $\mathbb{R}^d$, we want to determine the center of the class $\mathbf{a} \in \mathbb{R}^d$ and the corresponding radius R, by minimizing:

$$F(R, \mathbf{a}) = R^2 \tag{2}$$

such that all the training data are enclosed in the hypersphere, i.e.:

$$\|\mathbf{Q}\mathbf{x}_i - \mathbf{a}\|_2^2 \leq R^2,\ \ i = 1, \ldots, N. \tag{3}$$

In order to define a tighter class boundary (and possibly handle the situation of outliers in the training data), a relaxed version of the above criterion is solved by introducing a set of slack variables ξ_i. That is, the optimization function to minimize becomes:

$$F(R, \mathbf{a}) = R^2 + C \sum_{i=1}^{N} \xi_i \tag{4}$$

under the constraints that most of the training data should lie inside the hyper-sphere, i.e.:

$$\|\mathbf{Q}\mathbf{x}_i - \mathbf{a}\|_2^2 \leq R^2 + \xi_i,\ \ i = 1, \ldots, N \tag{5}$$

$$\xi_i \geq 0, \qquad i = 1, \ldots, N. \tag{6}$$

The parameter $C > 0$ in (4) is a regularization parameter which controls the trade-off between the volume of hyper-sphere and the training error caused by allowing outliers in the class description. C is inversely proportional to the fraction of the expected outliers in the training set. Increasing the value of C will allow more training samples to fall outside the class boundary.

The optimization problem in (4), under the constraints in (5) and (6) corresponds to the original SVDD optimization problem optimized with respect to an additional parameter $\mathbf{Q}$ that is used to define the optimal data representations for one-class classification. In order to find the optimal parameter values, we apply Lagrange-based optimization. The Lagrangian function is given by:

$$\begin{aligned} L(R, \mathbf{a}, \alpha_i, \xi_i, \gamma_i, \mathbf{Q}) &= R^2 + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \gamma_i \xi_i \\ &- \sum_{i=1}^{N} \alpha_i \Big(R^2 + \xi_i - \mathbf{x}_i^T \mathbf{Q}^T \mathbf{Q} \mathbf{x}_i \\ &+ 2\mathbf{a}^T \mathbf{Q}\mathbf{x}_i - \mathbf{a}^T \mathbf{a} \Big)) \end{aligned} \tag{7}$$

and should be maximized with respect to Lagrange multipliers $\alpha_i \geq 0, \gamma_i \geq 0$ and minimized with respect to radius R, center $\mathbf{a}$, slack variables ξ_i and projection matrix $\mathbf{Q}$.

By setting the partial derivative to zero, we get:

$$\frac{\partial L}{\partial R} = 0 \ \Rightarrow \ \sum_{i=1}^{N} \alpha_i = 1 \tag{8}$$

$$\frac{\partial L}{\partial \mathbf{a}} = 0 \ \Rightarrow \ \mathbf{a} = \sum_{i=1}^{N} \alpha_i \mathbf{Q}\mathbf{x}_i \tag{9}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \ \Rightarrow \ C - \alpha_i - \gamma_i = 0 \tag{10}$$

$$\frac{\partial L}{\partial \mathbf{Q}} = 0 \ \Rightarrow \ \mathbf{Q} = \left(\sum_{i=1}^{N} \alpha_i \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \left(\sum_{i=1}^{N} \alpha_i \mathbf{x}_i \mathbf{a}^T \right) \tag{11}$$

From (8)-(11), we can observe that the optimization parameters α_i and $\mathbf{Q}$ are inter-connected and, thus, they cannot be jointly optimized. In order to optimize (7) with respect to both α_i and $\mathbf{Q}$, we apply an iterative optimization process where, at each step, we fix one parameter and optimize the other, as will be described in the following subsections.

A. Class description

Given a data projection matrix $\mathbf{Q}$, the data description step follows the standard SVDD-based solution. That is, substituting (1), (8), (9) and (10) in (7) we obtain:

$$L = \sum_{i=1}^{N} \alpha_i \mathbf{y}_i^T \mathbf{y}_i - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \mathbf{y}_i^T \mathbf{y}_j \alpha_j. \tag{12}$$

Now, maximizing (12) gives the set of $\alpha_i,\ i = 1, \ldots, N$. The samples $\mathbf{y_i} = \mathbf{Q}\mathbf{x}_i$ corresponding to values $\alpha_i > 0$ are the support vectors defining the data description. The samples $\mathbf{y_i}$ corresponding to values $0 < \alpha_i < C$ are on the boundary of the corresponding hypersphere, while those outside the boundary will correspond to values $\alpha_i = C$. For the samples $\mathbf{y_i}$ inside the boundary, the corresponding values of α_i will be equal to zero [11]. Here we should note that whether a sample is a support vector or not, it is affected by the selection of the data projection matrix $\mathbf{Q}$, which is optimized based on the process described next.

B. SVDD-based subspace learning

After determining the optimal set of α_i, $i = 1, \ldots, N$, we optimize an augmented version of the Lagrangian function in (12):

$$L = \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^T \mathbf{Q}^T \mathbf{Q} \mathbf{x}_i - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \mathbf{x}_i^T \mathbf{Q}^T \mathbf{Q} \mathbf{x}_j \alpha_j + \beta \Psi, \quad (13)$$

where Ψ is a regularization term expressing the class variance in the low-dimensional space having the form:

$$\Psi = tr(\mathbf{Q}\mathbf{X}\lambda\lambda^T\mathbf{X}^T\mathbf{Q}^T). \quad (14)$$

β is a regularization parameter controlling the importance of the regularization term in the update and $tr(\cdot)$ is the trace operator. We additionally impose the constraint $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$, in order to obtain a orthogonal projection. $\lambda \in \mathbb{R}^N$ is a vector controlling the contribution of each training sample in the regularization term and can take the following values:

1) $\lambda_i = 0, i = 1, \ldots, N$: In this case the regularization term Ψ becomes obsolete and $\mathbf{Q}$ is optimized using (12). This case is referred to as Ψ_1 here-after.
2) $\lambda_i = 1$, $i = 1, \ldots, N$: In this case all training samples contribute to the regularization term Ψ equally. That is, all samples are used in order to describe the variance of the class. This case is referred to as Ψ_2 here-after.
3) $\lambda_i = \alpha_i$, $i = 1, \ldots, N$: In this case the samples belonging to the class boundary, as well as the outliers, are used to describe the class variance and regularize the update of $\mathbf{Q}$. This case is referred to as Ψ_3 here-after.
4) $\lambda_i = \alpha_i^C$, $i = 1, \ldots, N$, where $\alpha^C \in \mathbb{R}^N$ is a vector with values $\alpha_i^C = \alpha_i$, if $\mathbf{Q}\mathbf{x}_i$ is a support vector, and $\alpha_i^C = 0$, otherwise. This case is referred to as Ψ_4 here-after.

We update Q by using the gradient of L, i.e.:

$$\Delta L = 2 \sum_{i=1}^{N} \alpha_i \mathbf{Q}\mathbf{x}_i\mathbf{x}_i^T - 2 \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbf{Q}\mathbf{x}_i\mathbf{x}_j^T \alpha_i \alpha_j + \beta \Delta\Psi, \quad (15)$$

where $\Delta\Psi$ is the derivative of (14) with respect to $\mathbf{Q}$, i.e.:

$$\Delta\Psi = 2\mathbf{Q}\mathbf{X}\lambda\lambda^T\mathbf{X}^T. \quad (16)$$

C. S-SVDD optimization

In order to define both an optimized data projection matrix $\mathbf{Q}$ and the optimal data description in the resulting subspace, we iteratively apply the two processing steps described in subsections II-A and II-B, as described in Algorithm 1. The α_i's computed by maximizing (12) are used in (15) to update $\mathbf{Q}$ through a gradient step using a learning rate parameter η. The projection matrix $\mathbf{Q}$ is orthogonalized and normalized in each iteration to force the orthogonality constraint before applying the data mapping.

Algorithm 1: S-SVDD optimization

Input : $\mathbf{X}, \beta, \eta, d, C$
Output: $\mathbf{Q}$, R, $\boldsymbol{\alpha}$

// Initialize $\mathbf{Q}$
Random initialization of $\mathbf{Q}$;
Orthogonalize $\mathbf{Q}$ using QR decomposition;
Row normalize $\mathbf{Q}$ using l_2 norm;
Initialize $k = 1$;

while $k < k_{max}$ **do**

// SVDD in the subspace defined by $\mathbf{Q}$
Calculate $\mathbf{Y}$ using (1);
Calculate α_i, $i = 1, \ldots, N$ using (12);

// Update $\mathbf{Q}$ based on the SVDD solution
Calculate ΔL using (14)-(16);
Update $\mathbf{Q} \leftarrow \mathbf{Q} - \eta \Delta L$;

// Normalize the updated $\mathbf{Q}$
Orthogonalize $\mathbf{Q}$ using QR decomposition;
Row normalize $\mathbf{Q}$ using l_2 norm;
$k \leftarrow k + 1$

end

// SVDD in the optimized subspace
Calculate $\mathbf{Y}$ using (1);
Calculate α_i, $i = 1, \ldots, N$ using (12);

D. Non-linear data description

In order to exploit nonlinear mappings from $\mathbb{R}^D$ to $\mathbb{R}^d$ for one-class classification using the proposed S-SVDD, we follow the standard kernel-based learning approach [13]. That is, the original data representations $\mathbf{x}_i \in \mathbb{R}^D$, $i = 1, \ldots, N$ are nonlinearly mapped to the so-called kernel space $\mathcal{F}$ using a nonlinear function $\phi(\cdot)$, such that $\mathbf{x}_i \in \mathbb{R}^D \rightarrow \phi(\mathbf{x}_i) \in \mathcal{F}$. In $\mathcal{F}$, a linear projection of all the training data to $\mathbb{R}^d$ is given by:

$$\mathbf{y}_i = \mathbf{Q}\phi(\mathbf{x}_i), \quad , i = 1, \ldots, N, \quad (17)$$

where $\mathbf{Q} \in \mathbb{R}^{d \times |\mathcal{F}|}$ is a projection matrix of arbitrary dimensions [13]. In order to calculate the data representations $\mathbf{y}_i$, $i = 1, \ldots, N$, we employ the kernel trick stating that $\mathbf{Q}$ can be expressed as a linear combination of the training data representations in $\mathcal{F}$ leading to:

$$\mathbf{y}_i = \mathbf{W}\mathbf{\Phi}^T\phi(\mathbf{x}_i) = \mathbf{W}\mathbf{k}_i, \; i = 1, \ldots, N, \quad (18)$$

where $\mathbf{\Phi} \in \mathbb{R}^{|\mathcal{F}| \times N}$ is a matrix formed by the training data representations in $\mathcal{F}$, $\mathbf{W} \in \mathbb{R}^{d \times N}$ is a matrix containing the reconstruction weights of $\mathbf{W}$ with respect to $\mathbf{\Phi}$ and $\mathbf{k}_i$ is the i-th column of the so-called kernel matrix $\mathbf{K} \in \mathbb{R}^{N \times N}$ having elements equal to $\mathbf{K}_{ij} = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$. In our experiments we use the RBF kernel, given by:

$$\mathbf{K}_{ij} = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{\sigma^2}\right) \quad (19)$$

where $\sigma > 0$ is a hyper-parameter scaling the Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_j$.

In order to exploit the above-described nonlinear data mapping within the proposed S-SVDD method, we work as follows: for a given matrix $\mathbf{W}$, the training data $\mathbf{x}_i, i = 1, \ldots, N$ are mapped to $\mathbf{y}_i$, $i = 1, \ldots, N$ using (18) and α_i, $i = 1, \ldots, N$ are calculated by optimizing (12). Subsequently, $\mathbf{W}$ is updated using:

$$\Delta L = 2\sum_{i=1}^{N} \alpha_i \mathbf{W}\mathbf{k}_i\mathbf{k}_i^T - 2\sum_{i=1}^{N}\sum_{j=1}^{N} \mathbf{W}\mathbf{k}_i\mathbf{k}_j^T \alpha_i\alpha_j + \beta\Delta\Psi, \tag{20}$$

$$\Delta\Psi = 2\mathbf{W}\mathbf{K}\lambda\lambda^T\mathbf{K}^T. \tag{21}$$

E. Test phase

During testing, a sample $\mathbf{x}_* \in \mathbb{R}^D$ is mapped to its representation in the low-dimensional space $\mathbf{y}_* \in \mathbb{R}^d$ using (1) (or (18) for the non-linear case) and its distance from the hypersphere center is calculated:

$$\|\mathbf{y}_* - \mathbf{a}\|_2^2 = \mathbf{y}_*^T\mathbf{y}_* - 2\sum_{i=1}^{N} \alpha_i \mathbf{y}_*^T\mathbf{y}_i + \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_j \mathbf{y}_i^T\mathbf{y}_j. \tag{22}$$

$\mathbf{y}_*$ is classified as positive when $\|\mathbf{y}_* - \mathbf{a}\|_2^2 \leq R^2$ and as negative, otherwise.

Table I: List of datasets used

No.	Dataset name	*N*	*D*	Target class
1	Balance scale	625	4	Left
2	Iris	150	4	Iris-virginica
3	Lenses	24	4	No contact lenses
4	Seeds	210	7	Kama
5	Haberman's survival	306	3	Survived
6	Qualitative bankruptcy	250	7	Bankrupt
7	User knowledge modeling	403	5	Low
8	Pima Indians diabetes	768	8	No diabetes
9	Banknote authentication	1372	5	No
10	TA evaluation	151	5	High
11	PDelft pump	1500	64	Normal
12	Vehicle Opel	864	18	Opel
13	Sonar	208	60	Mines
14	Breast Wisconsin	699	9	Malignant

III. Experiments

A. Data-sets, evaluation criteria and experimental setup

We performed experiments on the datasets listed in Table I. Datasets 1-10 are downloaded from UCI website [21], while datasets 11-14 are downloaded from TU delft pattern recognition lab website [22]. The datasets with more than two classes were converted to a positive class and a negative class by considering the class with the majority of samples as the positive class and others as the negative class. Table I shows the target class of the each dataset in the last column.

In binary classification, a machine learning model can make two kinds of errors during testing. It can either wrongly predict a data sample from the positive class as negative or a negative data sample as positive. In one-class classification, the focus is on the target class and usually it is of greater interest to predict the positive class accurately. Recall, also called sensitivity, hit rate, or true positive rate, is the proportion of correctly classified positive samples during the test:

$$Recall = \frac{tp}{p}, \tag{23}$$

where tp is the total number of correctly classified positive samples and p is the total number of positive samples in the data. Recall is used to evaluate classification results in cases, where it is more important to predict the positive class accurately. Another metric used to evaluate machine learning algorithm is precision, which is the proportion of correctly classified samples among those classified into the positive class:

$$Precision = \frac{tp}{tp + fp}, \tag{24}$$

where fp is an acronym for false positives, i.e., the number of samples incorrectly predicted as positive during the test. A perfect precision score of 1.0 means that every sample classified as positive is from the positive class. In other words, a low precision score indicates a large number of false positives. F1 measure takes into account both precision and recall. It is defined as their harmonic mean as

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}. \tag{25}$$

We use (25) for evaluating and comparing performance of the proposed algorithm with competing methods.

To perform our experiments we divided our datasets into train and test sets. We performed our experiments on each dataset by selecting 70 percent of the data for training and the remaining 30 percent for testing. The 70-30 train and test sets were selected randomly 5 times to check the performance of each model robustly. Thus, in total we created 5 train-test (70-30%) partitions for each dataset. The proportion of each class in each set follows the original proportions. Also for the datasets having originally more than two classes, the positive and negative class labels were assigned after the subset for training and testing were created as described.

We selected the parameters for the proposed method by 5-fold cross-validation over each training set according to the best average F1 measure and then used them to train the final model using the whole training set. Whenever we trained a model, only positive samples were used. We selected the value of the parameter β as 10^l, where $l = -4, \ldots, 4$, σ is the scaled version of the mean distance between the training samples using a scaling factor 10^l, where $l = -3, \ldots, 3$, and C from $[0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6]$. The subspace dimension d for datasets having more than 10 dimensional feature space was restricted to a maximum of 10, i.e., $d = 1, \ldots, 10$. For datasets with $D \leq 10$, we set $d = 1, \ldots, D$.

We compared our results with the original SVDD (linear and kernel), OC-SVM (linear and kernel), GE-OC-SVM and GE-SVDD. The parameters were selected using a similar 5-fold cross-validation approach and the common parameters were selected from the ranges given above. Other parameters were selected as in the corresponding research papers.

Table II: F1 measures on 14 datasets

Dataset	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Linear														
SVDD	0.703	0.762	0.609	0.774	0.834	0.686	0.634	0.791	0.764	0.485	0.846	0.853	0.625	0.958
OC-SVM	0.688	0.612	0.394	0.619	0.644	0.562	0.532	0.529	0.657	**0.532**	0.632	0.590	0.535	0.660
S-SVDD Ψ_1	**0.907**	**0.899**	0.620	0.756	0.836	0.692	**0.960**	0.786	**0.908**	0.482	0.856	**0.855**	0.618	0.957
S-SVDD Ψ_2	0.898	0.897	**0.724**	**0.827**	0.839	0.720	0.957	**0.793**	0.889	0.502	0.857	**0.855**	0.599	**0.960**
S-SVDD Ψ_3	0.896	0.881	0.649	0.798	**0.841**	**0.722**	0.946	0.787	0.886	0.507	0.856	**0.855**	0.633	**0.960**
S-SVDD Ψ_4	0.896	0.868	0.694	0.778	0.821	0.715	0.954	0.784	0.852	0.458	**0.857**	0.854	**0.638**	0.953
Non-linear														
SVDD	0.734	0.827	0.413	**0.858**	0.835	0.605	0.651	0.785	0.804	0.396	0.836	0.852	0.609	0.962
OC-SVM	0.544	0.673	0.523	0.444	0.743	0.550	0.409	0.786	0.700	0.274	0.661	0.679	0.530	0.630
GE-SVDD	0.757	0.857	0.314	0.799	0.811	0.554	0.654	**0.790**	0.797	0.484	0.830	0.847	0.550	**0.966**
GE-OC-SVM	**0.815**	**0.869**	0.398	0.800	0.816	0.594	**0.658**	0.667	**0.930**	**0.498**	0.613	0.788	0.593	0.962
S-SVDD Ψ_1	0.635	0.725	**0.736**	0.727	0.842	0.700	0.518	0.786	0.728	0.472	0.836	**0.858**	0.504	0.961
S-SVDD Ψ_2	0.662	0.573	0.603	0.540	**0.845**	**0.762**	0.523	**0.790**	0.717	0.473	**0.856**	**0.858**	**0.637**	0.783
S-SVDD Ψ_3	0.734	0.694	0.624	0.719	0.838	0.620	0.578	0.785	0.720	0.417	**0.856**	**0.858**	**0.637**	0.902
S-SVDD Ψ_4	0.495	0.700	**0.736**	0.774	0.841	0.632	0.562	0.572	0.703	0.474	0.832	**0.858**	**0.637**	0.951

Table III: Standard deviation of the F1 scores

Dataset	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Linear														
SVDD	0.014	0.041	0.152	0.041	0.009	0.072	0.032	0.009	0.010	0.025	0.005	0.003	0.033	0.002
OC-SVM	0.074	0.143	0.257	0.171	0.055	0.082	0.071	0.093	0.015	0.091	0.027	0.024	0.083	0.040
S-SVDD Ψ_1	0.022	0.034	0.154	0.041	0.007	0.046	0.017	0.009	0.026	0.088	0.002	0.004	0.022	0.004
S-SVDD Ψ_2	0.026	0.032	0.136	0.052	0.017	0.012	0.019	0.006	0.031	0.049	0.001	0.003	0.058	0.012
S-SVDD Ψ_3	0.029	0.061	0.157	0.057	0.009	0.016	0.016	0.016	0.039	0.052	0.001	0.003	0.048	0.004
S-SVDD Ψ_4	0.024	0.063	0.118	0.030	0.038	0.016	0.031	0.013	0.110	0.078	0.001	0.003	0.022	0.016
Non-linear														
SVDD	0.020	0.020	0.276	0.066	0.011	0.046	0.027	0.010	0.011	0.224	0.008	0.005	0.042	0.008
OC-SVM	0.164	0.158	0.331	0.275	0.139	0.107	0.233	0.014	0.073	0.173	0.114	0.146	0.075	0.354
GE-SVDD	0.029	0.022	0.312	0.064	0.045	0.045	0.052	0.021	0.023	0.101	0.007	0.006	0.042	0.009
GE-OC-SVM	0.039	0.056	0.368	0.071	0.026	0.131	0.058	0.261	0.019	0.063	0.188	0.121	0.090	0.009
S-SVDD Ψ_1	0.006	0.058	0.060	0.178	0.010	0.029	0.036	0.004	0.039	0.029	0.047	0.000	0.282	0.018
S-SVDD Ψ_2	0.053	0.124	0.340	0.089	0.004	0.048	0.051	0.002	0.012	0.050	0.002	0.000	0.000	0.100
S-SVDD Ψ_3	0.013	0.027	0.353	0.059	0.008	0.135	0.074	0.029	0.017	0.133	0.002	0.000	0.000	0.079
S-SVDD Ψ_1	0.369	0.035	0.060	0.049	0.007	0.089	0.058	0.330	0.021	0.037	0.057	0.000	0.000	0.021

B. Experimental results

Fig. 1, illustrates an example transformation of all the data samples of dataset 5 (Haberman's survival) from the original 3-dimensional feature space to a lower 2-dimensional feature space using the non-linear version of the proposed S-SVDD method with the constraint Ψ_1 (see subsection II-B). The figure shows the capability of the proposed method to transform the data to a compact form which is more suitable to be enclosed by a hypersphere.

In Tables II and III, we report the average F1 measure and the standard deviation of F1 measure for the evaluated linear and non-linear methods. The linear version of the proposed S-SVDD clearly outperforms all other linear methods. Only for dataset 10, OC-SVM achieves a higher performance. The non-linear version of S-SVDD outperformed other non-linear methods on datasets 3, 5, 6, 11, 12 and 13. For dataset 8, GE-SVDD and S-SVDD Ψ_2 achieved the same results. GE-OC-SVM obtained the best results on datasets 1, 2, 7, 9 and 10. Compared to the baseline methods (SVDD and OC-SVM), S-SVDD shows a clear improvement.

For datasets 12 and 13, the non-linear versions of S-SVDD (except for Ψ_1 for dataset 13) have zero standard deviation. A closer inspection of the results shows that, in these cases, the obtained mapping and data description classify all the test samples as positive, due to the selection of small values for the hyper-parameter C [23]. A tighter fitted hypersphere on the training data may possibly lead to more meaningful results, which could be achieved by restricting the range of the C values used during the cross-validation process applied on the training data for hyper-parameter selection of the proposed method.

When comparing the different regularization terms Ψ used with the proposed method, Ψ_2 achieves the best performance most often with both linear and non-linear versions. In Ψ_2, all training samples contribute to the regularization term equally.

IV. CONCLUSION

In this paper, we proposed a new method for one-class classification. The proposed S-SVDD method maps the original data to a lower dimensional feature space, which is more suitable for one-class classification. The method iteratively optimizes the mapping to the new subspace and the data description in that feature space. Both linear and non-linear versions were defined along with four different regularization terms.

We performed experiments on 14 different publicly available datasets. Our experiments showed that the proposed

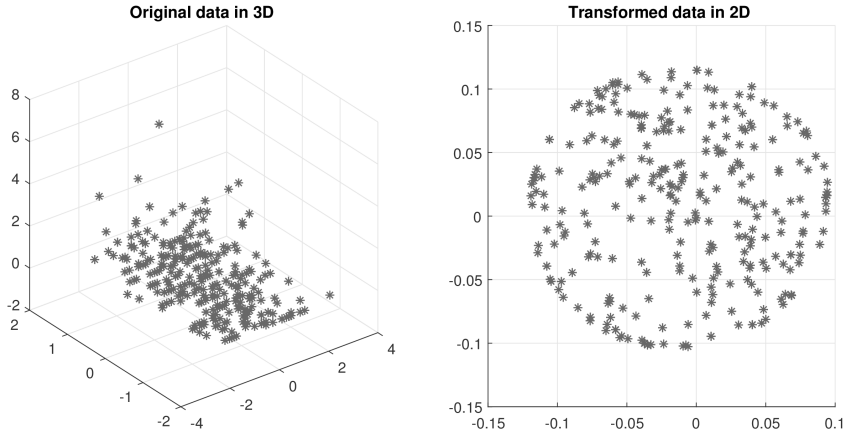

Figure 1: Transforming dataset 5 (Haberman's survival) from 3-dimensional feature space to 2-dimensional feature space using the proposed method (kernel S-SVDD Ψ_1)

method yields better results than the baselines and competing one-class classification methods in majority of the cases. A constraint that uses all samples for describing the data variance leads to the best results for S-SVDD.

In the future, we intend to try the proposed S-SVDD method with different kernels and design new regularization terms. We will also evaluate a similar mapping approach in combination with other already established one-class classification methods.

ACKNOWLEDGEMENT

This work was supported by the NSF-TEKES Center for Visual and Decision Informatics project Co-Botics, jointly sponsored by Tieto Oy Finland and CA Technologies.

REFERENCES

[1] C.-L. Liu and H. Sako, "Class-specific feature polynomial classifier for pattern classification and its application to handwritten numeral recognition," *Pattern recognition*, vol. 39, no. 4, pp. 669–681, 2006.

[2] A. Iosifidis and M. Gabbouj, "Class-specific kernel discriminant analysis revisited: Further analysis and extensions," *IEEE Transactions on Cybernetics*, vol. 47, no. 12, pp. 4485–4496, 2017.

[3] A. Iosifidis, A. Tefas, and I. Pitas, "Class-specific reference discriminant analysis with application in human behavior analysis," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 3, pp. 315–326, 2015.

[4] A. Iosifidis and M. Gabbouj, "Scaling up class-specific kernel discriminant analysis for large-scale face verificationn," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 11, pp. 2453–2465, 2016.

[5] L. C. M.A. Pimentel, D.A. Clifton and L. Tarassenko, "A review of novelty detection," *Signal Processing*, vol. 99, pp. 215–249, 2014.

[6] D. Martinus and J. Tax, "One-class classification: Concept-learning in the absence of counterexamples," Ph.D. dissertation, PhD thesis, Delft University of Technology, 2001.

[7] C. M. Bishop, *Neural networks for pattern recognition*. Oxford university press, 1995.

[8] E. Parzen, "On estimation of a probability density function and mode," *The annals of mathematical statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[9] M. GhasemiGol, M. Sabzekar, R. Monsefi, M. Naghibzadeh, and H. S. Yazdi, "A new support vector data description with fuzzy constraints," in *Intelligent Systems, Modelling and Simulation (ISMS), 2010 International Conference on*. IEEE, 2010, pp. 10–14.

[10] T. Kohonen, "Learning vector quantization," in *Self-Organizing Maps*. Springer, 1995, pp. 175–189.

[11] D. M. Tax and R. P. Duin, "Support vector data description," *Machine learning*, vol. 54, no. 1, pp. 45–66, 2004.

[12] B. Schölkopf, R. Williamson, A. Smola, and J. Shawe-Taylor, "Sv estimation of a distribution's support," *Advances in neural information processing systems*, vol. 12, 1999.

[13] B. Scholkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2001.

[14] T. Le, D. Tran, W. Ma, and D. Sharma, "A unified model for support vector machine and support vector data description," in *Neural Networks (IJCNN), The 2012 International Joint Conference on*. IEEE, 2012, pp. 1–8.

[15] V. Mygdalis, A. Iosifidis, A. Tefas, and I. Pitas, "Graph embedded one-class classifiers for media data classification," *Pattern Recognition*, vol. 60, pp. 585–595, 2016.

[16] R. Razavi-Far, M. Farajzadeh-Zanjani, S. Zare, M. Saif, and J. Zarei, "One-class classifiers for detecting faults in induction motors," in *Electrical and Computer Engineering (CCECE), 2017 IEEE 30th Canadian Conference on*. IEEE, 2017, pp. 1–5.

[17] C. Sanchez-Hernandez, D. S. Boyd, and G. M. Foody, "One-class classification for mapping a specific land-cover class: Svdd classification of fenland," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 4, pp. 1061–1073, 2007.

[18] W. Sakla, A. Chan, J. Ji, and A. Sakla, "An svdd-based algorithm for target detection in hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 2, pp. 384–388, 2011.

[19] M. Popescu and A. Mahnot, "Acoustic fall detection using one-class classifiers," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 3505–3508.

[20] A. Iosifidis, V. Mygdalis, A. Tefas, and I. Pitas, "One-class classification based on extreme learning and geometric class information," *Neural Processing Letters*, vol. 45, no. 2, pp. 577–592, 2017.

[21] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[22] "Technical University Delft pattern recognition lab, one-class classifier." [Online]. Available: http://homepage.tudelft.nl/n9d04/occ/index.html

[23] C. L. W.C. Chang and C. Lin, "A revisit to support vector data description (svdd)," *Technical report*, 2013.

Publication II

Ellipsoidal Subspace Support Vector Data Description

F. Sohrab, J. Raitoharju, A. Iosifidis and M. Gabbouj

IEEE Access 8.(2020), 122013–122025
DOI: 10.1109/ACCESS.2020.3007123

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 17, 2020, accepted June 30, 2020, date of publication July 6, 2020, date of current version July 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3007123

Ellipsoidal Subspace Support Vector Data Description

FAHAD SOHRAB[1], (Graduate Student Member, IEEE),
JENNI RAITOHARJU[2], (Member, IEEE),
ALEXANDROS IOSIFIDIS[3], (Senior Member, IEEE),
AND MONCEF GABBOUJ[1], (Fellow, IEEE)

[1]Faculty of Information Technology and Communication Sciences, Tampere University, 33720 Tampere, Finland
[2]Finnish Environment Institute, 40500 Jyväskylä, Finland
[3]Department of Engineering, Aarhus University, 8200 Aarhus, Denmark

Corresponding author: Fahad Sohrab (fahad.sohrab@tuni.fi)

This work was supported in part by the National Science Foundation (NSF)-Business Finland Center for Visual and Decision Informatics (CVDI) Project Amalia, and in part by the Business Finland projects VIRPA D and Industrial Data Excellence (INDEX) (Digital, Internet, Materials and Engineering Co-Creation (DIMECC) Industrial Data program).

ABSTRACT In this paper, we propose a novel method for transforming data into a low-dimensional space optimized for one-class classification. The proposed method iteratively transforms data into a new subspace optimized for ellipsoidal encapsulation of target class data. We provide both linear and non-linear formulations for the proposed method. The method takes into account the covariance of the data in the subspace; hence, it yields a more generalized solution as compared to the data description in the subspace by hyperspherical encapsulation of target class data. We propose different regularization terms expressing the class variance in the projected space. We compare the results with classic and recently proposed one-class classification methods and achieve competing results and show clear improvement compared to the other support vector based methods. The proposed method is also noticed to converge much faster than recently proposed Subspace Support Vector Data Description.

INDEX TERMS Anomaly detection, ellipsoidal data description, machine learning, one-class classification, subspace learning.

I. INTRODUCTION

The ability of machines to make a concise description of information requires learning from previous experience. Researchers have been trying to develop techniques for accurately modeling data using supervised and unsupervised learning techniques for many decades. In unsupervised learning techniques, patterns are found without any knowledge of class labels [1]. In supervised learning, labeled training data are used to train models for classifying future instances into different categories [2]. A typical multi-class classification task can be decomposed into several binary classification tasks, where the aim is to decide to which of the two considered classes samples belong to [3]. In binary classification, the data from both classes are used to train a model. One-class classification is conceptually close to binary classification, but the models for classifying future instances are trained using data only from one particular target class [4], [5]. In practice, one-class classification is used when data from one of the classes is scarce.

In one-class classification, the class of interest to be modeled is called target or positive class, while samples from the other unknown class(es) are referred to as outliers or negative samples. Numerous attempts have been made to solve one-class classification tasks [6]. The three main approaches for solving one-class classification tasks are density based, reconstruction based, and border based methods [7]. In the density based approach, the description of the target class is based on its density [8], which is usually estimated by using popular density estimation methods such as Parzen density, Gaussian model, or mixture of Gaussians [9]. In reconstruction based approach, some assumptions about the data generating process are made. The underlying function which represents the target class is obtained by fitting a curve over the data by using prior information, such as data clustering characteristics. Self-organizing maps (SOM) [10] and least-squares quantization [11] are classic examples of reconstruction methods. In border based approaches, a model is

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

created by defining a closed boundary around the target class without estimating its density. One-class Support Vector Machine (OC-SVM) [12] and Support Vector Data Description (SVDD) [13] are among the popular boundary techniques for one-class classification. In OC-SVM, a hyperplane separating the target class is constructed so that the distance of the hyperplane from the origin is maximized. In SVDD, a hypersphere is formed around the target class data by minimizing the volume of hypersphere in a given feature space. Recently, there has been a rising trend to propose approaches based on regression and neural-networks as well [5], [14].

SVDD has been justified over time as a powerful data description method and it has been used in many different application domains for solving one-class classification problems. For example, in [15], SVDD is found to be an excellent choice for solving the problem of identification of freshness of eggs using near infrared spectroscopy (NIR) with an imbalanced number of training samples. In [16], a terrain classification method for ensuring navigation safety of mobile service robots based on SVDD is proposed. To enhance the performance of SVDD, numerous extensions and hybridization techniques have been proposed [8], [17]–[21]. The main extensions of SVDD can be categorized into four main categories. In the first category of extensions, the techniques are focused on manipulating the structure of data, such as associating a confidence coefficient with all training instances which deals with the uncertainty of data [22]. In the second category, the performance is enhanced by proposing new non-linear methods and reducing the complexity of algorithms [23], [24]. Techniques for handling non-stationary data in the context of one-class classification falls in the third category of extensions [25]. In the fourth category, different changes are proposed in the shape of the boundary encapsulating the target data [26].

A popular alternative to the spherical SVDD is Ellipsoidal SVDD (E-SVDD) [26], [27]. E-SVDD forms a unique hyperellipsoid with a minimum volume covering most of the target data. An ellipsoid, unlike a hypersphere, takes into account the difference in variance for each dimension as well as covariance between them. A hypersphere, characterized only by a radius and a center will result in superfluous regions which do not contain any target objects in the input space [28]. Ellipsoids with a minimum volume containing the target data have applications spanning over many different fields. For example, in [29], it is used to detect intrusion in computer networks and, in [30], it is used to estimate the distance between a robot and its surrounding environment for obstacle collision avoidance. An ellipsoid is preferred for heterogeneous data in the input space because its shape is less conservative than a sphere. However, there are some difficulties in kernelizing the algorithms. The kernel trick cannot be applied directly to E-SVDD because its formulation includes outer products rather than inner products [31].

In this paper, we propose a novel subspace learning algorithm for ellipsoidal one-class classification. The proposed method takes into account the covariance of data in the subspace so that the boundary created around the target class is a better fit. The proposed method finds a projection along with a data description iteratively by minimizing the volume of the hyperellipsoid. We propose different variants of the proposed method by proposing different settings of the regularization term, which takes into account the concentration matrix. We also annexed the regularization term with different settings without taking into account the concentration matrix and report the results. The proposed method is called Ellipsoidal Subspace Support Vector Data Description (ES-SVDD), since it is analogous to Subspace Support Vector Data Description (S-SVDD) [32] but offers more flexibility by using hyperellipsoid instead of hypersphere. Our results show that using hyperellipsoid for data description in the subspace converges faster and produces better results than the data description in a subspace using hypersphere. Further, we see that hyperellipsoid in the subspace optimised for one-class classification provides provides a better data description as compared to the hyperellipsoid in the original feature space. We also propose a non-linear version of the algorithm by exploiting the non-linear projection trick (NPT) [33].

The rest of the paper is organized as follows. In Section II, we present an overview of related works. In Section III, a detailed derivation of the newly proposed method is presented. In Section IV, we provide and discuss the experimental protocol along with the obtained results and, finally, conclusions are drawn in Section V.

II. BACKGROUND AND RELATED WORK

One-class classification has been studied extensively in recent years and the approaches predominantly focus on data description in a given feature space [7], [13], [22]. On the other hand, feature selection and subspace learning have been an active research area in machine learning, primarily for challenges with data available for all categories [34], [35]. The aim is to avoid the curse of dimensionality in the original feature space by modeling the given data in a lower dimensional space.

In feature selection methods, a subset of representative features is selected by following some criterion [36]–[38]. The two main approaches for feature selection are the *filter* approach and the *wrappers* approach. In the filter approaches, the main focus is on the intrinsic characteristics of the data and they do not take into account any classification algorithm. On the other hand, the wrappers approaches are dependent only on a specific classification algorithm [39].

In subspace learning, the features are transformed from original feature space to a lower-dimensional subspace [40]. Most of the existing subspace learning methods, particularly for anomaly detection, follow three general steps [41], [42]: First, the features are selected randomly by applying random projections to the attributes. Second, classical algorithms are applied locally in each subspace and scores (e.g., voting) are computed. Finally, all the scores are aggregated to compute a global score for classification.

The focus of our paper is to find an optimized subspace for one-class classification. We review the classical one-class classification method, SVDD, in Section II-A and also provide an overview of S-SVDD and graph embedded one-class classifiers in Sections II-B and II-C, respectively.

A. SUPPORT VECTOR DATA DESCRIPTION

Let us denote the data points to be enclosed inside a closed boundary by a matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_N], \mathbf{x}_i \in \mathbb{R}^D$, where N is total number of instances and D is dimensionality of data in the original feature space. All the data samples represented by $\mathbf{X}$ belong to the same class.

SVDD finds a spherical boundary around the data by minimizing the volume of a hypersphere enclosing all the target class data:

$$\begin{aligned} &\min\ F(R, \mathbf{a}) = R^2 \\ &\text{s.t. } \|\mathbf{x}_i - \mathbf{a}\|_2^2 \leq R^2,\ \ \forall i \in \{1, \dots, N\}, \end{aligned} \tag{1}$$

where R is the radius of hypersphere and $\mathbf{a} \in \mathbb{R}^D$ is the center of the hypersphere in the given feature space. Slack variables $\xi_i,\ i = 1, \dots, N$ are introduced for allowing the possibility of data points being outliers, hence the optimization problem changes to

$$\begin{aligned} &\min\ F(R, \mathbf{a}) = R^2 + C \sum_{i=1}^{N} \xi_i \\ &\text{s.t. } \|\mathbf{x}_i - \mathbf{a}\|_2^2 \leq R^2 + \xi_i, \\ &\quad \xi_i \geq 0,\ \ \forall i \in \{1, \dots, N\}, \end{aligned} \tag{2}$$

where $C > 0$ is a hyperparameter which controls the trade-off between the volume of the sphere and the amount of data outside the sphere. The Lagrangian dual of (2) reduces to

$$L = \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^\intercal \mathbf{x}_i - \sum_{i}^{N} \sum_{j}^{N} \alpha_i \alpha_j \mathbf{x}_i^\intercal \mathbf{x}_j, \tag{3}$$

subject to $0 \leq \alpha_i \leq C$. Maximizing (3) gives a set of α_i for corresponding data points. The samples with $\alpha_i > 0$ are the support vectors defining the data description [13]. The samples corresponding to $0 < \alpha_i < C$ lie on the boundary of the hypersphere and those with $\alpha_i = C$ are outliers.

B. SUBSPACE SUPPORT VECTOR DATA DESCRIPTION

In S-SVDD [32], a projection matrix $\mathbf{Q}$ is determined to map data from the original space $\mathbb{R}^D$ to a new optimized lower dimensional space $\mathbb{R}^d,\ d < D$, so that the data are more suitable for one-class classification:

$$\begin{aligned} &\min\ F(R, \mathbf{a}) = R^2 + C \sum_{i=1}^{N} \xi_i \\ &\text{s.t. } \|\mathbf{Q}\mathbf{x}_i - \mathbf{a}\|_2^2 \leq R^2 + \xi_i, \\ &\quad \xi_i \geq 0,\ \ \forall i \in \{1, \dots, N\}, \end{aligned} \tag{4}$$

where $\mathbf{a} \in \mathbb{R}^d$ is the center of the hypersphere in lower d-dimensional space. The method iteratively solves the SVDD in the current subspace to obtain the data description parameters $\alpha_i,\ i = 1, \dots, N$, and then updates the subspace projection by optimizing an augmented version of the Lagrangian:

$$L = \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^\intercal \mathbf{Q}^\intercal \mathbf{Q} \mathbf{x}_i - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \mathbf{x}_i^\intercal \mathbf{Q}^\intercal \mathbf{Q} \mathbf{x}_j \alpha_j + \beta \psi, \tag{5}$$

where ψ is a regularization term expressing the class variance in the low dimensional space and β is a regularization parameter controlling the importance of the ψ, where

$$\psi = \mathrm{Tr}(\mathbf{Q}\mathbf{X}\boldsymbol{\lambda}\boldsymbol{\lambda}^\intercal \mathbf{X}^\intercal \mathbf{Q}^\intercal), \tag{6}$$

where Tr(.) is the trace operator and $\boldsymbol{\lambda} \in \mathbb{R}^N$ is a vector controlling the contribution of each training sample. $\mathbf{Q}$ is updated by using the gradient of (5), i.e.,

$$\mathbf{Q} \leftarrow \mathbf{Q} - \eta \Delta L, \tag{7}$$

where η is the learning rate. A non-linear version of S-SVDD employing the kernel trick is also proposed in [32].

C. GRAPH EMBEDDED ONE-CLASS CLASSIFIERS

Graph embedded one-class classifiers constitute extensions of the OC-SVM and SVDD by incorporating generic graph structures in their optimization process. The generic graph structures express geometric data relationships of the target class in the data. For example, Graph Embedded SVDD (GE-SVDD) [17] optimization problem is formulated as

$$\begin{aligned} &\min\ F(R, \mathbf{a}) = R^2 + C \sum_{i=1}^{N} \xi_i \\ &\text{s.t. } \big(\phi(\mathbf{x}_i) - \mathbf{a}\big)^\intercal \mathbf{S}^{-1} \big(\phi(\mathbf{x}_i) - \mathbf{a}\big) \leq R^2 + \xi_i, \\ &\quad \xi_i \geq 0, \quad \forall i \in \{1, \dots, N\}, \end{aligned} \tag{8}$$

where $\phi(.)$ is any non-linear function used for mapping the training samples from the input feature space to the kernel space. The matrix $\mathbf{S}$ contains the geometric data relationships. For example, in PCA, the scatter of training data can be expressed as

$$\mathbf{S} = \frac{1}{N} \Phi \Big(\mathbf{I} - \frac{1}{N} \mathbf{1}\mathbf{1}^\intercal \Big) \Phi^\intercal = \Phi \mathbf{L} \Phi^\intercal, \tag{9}$$

where $\mathbf{1} \in \mathbb{R}^N$ is a vector containing all values as ones, $\mathbf{I} \in \mathbb{R}^{N \times N}$ is an identity matrix, and Φ is a matrix that contains the training data representations in kernel space.

The Lagrangian of GE-SVDD is

$$L = \sum_{i=1}^{N} \alpha_i \phi(\mathbf{x}_i)^\intercal \mathbf{S}^{-1} \phi(\mathbf{x}_i) - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \phi(\mathbf{x}_i)^\intercal \mathbf{S}^{-1} \phi(\mathbf{x}_j) \alpha_j. \tag{10}$$

It has been shown in [17] that the optimization problem in (10) is equivalent to the problem of SVDD in a transformed feature space.

III. ELLIPSOID SUBSPACE SUPPORT VECTOR DATA DESCRIPTION

Our aim is to find a projection matrix $\mathbf{Q} \in \mathbb{R}^{d\times D}$ to be used for transforming the data to an optimized subspace suitable for one-class classification. In the following analysis, we assume that the data has been centered by setting $\mathbf{X} \leftarrow \mathbf{X} - \boldsymbol{\mu}$, where $\boldsymbol{\mu}$ is the mean of the given training data. The mapping from the original feature space with dimensionality D to a subspace with dimensionality $d \leq D$ is carried out. The mapping is done to transform the data so that it is more suitable to be encapsulated inside an ellipsoid with a minimum volume.

The optimization problem is formulated as

$$\begin{aligned} &\min\ F(R,\mathbf{a}) = R^2 + C\sum_{i=1}^{N}\xi_i \\ &\text{s.t. } (\mathbf{Q}\mathbf{x}_i - \mathbf{a})^\intercal \mathbf{E}^{-1}(\mathbf{Q}\mathbf{x}_i - \mathbf{a}) \le R^2 + \xi_i, \\ &\quad \xi_i \ge 0, \quad \forall i \in \{1,\ldots,N\}, \end{aligned} \tag{11}$$

where $\mathbf{a}$ is the center of the hyperellipsoid and $\mathbf{E} = \mathbf{Q}\mathbf{X}\mathbf{X}^\intercal\mathbf{Q}^\intercal$ is the covariance matrix of the data in d-dimensional space. The inverse of covariance matrix $\mathbf{E}$, also known as the concentration or precision matrix is symmetric and positive definite $\mathbf{E}^{-1} \in \mathbb{R}^{d\times d}$. By defining a new vector $\mathbf{u} = \mathbf{E}^{-\frac{1}{2}}\mathbf{a}$, (11) can be written as

$$\begin{aligned} &\min\ F(R,\mathbf{u}) = R^2 + C\sum_{i=1}^{N}\xi_i \\ &\text{s.t. } \|\mathbf{E}^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i - \mathbf{u}\|_2^2 \le R^2 + \xi_i, \\ &\quad \xi_i \ge 0, \quad \forall i \in \{1,\ldots,N\}. \end{aligned} \tag{12}$$

The data in the subspace is represented by

$$\mathbf{y}_i = \mathbf{Q}\mathbf{x}_i, \ \ i = 1,\ldots,N. \tag{13}$$

The constraints in (12) can be incorporated into its corresponding objective function by using Lagrange multipliers:

$$\begin{aligned} L = R^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\big(R^2 + \xi_i \\ -(\mathbf{E}^{-\frac{1}{2}}\mathbf{y}_i)^\intercal\mathbf{E}^{-\frac{1}{2}}\mathbf{y}_i + 2\mathbf{u}^\intercal\mathbf{E}^{-\frac{1}{2}}\mathbf{y}_i - \mathbf{u}^\intercal\mathbf{u}\big) - \sum_{i=1}^{N}\gamma_i\xi_i \end{aligned} \tag{14}$$

with Lagrange multipliers $\alpha_i \ge 0$ and $\gamma_i \ge 0$.

By setting partial derivatives with respect to R, $\mathbf{u}$ and ξ_i to zero, we get

$$\frac{\partial L}{\partial R} = 0 \Rightarrow \sum_{i=1}^{N}\alpha_i = 1 \tag{15}$$

$$\frac{\partial L}{\partial \mathbf{u}} = 0 \Rightarrow \mathbf{u} = \sum_{i=1}^{N}\alpha_i\mathbf{E}^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i \tag{16}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \Rightarrow C - \alpha_i - \xi_i = 0. \tag{17}$$

By substituting (15)-(17) into (14) we get

$$L = \sum_{i=1}^{N}\alpha_i\mathbf{x}_i^\intercal\mathbf{Q}^\intercal\mathbf{E}^{-1}\mathbf{Q}\mathbf{x}_i - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\mathbf{x}_i^\intercal\mathbf{Q}^\intercal\mathbf{E}^{-1}\mathbf{Q}\mathbf{x}_j\alpha_j. \tag{18}$$

We can use SVDD to solve (18) for getting α_i values. The concentration matrix $\mathbf{E}^{-1}$ is equivalent to

$$\mathbf{E}^{-1} = (\mathbf{Q}\mathbf{X}\mathbf{X}^\intercal\mathbf{Q}^\intercal)^{-1}. \tag{19}$$

By putting (19) in (18) we get

$$\begin{aligned} L = \sum_{i=1}^{N}\alpha_i\mathbf{x}_i^\intercal\mathbf{Q}^\intercal(\mathbf{Q}\mathbf{X}\mathbf{X}^\intercal\mathbf{Q}^\intercal)^{-1}\mathbf{Q}\mathbf{x}_i \\ - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\mathbf{x}_i^\intercal\mathbf{Q}^\intercal(\mathbf{Q}\mathbf{X}\mathbf{X}^\intercal\mathbf{Q}^\intercal)^{-1}\mathbf{Q}\mathbf{x}_j\alpha_j. \end{aligned} \tag{20}$$

We add an extra term Υ to (20) as a regularization term expressing the class variance in the projected space, also taking into account the concentration matrix. Hence, (20) now becomes

$$\begin{aligned} L = \sum_{i=1}^{N}\alpha_i\mathbf{x}_i^\intercal\mathbf{Q}^\intercal(\mathbf{Q}\mathbf{X}\mathbf{X}^\intercal\mathbf{Q}^\intercal)^{-1}\mathbf{Q}\mathbf{x}_i \\ - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\mathbf{x}_i^\intercal\mathbf{Q}^\intercal(\mathbf{Q}\mathbf{X}\mathbf{X}^\intercal\mathbf{Q}^\intercal)^{-1}\mathbf{Q}\mathbf{x}_j\alpha_j + \beta\Upsilon, \end{aligned} \tag{21}$$

where β controls the importance of regularization term and is used as a hyperparameter. Υ is defined as follows:

$$\Upsilon = \mathrm{Tr}(\mathbf{E}^{-\frac{1}{2}}\mathbf{Q}\mathbf{X}\boldsymbol{\lambda}\boldsymbol{\lambda}^\intercal\mathbf{X}^\intercal\mathbf{Q}^\intercal\mathbf{E}^{-\frac{\intercal}{2}}), \tag{22}$$

where $\boldsymbol{\lambda}$ can take three different forms. In the first form, all elements in $\boldsymbol{\lambda}$ take the value of 1 and, hence, all the samples are used to describe the covariance of the class. In the second form, $\boldsymbol{\lambda}$ is replaced by $\boldsymbol{\alpha}$, which means that the samples belonging to the boundary and outside the boundary are used to describe the covariance of the class. In the third form, the λ_i values are replaced by α_i values of the samples belonging to the boundary and zero for other instances. The first, second and third forms of the regularization terms are expressed as Υ_1, Υ_2, and Υ_3 hereinafter.

In our experiments, we also consider the regularization term expressing the class variance in the projected space without taking into account the concentration matrix. This is achieved by replacing the covariance matrix $\mathbf{E}$ with the identity matrix $\mathbf{I}$ in (22). By doing so, the regularization term Υ becomes equivalent to ψ as described in (6). Analogous to regularization term Υ, ψ can also take different forms by changing $\boldsymbol{\lambda}$ and similarly hereinafter we refer to all those cases by ψ_1, ψ_2 and ψ_3. The methods used with ψ and Υ are denoted by ES-SVDDψ_m and ES-SVDDΥ_m ($m = 1, 2, 3$), respectively. We refer to the case, where no regularization term is used in ES-SVDD, as ES-SVDD$\psi_0\Upsilon_0$.

Equation (21) can be further simplified and written as

$$L = \mathrm{Tr}((\mathbf{Q}\mathbf{X}\mathbf{X}^\intercal\mathbf{Q}^\intercal)^{-1}\mathbf{Q}\mathbf{X}(\mathbb{A} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal)\mathbf{X}^\intercal\mathbf{Q}^\intercal) + \beta\Upsilon, \tag{23}$$

where $\mathbb{A}$ is a diagonal matrix having α_i values in its diagonal and $\boldsymbol{\alpha}$ is a vector of α_i's. We use gradient of (23) to update the projection matrix. The gradient can be solved using identity 126 in [43]:

$$\Delta L = 2\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}(\mathbb{A} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal)\mathbf{X}^\intercal - 2\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}(\mathbb{A} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal)\mathbf{X}^\intercal\mathbf{Q}^\intercal\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}\mathbf{X}^\intercal + \beta\Delta\Upsilon, \quad (24)$$

where

$$\Delta\Upsilon = 2\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}\boldsymbol{\lambda}\boldsymbol{\lambda}^\intercal\mathbf{X}^\intercal - 2\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}\boldsymbol{\lambda}\boldsymbol{\lambda}^\intercal\mathbf{X}^\intercal\mathbf{Q}^\intercal\mathbf{E}^{-1}\mathbf{Q}\mathbf{X}\mathbf{X}^\intercal. \quad (25)$$

When ψ is used as a regularization term, we use $\Delta\psi$ instead of $\Delta\Upsilon$ in (24):

$$\Delta\psi = 2\mathbf{Q}\mathbf{X}\boldsymbol{\lambda}\boldsymbol{\lambda}^\intercal\mathbf{X}^\intercal. \quad (26)$$

We obtain an optimised data projection matrix along with optimised data description in a two-step iterative process. In the first step, the α_i values are computed by maximizing (18). In the second step, $\mathbf{Q}$ is updated through the gradient descent after computing the gradient by using (23). In order to obtain an orthogonal projection, we impose the orthogonality constraint $\mathbf{Q}\mathbf{Q}^\intercal = \mathbf{I}$. We orthogonalize and normalize $\mathbf{Q}$ during the two-step iterative process. Algorithm 1 presents the whole algorithm.

Algorithm 1 Linear ES-SVDD Optimization

Input : $\mathbf{X}, \beta, \eta, d, C, k_{max}$
Output: $\mathbf{Q}, R, \boldsymbol{\alpha}$
Random initialization of $\mathbf{Q}$;
Initialize $k = 1$;
while $k < k_{max}$ **do**
  Compute concentration matrix $\mathbf{E}^{-1}$ using (19) ;
  Solve $\alpha_i,\ i = 1, \ldots, N$ with SVDD using (18);
  Calculate ΔL using (24);
  Update $\mathbf{Q} \leftarrow \mathbf{Q} - \eta\Delta L$;
  Orthogonalize $\mathbf{Q}$ using QR decomposition;
  Row normalize $\mathbf{Q}$ using l_2 norm;
  $k \leftarrow k + 1$
end
// Data description in the optimized subspace
Compute concentration matrix $\mathbf{E}^{-1}$ using (19)
Calculate $\alpha_i,\ i = 1, \ldots, N$ with SVDD using (18) ;

A. NON-LINEAR ELLIPSOIDAL SUBSPACE SUPPORT VECTOR DATA DESCRIPTION

The non-linear ellipsoidal subspace SVDD is not trivial, because the kernel trick cannot be applied directly due to the outer products involved in its derivation. To avoid this problem, we follow the NPT based solution described below [33]. We first compute a noncentered kernel matrix $\mathbf{K} = \Phi^\intercal\Phi$ using the radial basis function kernel as

$$\mathbf{K}_{ij} = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2\sigma^2}\right), \quad (27)$$

where σ is a hyperparameter scaling the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$. The kernel matrix is centered as

$$\hat{\mathbf{K}} = (\mathbf{I} - \mathbf{J})\mathbf{K}(\mathbf{I} - \mathbf{J}), \quad (28)$$

where $\mathbf{J} \in \mathbb{R}^{N\times N}$ is a matrix defined as

$$\mathbf{J} = \frac{1}{N}\mathbf{1}\mathbf{1}^\intercal. \quad (29)$$

The centered kernel matrix is decomposed by using eigendecomposition:

$$\hat{\mathbf{K}} = \mathbf{U}\mathbf{A}\mathbf{U}^\intercal, \quad (30)$$

where $\mathbf{A}$ contains the non-negative eigenvalues of the centered kernel matrix in its diagonal and the columns of $\mathbf{U}$ contain the corresponding eigenvectors. Finally, the data in the reduced dimensional kernel space is obtained as

$$\Phi = (\mathbf{A}^{\frac{1}{2}})^+\mathbf{U}^+\hat{\mathbf{K}}, \quad (31)$$

where $+$ sign in the superscript denotes the pseudo-inverse.

After applying NPT, we continue by considering Φ as our input data. This allows as to use the linear E-SVDD formulation to obtain a non-linear transformation.

B. TEST PHASE

During the test phase of the linear case, a test instance $\mathbf{x}_*$ is first mapped to the optimized lower d-dimensional space as

$$\mathbf{y}_* = \mathbf{Q}\mathbf{x}_*. \quad (32)$$

The decision to classify the instance as target or outlier is taken on the basis of its distance from the center of data description in the d-dimensional space. The distance is calculated as follows:

$$\|\mathbf{E}^{-\frac{1}{2}}\mathbf{y}_* - \mathbf{u}\|_2^2 = (\mathbf{E}^{-\frac{1}{2}}\mathbf{y}_*)^\intercal\mathbf{E}^{-\frac{1}{2}}\mathbf{y}_* - 2(\mathbf{E}^{-\frac{1}{2}}\mathbf{y}_*)^\intercal\mathbf{u} + \mathbf{u}^\intercal\mathbf{u}, \quad (33)$$

where $\mathbf{u}$ can be solved with (16). If $\|\mathbf{E}^{-\frac{1}{2}}\mathbf{y}_* - \mathbf{u}\|_2^2 \leq R^2$, the test instance is classified as positive, as it will fall inside the boundary of the data description. The test instance is classified as negative if $\|\mathbf{E}^{-\frac{1}{2}}\mathbf{y}_* - \mathbf{u}\|_2^2 > R^2$. The threshold R^2 for taking the decision is calculated as follows:

$$R^2 = (\mathbf{E}^{-\frac{1}{2}}\mathbf{s})^\intercal\mathbf{E}^{-\frac{1}{2}}\mathbf{s} - 2\mathbf{u}^\intercal\mathbf{s} + \mathbf{u}^\intercal\mathbf{u}, \quad (34)$$

where $\mathbf{s}$ is any support vector with $0 < \alpha_i < C$.

During the test phase for non-linear ES-SVDD, we use NPT by first computing the kernel vector as

$$\mathbf{k}_* = \Phi^\intercal\phi(\mathbf{x}_*). \quad (35)$$

The kernel vector is then centered as

$$\hat{\mathbf{k}}_* = (\mathbf{I} - \mathbf{J})[\mathbf{k}_* - \frac{1}{N}\mathbf{K}\mathbf{1}]. \quad (36)$$

The centered kernel vector is then mapped to

$$\phi_* = (\Phi^\intercal)^+\hat{\mathbf{k}}_* \quad (37)$$

We now consider ϕ_* as the test input $\mathbf{x}_*$ and follow all the steps described for the linear test.

TABLE 1. Datasets used in the experiments.

Abbreviation	Dataset Name (Target Class)	Total Samples	Target Samples	D
S-K	Seeds (Kama)	210	70	7
S-R	Seeds (Rosa)	210	70	7
S-C	Seeds (Canadian)	210	70	7
QB-B	Qualitative bankruptcy (bankruptcy)	250	107	6
QB-N	Qualitative bankruptcy (non-bankruptcy)	250	143	6
SH-H	Somerville happiness (happy)	143	77	6
SH-U	Somerville happiness (un-happy)	143	66	6
I-S	Iris (Setosa)	150	50	4
I-VC	Iris (Versicolor)	150	50	4
I-V	Iris (Virginica)	150	50	4
IS-B	Ionosphere (bad)	351	126	34
IS-G	Ionosphere (good)	351	225	34
SR-R	Sonar (rock)	208	97	60
SR-M	Sonar (mines)	208	111	60

IV. EXPERIMENTS

A. DATASETS AND EXPERIMENTAL SETUP

We evaluated the proposed and competing methods over different datasets downloaded from UCI machine learning repository [44]. Since one-class classification methods inherently are suited for binary (target and outliers) imbalanced classification problems, we converted the datasets to one-class datasets by considering a single class in a dataset at a time as the target class and all other classes as outliers. Naturally, only the target class samples were used for training the models, while all the classes were considered in the validation and test phases. The total number of samples, number of target class samples, and number of dimensions in the original feature space are given in Table 1.

In each dataset, 70% of the data was used for training and the remaining 30% for testing. The train and test sets were selected randomly by keeping the proportions of classes similar to the full dataset. Each experiment was repeated five times using different random train/test splits, while the same five splittings were used for all the considered methods. We report the average test performance over the five splittings. During training, a 5-fold cross-validation technique was used to select the best hyperparameters with the best evaluation score. We used only the training sets for selecting the hyperparameters. We used Geometric mean (*Gmean*) as the evaluation metric for all the methods. *Gmean* is defined as

$$Gmean = \sqrt{tpr \times tnr}, \tag{38}$$

where *tpr* is true positive rate (also known as sensitivity) and *tnr* is true negative rate (also known as specificity). For the proposed ES-SVDD method, we chose the hyperparameters from the following values

- $\beta \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}, 10^{4}\}$,
- $C \in \{0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$,
- $\sigma \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$,
- $d \in \{1, 2, 3, 4, 5, 10, 20, 50, 100\}$,
- $\eta \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$.

For all the competing methods, the hyperparameters corresponding to ES-SVDD hyperparameters were selected from the above values. For other hyperparameters, the same ranges were used as provided in the corresponding work or stated otherwise. We used the target class samples of the full training set with the optimal hyperparameters for the final training and then tested with the test set.

We compared the proposed ES-SVDD with other support vector (SV)-based and non-SV-based methods. The SV-based one-class classification methods essentially create a model by defining a boundary. The SV-based methods used for comparison were S-SVDD [32], OC-SVM [12], SVDD [13], and E-SVDD. The non-SV-based methods used for comparison were density-based, reconstruction-based, and regression-based one-class classification approaches. The density-based methods used for comparison were Parzen density-based data description [7] and Gaussian density-based data description [7]. As reconstruction-based methods, we used SOM data description [7] and K-means data description [7]. The regression-based method used for comparison was Graph Embedded One-Class Extreme Learning Machines (GE-OC-ELM) which exploits geometric class information [5].

We used maximum likelihood estimation for finding the optimum smoothing parameter in the Parzen density-based data description. The grid-size in SOM was fixed to $5\sqrt{N_t}$, where N_t is the size of training data for a given dataset [45]. We chose the number of clusters (N_c) for K-means from $N_c = \{1, 2, 3\}$ and report the best outcome. For non-linear methods, we employed NPT for ES-SVDD and S-SVDD, kernel whitening for Gaussian data description [46], and the kernel trick for other methods. Since the closest counterpart of the proposed method is S-SVDD and different regularization terms for S-SVDD were proposed [32], we report the results with all the previously proposed variants of S-SVDD. We used LIBSVM [47] toolbox implementation for OC-SVM and SVDD and DD-toolbox [48] for SOM, K-means data description, Parzen density, and Gaussian density-based data description. The implementation of GE-OC-ELM is publicly available.[1] The proposed ES-SVDD, along with S-SVDD

[1] https://sites.google.com/view/iosifidis/codes-and-datasets

TABLE 2. ***Gmean*** **results for linear methods over different datasets.**

Dataset	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.
ES-SVDD $\psi_0\Upsilon_0$	0.83	0.91	0.77	0.84	0.84	0.26	0.55	0.41	0.51	0.46
ES-SVDD ψ_1	0.83	0.89	0.89	0.87	0.76	0.46	**0.61**	0.47	0.47	0.47
ES-SVDD ψ_2	0.82	0.89	0.87	0.86	0.85	0.18	0.51	0.53	0.51	0.52
ES-SVDD ψ_3	0.79	0.90	0.87	0.86	0.90	0.23	0.57	0.46	0.35	0.40
ES-SVDD Υ_1	0.82	0.92	0.90	0.88	0.85	0.32	0.58	0.46	0.52	0.49
ES-SVDD Υ_2	0.84	0.91	**0.91**	**0.89**	0.81	0.30	0.56	**0.55**	**0.53**	**0.54**
ES-SVDD Υ_3	**0.85**	0.88	0.88	0.87	0.87	0.33	0.60	0.49	0.47	0.48
S-SVDD ψ_0	0.79	0.86	0.81	0.82	0.72	0.50	**0.61**	0.49	0.48	0.48
S-SVDD ψ_1	0.72	0.76	0.77	0.75	0.85	0.34	0.59	0.46	0.46	0.46
S-SVDD ψ_2	0.81	0.82	0.77	0.80	0.75	0.40	0.58	0.47	0.48	0.48
S-SVDD ψ_3	0.80	**0.93**	0.75	0.83	0.72	0.41	0.57	0.49	0.46	0.48
SVDD	0.82	0.92	0.86	0.87	0.83	0.04	0.43	0.54	0.48	0.51
E-SVDD	0.80	0.87	0.86	0.84	**0.96**	0.20	0.58	0.54	0.41	0.48
OC-SVM	0.43	0.46	0.58	0.49	0.46	**0.55**	0.51	0.45	0.42	0.44
Non-support-vector-based methods										
K-means	**0.86**	**0.94**	**0.91**	**0.90**	0.71	0.41	0.56	0.56	0.39	0.47
Parzen	0.49	0.33	0.58	0.47	**0.98**	**0.60**	**0.79**	**0.58**	**0.43**	**0.50**

Dataset	I-S	I-VC	I-V	Av.	IS-B	IS-G	Av.	SR-R	SR-M	Av.
ES-SVDD $\psi_0\Upsilon_0$	0.64	0.75	0.70	0.70	0.16	0.89	0.52	**0.50**	0.64	0.57
ES-SVDD ψ_1	0.92	0.86	0.77	0.85	**0.52**	0.85	**0.69**	**0.50**	0.56	0.53
ES-SVDD ψ_2	0.87	0.82	0.79	0.83	0.31	0.87	0.59	0.48	**0.67**	**0.58**
ES-SVDD ψ_3	0.93	0.87	0.71	0.84	0.35	0.89	0.62	0.48	0.65	0.57
ES-SVDD Υ_1	0.85	0.84	0.86	0.85	0.26	0.87	0.57	0.47	**0.67**	0.57
ES-SVDD Υ_2	**0.96**	0.83	0.74	0.84	0.31	0.89	0.60	0.47	0.65	0.56
ES-SVDD Υ_3	0.80	0.85	0.79	0.81	0.35	**0.90**	0.62	0.49	0.65	0.57
S-SVDD ψ_0	0.87	0.75	0.64	0.76	0.16	0.75	0.46	0.37	0.37	0.37
S-SVDD ψ_1	0.88	0.81	0.75	0.81	0.50	0.71	0.61	0.44	0.36	0.40
S-SVDD ψ_2	0.87	0.84	0.58	0.76	0.43	0.72	0.58	0.46	0.40	0.43
S-SVDD ψ_3	0.81	0.68	0.63	0.71	0.27	0.66	0.46	0.46	0.41	0.43
SVDD	0.94	**0.90**	**0.89**	**0.91**	0.04	0.73	0.39	**0.50**	0.52	0.51
E-SVDD	0.89	0.88	0.86	0.88	0.33	0.00	0.17	0.00	0.00	0.00
OC-SVM	0.50	0.52	0.39	0.47	0.47	0.45	0.46	0.44	0.52	0.48
Non-support-vector-based methods										
K-means	**0.94**	**0.92**	**0.89**	**0.91**	**0.37**	**0.88**	**0.63**	**0.49**	**0.68**	**0.58**
Parzen	0.85	0.68	0.79	0.77	0.32	0.25	0.28	0.00	0.00	0.00

and E-SVDD, was implemented by the authors using Matlab by leveraging LIBSVM.

B. EXPERIMENTAL RESULTS AND DISCUSSION

In Tables 2 and 3, we report the average test results for each dataset for the linear and non-linear cases, respectively. In each experiment, a single class was selected as the target class and the rest of the data as outliers (see Table 1). We also report the average performance of the proposed and competing methods in the average (Av.) column by averaging the results for a given dataset. For example, the performance over S-K, S-R, and S-C is averaged and provided in the Av. column as the overall performance for Seeds dataset. In this way, we can get an idea of the overall performance for each algorithm over the full dataset. For ES-SVDD and S-SVDD, we report the test results after 10 training iterations.

When compared to SV-based methods, our proposed methods achieved the best average results on all but Iris dataset in the linear case and on half of the datasets in the non-linear case. We note that the average results for the non-linear methods are generally better than those of the linear ones for the majority of the datasets. Overall, the proposed (linear and non-linear) methods achieved the best average results in 4 out of 6 datasets among the SV-based methods. In general, the best performing methods vary for different datasets, but we can see that there is no case, where the proposed method would fail completely, unlike most of the competing methods. In the linear case, other SV-based competing methods outperformed ES-SVDD only with Iris dataset, which has the lowest original dimensionality and also a low number of samples. Also in the non-linear case, other SV-based methods outperformed ES-SVDD most clearly on the 2 smallest datasets. Thus, it seems that the proposed method is more beneficial when the data dimensionality is higher.

When compared with also non-SV-based methods, we see that the proposed method gave the best average performance on 3 out of 6 datasets in the linear case. In the non-linear case, the ranking of the methods varies more and only GE-SVM achieved the best average results on more than one (2) datasets. The proposed method outperformed the other methods on Ionosphere dataset, which is the largest considered dataset. Furthermore, the stable performance of the proposed method makes it a viable solution also in the non-linear case.

Comparing regularization terms for linear ES-SVDD, we notice that ES-SVDD performed better in majority of cases with regularization term Υ_2 which uses samples belonging to the boundary and outside the boundary to describe the

TABLE 3. ***Gmean*** **results for non-linear methods over different datasets.**

Dataset	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.
ES-SVDD $\psi_0\Upsilon_0$	0.78	0.88	**0.93**	0.87	0.83	0.61	0.72	0.52	0.42	0.47
ES-SVDD ψ_1	0.80	0.88	0.88	0.85	0.80	0.34	0.57	0.51	0.42	0.47
ES-SVDD ψ_2	0.80	0.90	**0.93**	0.87	0.90	0.35	0.62	0.52	0.33	0.42
ES-SVDD ψ_3	0.82	0.86	0.72	0.80	0.89	0.64	**0.76**	0.52	0.45	0.49
ES-SVDD Υ_1	0.85	**0.92**	0.91	**0.89**	0.87	0.47	0.67	0.47	0.38	0.43
ES-SVDD Υ_2	0.82	0.88	0.89	0.86	0.84	**0.68**	**0.76**	0.52	0.34	0.43
ES-SVDD Υ_3	0.85	0.88	0.91	0.88	0.87	0.54	0.71	0.52	0.45	0.49
S-SVDD ψ_0	0.74	0.74	0.83	0.77	0.23	0.49	0.36	0.45	0.29	0.37
S-SVDD ψ_1	0.71	0.78	0.81	0.77	0.11	0.08	0.09	0.39	0.32	0.35
S-SVDD ψ_2	0.72	0.85	0.83	0.80	0.36	0.37	0.37	0.47	0.32	0.40
S-SVDD ψ_3	0.60	0.76	0.76	0.71	0.36	0.40	0.38	0.46	0.29	0.37
SVDD	**0.86**	0.91	0.88	0.88	0.88	0.55	0.71	0.54	**0.48**	0.51
E-SVDD	0.84	0.85	0.85	0.85	**0.96**	0.51	0.74	0.55	0.42	0.48
GE-SVDD	0.84	0.90	0.76	0.83	0.94	0.17	0.55	0.54	0.47	0.50
OC-SVM	0.79	0.60	0.63	0.67	0.67	0.52	0.59	**0.57**	**0.48**	**0.52**
GE-SVM	0.83	0.88	0.89	0.87	0.88	0.58	0.73	**0.57**	0.42	0.50
Non-support-vector-based methods										
SOM	0.80	0.90	0.89	0.86	0.79	0.37	0.58	0.28	0.26	0.27
Gaussian	**0.85**	**0.95**	**0.94**	**0.91**	0.63	0.46	0.54	**0.52**	**0.42**	**0.47**
GE-OC-ELM	**0.85**	0.93	0.89	0.89	**1.00**	**0.80**	**0.90**	0.31	0.31	0.31

Dataset	I-S	I-VC	I-V	Av.	IS-B	IS-G	Av.	SR-R	SR-M	Av.
ES-SVDD $\psi_0\Upsilon_0$	0.93	0.84	**0.86**	0.88	0.44	0.89	0.67	0.41	0.67	0.54
ES-SVDD ψ_1	**0.94**	0.81	0.74	0.83	**0.71**	0.90	**0.80**	0.48	0.55	0.51
ES-SVDD ψ_2	0.91	0.87	0.83	0.87	0.31	0.87	0.59	0.47	0.66	0.56
ES-SVDD ψ_3	0.89	0.84	0.74	0.82	0.32	0.88	0.60	0.47	0.66	0.57
ES-SVDD Υ_1	0.81	0.89	0.70	0.80	0.47	0.86	0.67	0.53	0.65	0.59
ES-SVDD Υ_2	0.91	0.83	0.81	0.85	0.68	0.86	0.77	0.47	**0.70**	0.58
ES-SVDD Υ_3	**0.94**	0.88	0.83	0.88	0.45	0.85	0.65	0.41	**0.70**	0.55
S-SVDD ψ_0	0.92	0.85	0.78	0.85	0.24	0.53	0.38	0.43	0.41	0.42
S-SVDD ψ_1	0.89	0.89	0.63	0.80	0.68	0.64	0.66	0.20	0.48	0.34
S-SVDD ψ_2	0.91	0.84	0.77	0.84	0.21	0.61	0.41	0.40	0.52	0.46
S-SVDD ψ_3	0.92	0.85	0.73	0.83	0.35	0.62	0.49	0.37	0.16	0.27
SVDD	**0.94**	**0.91**	0.84	**0.89**	0.31	0.80	0.55	0.53	0.66	0.59
E-SVDD	0.89	0.84	**0.86**	0.86	0.30	0.00	0.15	0.00	0.00	0.00
GE-SVDD	0.91	0.88	0.85	0.88	0.26	0.81	0.54	**0.56**	0.66	**0.61**
OC-SVM	0.45	0.65	0.66	0.59	0.27	0.63	0.45	0.51	0.58	0.54
GE-SVM	0.92	0.90	**0.86**	**0.89**	0.39	**0.91**	0.65	0.54	0.67	**0.61**
Non-support-vector-based methods										
SOM	0.91	0.84	**0.88**	**0.88**	0.06	**0.87**	0.47	0.46	0.32	0.39
Gaussian	0.97	0.86	0.80	**0.88**	0.33	0.50	0.42	**0.47**	**0.59**	**0.53**
GE-OC-ELM	**0.99**	**0.89**	0.78	**0.88**	**0.48**	0.81	**0.65**	0.38	0.55	0.47

covariance of the class. Regularization term ψ_1, which uses all training samples to describe the covariance of the class, also performed well. Both of these regularization terms produced 2 out of 6 best results in the linear case. We also noticed that ES-SVDD without any regularization term performs the worst as compared to ES-SVDD with regularization terms.

For non-linear ES-SVDD, the regularization terms ψ_1 and Υ_3 resulted in the best results for most of the datasets. However, ψ_1 is also noticed to perform worse than the others in a few datasets. ψ_1 uses all target training samples in describing the covariance of the class without taking into account the concentration matrix. In Υ_3, the $\boldsymbol{\lambda}$ values take the values of α_i values of the boundary samples and zero for non-boundary samples. In the non-linear case for high dimensional datasets, we notice that using all the training data for describing the covariance of the data in a projected space, with or without using the concentration matrix (i.e., ψ_1 or Υ_1), yielded the best results for ES-SVDD.

We further notice that by considering the class variance taking into account the concentration matrix in the regularization term, ES-SVDD performed better in most datasets as compared to the regularization terms without considering the concentration matrix. Overall Υ_2 is found to be more robust that other regularization terms. Hence, we recommend to use samples belonging to the boundary and outside the boundary to describe the covariance of the class while taking into account the concentration matrix.

We also show the performance of the proposed ES-SVDD and the recently proposed S- SVDD on the test set after every training iteration for the linear and non-linear cases. We compare the performances of these methods with different regularization terms Υ and ψ. The average *Gmean* value is calculated for each iteration over the 5 test splits for the different datasets, see Figures 1-6.

It can clearly be seen from the figures that for both the linear and non-linear methods, ES-SVDD achieves its best performance much earlier than the recently proposed counterpart S-SVDD. This is not surprising, because the ellipsoidal description can fit a larger variety of data distributions,

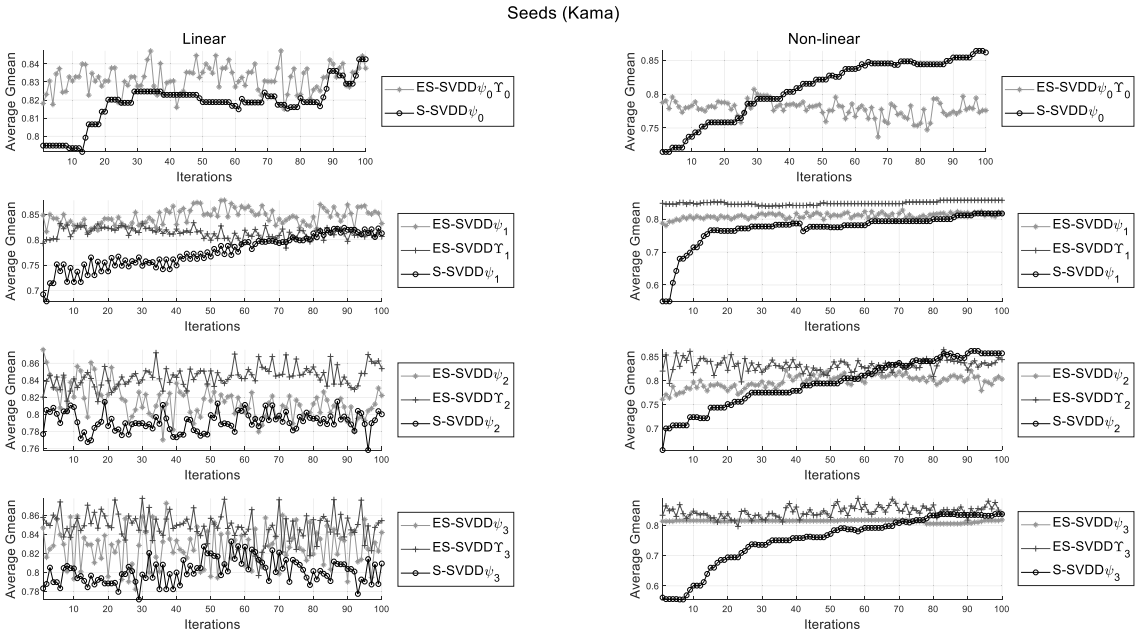

FIGURE 1. Comparison of different regularization terms for ES-SVDD and S-SVDD on dataset S-K.

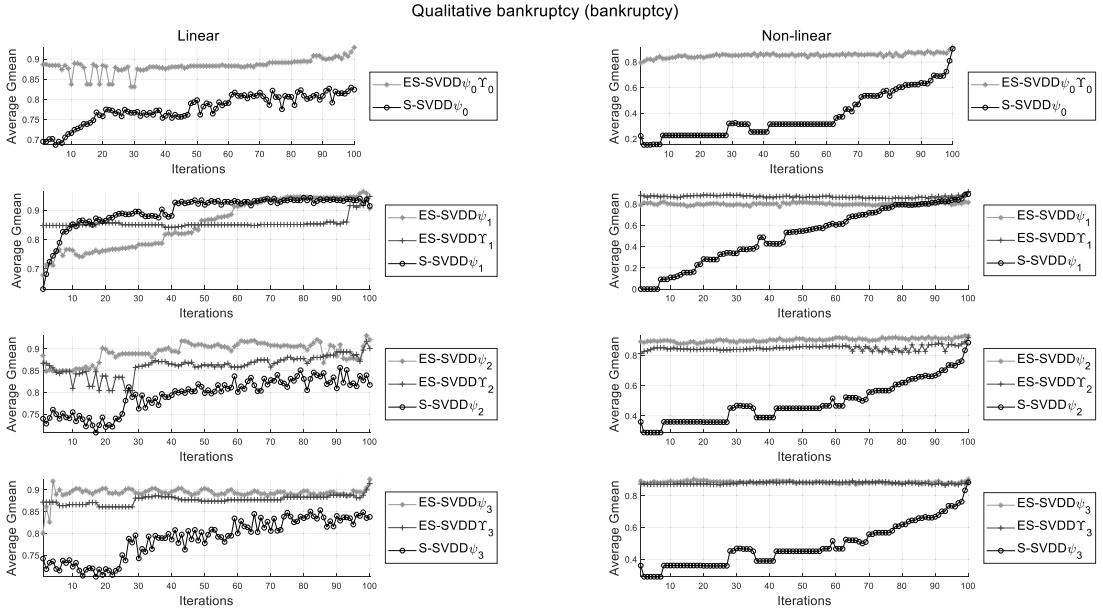

FIGURE 2. Comparison of different regularization terms for ES-SVDD and S-SVDD on dataset QB-B.

while the optimal spherical description gets successful only after the data variance for different dimensions has been equalized. Using the ellipsoidal data description in the proposed method makes it converge faster to an optimal solution. We also notice that for high dimensional datasets ES-SVDDψ_1 and ES-SVDDΥ_1 are more stable as compared

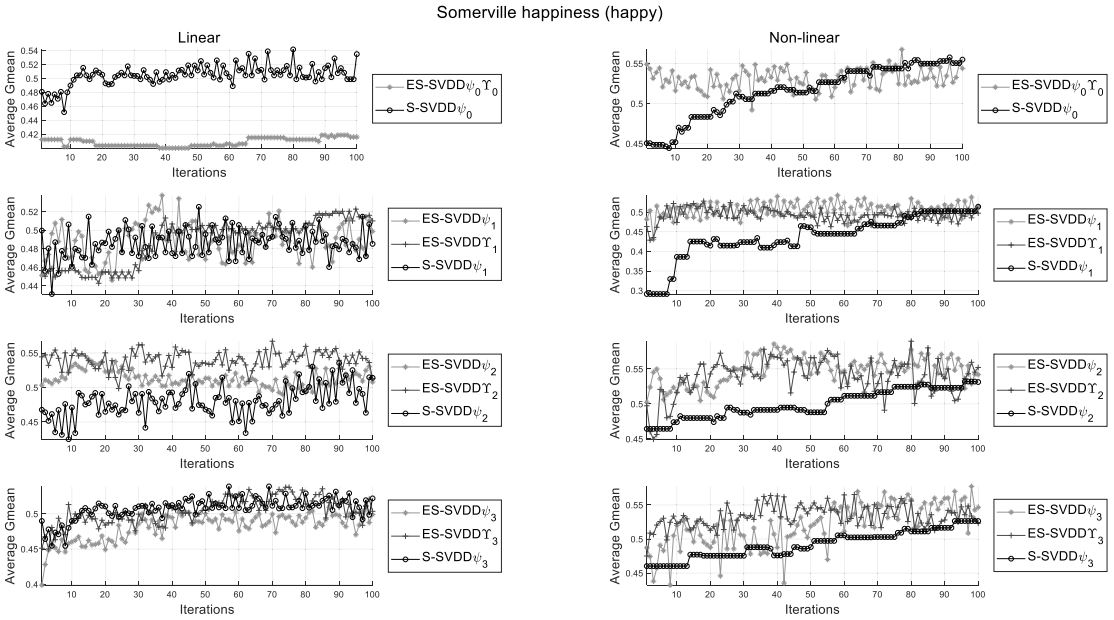

FIGURE 3. Comparison of different regularization terms for ES-SVDD and S-SVDD on dataset SH-H.

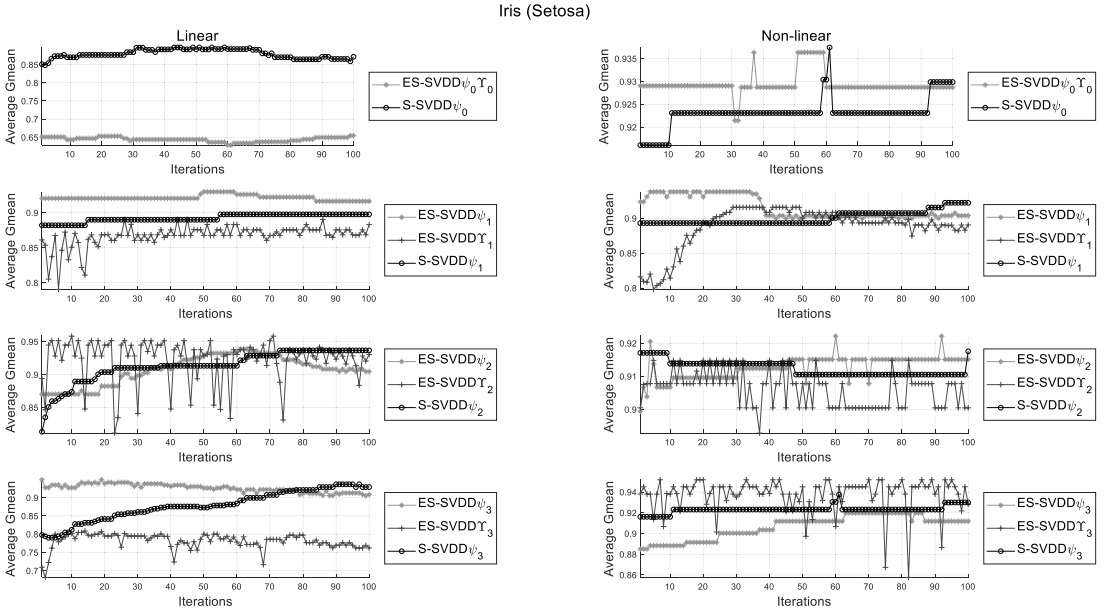

FIGURE 4. Comparison of different regularization terms for ES-SVDD and S-SVDD on dataset I-S.

to the other proposed linear and non-linear methods. Overall, the trend of faster convergence and higher stability in terms of producing consistent results for different range of iterations for ES-SVDD can be observed both in the linear and non-linear methods for all regularization terms in the majority of the cases.

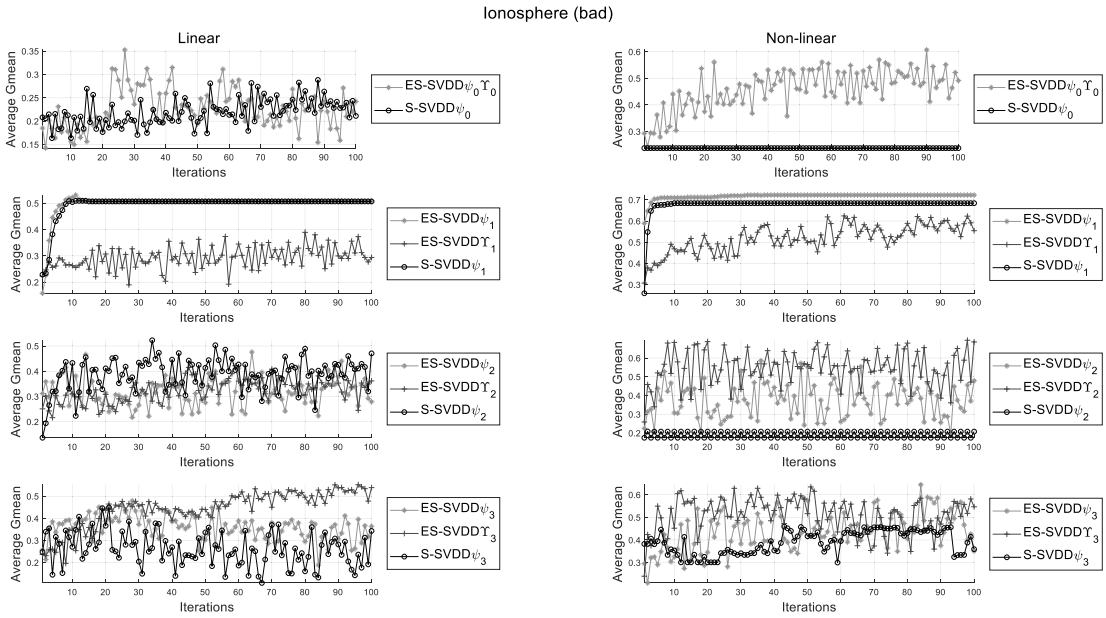

FIGURE 5. Comparison of different regularization terms for ES-SVDD and S-SVDD on dataset IS-B.

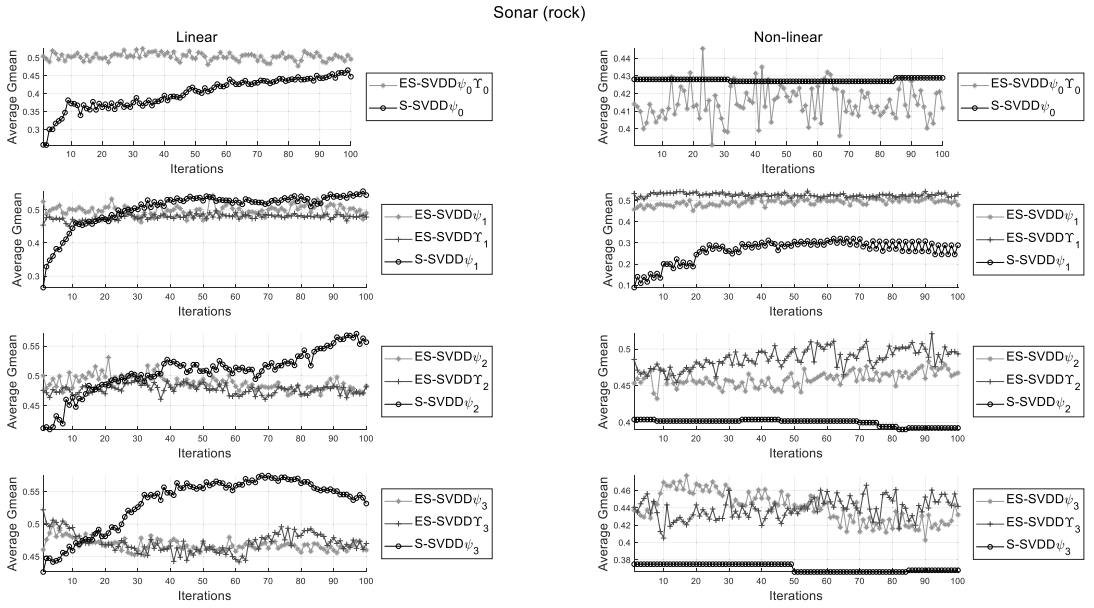

FIGURE 6. Comparison of different regularization terms for ES-SVDD and S-SVDD on dataset SR-R.

V. CONCLUSION

In this paper, a novel method, ES-SVDD, for one-class classification is proposed. The proposed method projects the data from an input feature space to a new optimized subspace suitable for one-class classification. The proposed method generalizes S-SVDD for a hypersphere by using ellipsoidal data description. We proposed different regularization terms along with linear and non-linear formulations of the method.

In most cases, the proposed ES-SVDD variants outperform the competing SV-based methods and converge faster than in the case of data description without ellipsoidal encapsulation.

In the future, we intend to use other kernel types in the non-linear case of ES-SVDD. We also plan to devise a strategy for early exit in the training process to reduce the training time. We will also experiment with finetuning hyperparameters according to different criteria, such as area under receiver operating characteristic curve. Additionally, we plan to formulate and implement a neural network based version of the proposed method and compare its performance with deep neural networks.

REFERENCES

[1] T. Hastie, R. Tibshirani, and J. Friedman, "Unsupervised learning," in *The Elements of Statistical Learning*. New York, NY, USA: Springer, 2009, pp. 485–585.

[2] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerg. Artif. Intell. Appl. Comput. Eng.*, vol. 160, no. 1, pp. 3–24, 2007.

[3] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Amsterdam, The Netherlands: Elsevier, 2013.

[4] Y. Yang, C. Hou, Y. Lang, G. Yue, and Y. He, "One-class classification using generative adversarial networks," *IEEE Access*, vol. 7, pp. 37970–37979, 2019.

[5] A. Iosifidis, V. Mygdalis, A. Tefas, and I. Pitas, "One-class classification based on extreme learning and geometric class information," *Neural Process. Lett.*, vol. 45, no. 2, pp. 577–592, Apr. 2017.

[6] V. Mygdalis, A. Iosifidis, A. Tefas, and I. Pitas, "Laplacian one class extreme learning machines for human action recognition," in *Proc. IEEE 18th Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2016, pp. 1–5.

[7] D. M. J. Tax, "One-class classification: Concept learning in the absence of counter-examples," ASCI dissertation, Delft Univ. Technol., Delft, The Netherlands, 2001, vol. 65, pp. 1–190.

[8] K. Lee, D.-W. Kim, K. H. Lee, and D. Lee, "Density-induced support vector data description," *IEEE Trans. Neural Netw.*, vol. 18, no. 1, pp. 284–289, Jan. 2007.

[9] C. Fraley and A. E. Raftery, "Model-based clustering, discriminant analysis, and density estimation," *J. Amer. Stat. Assoc.*, vol. 97, no. 458, pp. 611–631, Jun. 2002.

[10] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[11] T. Kohonen and P. Somervuo, "Self-organizing maps of symbol strings," *Neurocomputing*, vol. 21, nos. 1–3, pp. 19–30, Nov. 1998.

[12] B. Schölkopf, R. Williamson, A. Smola, and J. Shawe-Taylor, "SV estimation of a distribution's support," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 12, 1999, pp. 1–7.

[13] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, no. 1, pp. 45–66, Jan. 2004.

[14] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, "Deep one-class classification," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4393–4402.

[15] J. Zhao, H. Lin, Q. Chen, X. Huang, Z. Sun, and F. Zhou, "Identification of egg's freshness using NIR and support vector data description," *J. Food Eng.*, vol. 98, no. 4, pp. 408–414, Jun. 2010.

[16] H. Lee and W. Chung, "Terrain classification for mobile robots on the basis of support vector data description," *Int. J. Precis. Eng. Manuf.*, vol. 19, no. 9, pp. 1305–1315, Sep. 2018.

[17] V. Mygdalis, A. Iosifidis, A. Tefas, and I. Pitas, "Graph embedded one-class classifiers for media data classification," *Pattern Recognit.*, vol. 60, pp. 585–595, Dec. 2016.

[18] J. Wang, W. Liu, K. Qiu, H. Xiong, and L. Zhao, "Dynamic hypersphere SVDD without describing boundary for one-class classification," *Neural Comput. Appl.*, vol. 31, no. 8, pp. 3295–3305, Aug. 2019.

[19] T. Kenaza, K. Bennaceur, and A. Labed, "An efficient hybrid SVDD/clustering approach for anomaly-based intrusion detection," in *Proc. 33rd Annu. ACM Symp. Appl. Comput. (SAC)*, 2018, pp. 435–443.

[20] V. Mygdalis, A. Iosifidis, A. Tefas, and I. Pitas, "Semi-supervised subclass support vector data description for image and video classification," *Neurocomputing*, vol. 278, pp. 51–61, Feb. 2018.

[21] M. Rahmanimanesh, J. A. Nasiri, S. Jalili, and N. M. Charkari, "Adaptive three-phase support vector data description," *Pattern Anal. Appl.*, vol. 22, no. 2, pp. 491–504, May 2019.

[22] B. Liu, Y. Xiao, L. Cao, Z. Hao, and F. Deng, "SVDD-based outlier detection on uncertain data," *Knowl. Inf. Syst.*, vol. 34, no. 3, pp. 597–618, Mar. 2013.

[23] A. Banerjee, P. Burlina, and R. Meth, "Fast hyperspectral anomaly detection via SVDD," in *Proc. IEEE Int. Conf. Image Process.*, vol. 4, 2007, p. IV-101.

[24] Y.-H. Liu, Y.-C. Liu, and Y.-J. Chen, "Fast support vector data descriptions for novelty detection," *IEEE Trans. Neural Netw.*, vol. 21, no. 8, pp. 1296–1313, Aug. 2010.

[25] F. Camci and R. B. Chinnam, "General support vector representation machine for one-class classification of non-stationary classes," *Pattern Recognit.*, vol. 41, no. 10, pp. 3021–3034, Oct. 2008.

[26] Y. Forghani, S. Effati, H. S. Yazdi, and R. S. Tabrizi, "Support vector data description by using hyper-ellipse instead of hyper-sphere," in *Proc. 1st Int. eConf. Comput. Knowl. Eng. (ICCKE)*, Oct. 2011, pp. 22–27.

[27] M. GhasemiGol, R. Monsefi, and H. S. Yazdi, "Ellipse support vector data description," in *Proc. Int. Conf. Eng. Appl. Neural Netw.* Berlin, Germany: Springer, 2009, pp. 257–268.

[28] K. Wang and H. Xiao, "Ellipsoidal data description," *Neurocomputing*, vol. 238, pp. 328–339, May 2017.

[29] M. Ghasemigol, R. Monsefi, and H. Sadoghi-Yazdi, "Intrusion detection by ellipsoid boundary," *J. Netw. Syst. Manage.*, vol. 18, no. 3, pp. 265–282, Sep. 2010.

[30] E. Rimon and S. P. Boyd, "Obstacle collision detection using best ellipsoid fit," *J. Intell. Robot. Syst.*, vol. 18, no. 2, pp. 105–126, 1997.

[31] K. Wang, H. Xiao, and Y. Fu, "Ellipsoidal support vector data description in kernel PCA subspace," in *Proc. 3rd Int. Conf. Digit. Inf. Process., Data Mining, Wireless Commun. (DIPDMWC)*, Jul. 2016, pp. 13–18.

[32] F. Sohrab, J. Raitoharju, M. Gabbouj, and A. Iosifidis, "Subspace support vector data description," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 722–727.

[33] N. Kwak, "Nonlinear projection trick in kernel methods: An alternative to the kernel trick," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 12, pp. 2113–2119, Dec. 2013.

[34] R. Chen, N. Sun, X. Chen, M. Yang, and Q. Wu, "Supervised feature selection with a stratified feature weighting method," *IEEE Access*, vol. 6, pp. 15087–15098, 2018.

[35] H. Luo and J. Han, "Trace ratio criterion based large margin subspace learning for feature selection," *IEEE Access*, vol. 7, pp. 6461–6472, 2019.

[36] M. U. Chaudhry and J.-H. Lee, "Feature selection for high dimensional data using Monte Carlo tree search," *IEEE Access*, vol. 6, pp. 76036–76048, 2018.

[37] Z. Zhao, L. Wang, H. Liu, and J. Ye, "On similarity preserving feature selection," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 3, pp. 619–632, Mar. 2013.

[38] N. Zhou, Y. Xu, H. Cheng, J. Fang, and W. Pedrycz, "Global and local structure preserving sparse subspace learning: An iterative approach to unsupervised feature selection," *Pattern Recognit.*, vol. 53, pp. 87–101, May 2016.

[39] R. Vijayanand and D. Devaraj, "A novel feature selection method using whale optimization algorithm and genetic operators for intrusion detection system in wireless mesh network," *IEEE Access*, vol. 8, pp. 56847–56854, 2020.

[40] Q. Gu, Z. Li, and J. Han, "Joint feature selection and subspace learning," in *Proc. 22nd Int. Joint Conf. Artif. Intell.*, 2011, pp. 1294–1299.

[41] M. Bacher, I. Ben-Gal, and E. Shmueli, "Subspace selection for anomaly detection: An information theory approach," in *Proc. IEEE Int. Conf. Sci. Electr. Eng. (ICSEE)*, Nov. 2016, pp. 1–5.

[42] H. V. Nguyen, E. Muller, and K. Bohm, "4S: Scalable subspace search scheme overcoming traditional apriori processing," in *Proc. IEEE Int. Conf. Big Data*, Oct. 2013, pp. 359–367.

[43] K. B. Petersen and M. S. Pedersen. (Nov. 2012). *The Matrix Cookbook, Version 20121115*. [Online]. Available: https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf

[44] D. Dua and C. Graff. (2017). *UCI Machine Learning Repository*. [Online]. Available: http://archive.ics.uci.edu/ml

[45] J. Laaksonen and T. Honkela, *Advances in Self-Organizing Maps: 8th International Workshop, WSOM 2011, Espoo, Finland, June 13-15, 2011. Proceedings*, vol. 6731. Berlin, Germany: Springer, 2011.
[46] D. M. J. Tax and P. Juszczak, "Kernel whitening for one-class classification," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 17, no. 3, pp. 333–347, May 2003.
[47] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm
[48] D. M. J. Tax, "A MATLAB toolbox for data description, outlier and novelty detection version 1.7.5," Delft Univ. Technol., Delft, The Netherlands, Tech. Rep., Apr. 2010.

FAHAD SOHRAB (Graduate Student Member, IEEE) received the B.S. degree *(cum laude)* in telecommunication engineering from the National University of Computer and Emerging Sciences, Peshawar Pakistan, in August 2012, and the master's degree from Sabanci University, Istanbul, Turkey. He is currently pursuing the Ph.D. degree with Tampere University, Finland. In the fall of 2013, he joined the Computer Vision and Pattern Analysis Laboratory, Sabanci University. He was also affiliated with the Pattern Recognition Laboratory, Delft University of Technology, The Netherlands, during the second year of his Master's studies. He is currently with the Signal Analysis and Machine Intelligence Group, Tampere University. His research interests include machine learning, pattern recognition, subspace learning, and one-class classification.

JENNI RAITOHARJU (Member, IEEE) received the Ph.D. degree from the Tampere University of Technology, Finland, in 2017. Since then, she has worked as a Postdoctoral Research Fellow at the Faculty of Information Technology and Communication Sciences, Tampere University, Finland. In 2019, she started working as a Senior Research Scientist at the Finnish Environment Institute, Jyväskylä, Finland, after receiving Academy of Finland Postdoctoral Researcher funding for 2019–2022. She has coauthored 15 journal articles and 27 papers in international conferences. Her research interests include machine learning and pattern recognition methods along with applications in biomonitoring and autonomous systems. She has been the Chair of the Young Academy Finland, since 2019.

ALEXANDROS IOSIFIDIS (Senior Member, IEEE) received the B.Sc. degree in electrical and computer engineering and the M.Sc. degree with a specialisation in mechatronics from the Democritus University of Thrace, Greece, in 2008 and 2010, respectively, and the Ph.D. degree in computer science from the Aristotle University of Thessaloniki, Greece, in 2014. He is currently an Associate Professor of machine learning with the Department of Engineering, Aarhus University, Denmark. Before, he joined Aarhus University, he held a Postdoctoral Researcher positions at the Aristotle University of Thessaloniki and at the Tampere University of Technology, Finland, where he was an Academy of Finland Postdoctoral Research Fellow. He has contributed in more than 20 Research and Development projects financed by EU, Finnish and Danish funding agencies and companies. He has (co)authored 65 articles in international journals and 85 papers in international conferences proposing novel Machine Learning techniques and their application in a variety of problems. He has served as an Officer of the Finnish IEEE Signal Processing—Circuits and Systems Chapter from 2016 to 2018. His research interests include neural networks and statistical machine learning finding applications in computer vision, financial engineering, and graph mining problems. He is also a member of the EURASIP Technical Area Committee on Visual Information Processing. He serves as an Area/Associate Editor in *Neurocomputing*, *Signal Processing: Image Communications*, IEEE Access, and *BMC Bioinformatics* journals.

MONCEF GABBOUJ (Fellow, IEEE) received the B.S. degree in electrical engineering from Oklahoma State University, Stillwater, OK, USA, in 1985, and the M.S. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1986 and 1989, respectively. He was an Academy of Finland Professor, from 2011 to 2015. He is currently a Professor of signal processing with the Faculty of Information Technology and Communication Sciences, Tampere University, Finland. He guided 40 Ph.D. students and has published 650 articles. His current research interests include multimedia content-based analysis, indexing and retrieval, machine learning, nonlinear signal and image processing and analysis, voice conversion, and video processing and coding. He is a member of the Academia Europaea and the Finnish Academy of Science and Letters. He organized several tutorials and special sessions for major IEEE conferences and EUSIPCO. He is the past Chairman of the IEEE CAS TC on DSP and a Committee Member of the IEEE Fourier Award for Signal Processing. He has served as an Associate Editor and a Guest Editor for many IEEE international journals and a Distinguished Lecturer for the IEEE CASS.

PUBLICATION III

Multimodal subspace support vector data description

F. Sohrab, J. Raitoharju, A. Iosifidis and M. Gabbouj

Pattern Recognition110.(2021), 107648
DOI: https://doi.org/10.1016/j.patcog.2020.107648

Pattern Recognition 110 (2021) 107648

Contents lists available at ScienceDirect

Pattern Recognition

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

Multimodal subspace support vector data description

Fahad Sohrab[a,*], Jenni Raitoharju[a,b], Alexandros Iosifidis[c], Moncef Gabbouj[a]

[a] Faculty of Information Technology and Communication Sciences, Tampere University, FI-33720 Tampere, Finland
[b] Programme for Environmental Information, Finnish Environment Institute, FI-40500 Jyväskylä, Finland
[c] Department of Engineering, Electrical and Computer Engineering, Aarhus University, DK-8200 Aarhus, Denmark

ARTICLE INFO

Article history:
Received 21 August 2019
Revised 13 July 2020
Accepted 6 September 2020
Available online 10 September 2020

Keywords:
Feature transformation
Multimodal data
One-class classification
Support vector data description
Subspace learning

ABSTRACT

In this paper, we propose a novel method for projecting data from multiple modalities to a new subspace optimized for one-class classification. The proposed method iteratively transforms the data from the original feature space of each modality to a new common feature space along with finding a joint compact description of data coming from all the modalities. For data in each modality, we define a separate transformation to map the data from the corresponding feature space to the new optimized subspace by exploiting the available information from the class of interest only. We also propose different regularization strategies for the proposed method and provide both linear and non-linear formulations. The proposed Multimodal Subspace Support Vector Data Description outperforms all the competing methods using data from a single modality or fusing data from all modalities in four out of five datasets.

1. Introduction

In our surroundings, on a daily basis, we are exposed to information from many different sources. Different sensors are used to gather information about similar objects. Our brains usually perform well in combining the information from different sources to make a concise analysis of that particular entity. In order to analyze an entity, even a single source of information might be enough, but to make some critical decisions it is important to combine information from different sources in a systematic way. For example, if a person is walking in a crowd, the main information to not hit anything comes from visual cues, but people can warn each other also by voice or even by touch, and this extra information helps in understanding the environment in a better way. The smell could help to avoid unpleasant spots, too. As another example, while watching a movie, only visual information of the scenes may not be enough to understand the whole scenario, but the audio and/or captions combined together with the visuals information will provide the full information.

In machine learning techniques for predictive data modeling, training data are used to form a model that can accurately classify future instances into a predefined number of classes. In many cases, data comes from sensors and can be further processed to extract different features. The term *multimodal* is used to describe the data coming from different sensors (also referred to as mode or modality), however, it is also used as a synonym to *multi-view* when different features are extracted from the same sensor or when there are multiple similar sensors, e.g., cameras. The aim of multimodal machine learning algorithms is to build models that can process and relate information from more than one modality (or view).

The examples of multimodal representations are prevalent in different application areas. In [1], an active multimodal sensor system for target recognition and tracking is studied where information from three different sensors (visual, infrared, and hyperspectral) is used. In [2], a framework for vehicle tracking with multimodal data (velocity and images) is proposed where the outcome of velocity modality estimated by using a Kalman filter on the data obtained from motion sensors is fused with features learned from image modality by the color-faster R-CNN method. In [3], a multimodal data collection framework for mental stress monitoring is studied. In the proposed framework, physiological and motion sensor data of people under stress are collected.

* Corresponding author.
E-mail addresses: fahad.sohrab@tuni.fi (F. Sohrab), jenni.raitoharju@tuni.fi (J. Raitoharju), alexandros.iosifidis@eng.au.dk (A. Iosifidis), moncef.gabbouj@tuni.fi (M. Gabbouj).

https://doi.org/10.1016/j.patcog.2020.107648

The data in multimodal applications come from different modalities, where each modality has its own statistical properties and contains specific information. The different modalities usually share high-level concepts and semantic information, and all together contain more information than any single-modal data [4]. If we build a model separately for each modality, the relationship between the modalities cannot be exploited efficiently. In multimodal subspace learning, the goal is to infer a shared latent representation, that can accurately model data from each original modality and exploit the relationship between the modalities.

In traditional multiclass machine learning, an adequate amount of data are available for all the categories during training and, hence, the algorithm takes advantage of all available training data from all classes to train a model [5]. However, it is possible that during the training, data are highly imbalanced, or the only data available is from a single class. In such cases, one-class classification techniques are used. It is useful in many different cases, such as outlier detection, predicting specific events, or, in general, predicting a specific target class. While much effort has been put on solving one-class classification tasks for data of a single modality [6], much less effort has been put on solving one-class multimodal challenges in general, and we are not aware of any prior work in the field of multimodal learning for one-class classification. In one-class multimodal tasks, it is assumed that the only data available is from a single class in many different modalities.

In this paper, we propose a novel method for solving multimodal one-class classification tasks. The proposed method, Multimodal Subspace Support Vector Data Description (MS-SVDD), finds a transformation for each modality along with defining a common model for all modalities in a lower-dimensional subspace optimized for one-class classification. The rest of the paper is organized as follows. In Section 2, an overview of related work is presented. In Section 3, the newly proposed MS-SVDD is derived and discussed. In Section 4, we present the experimental setup and results, and finally, in Section 5, conclusions are drawn.

2. Background and related work

In this section, we briefly discuss the principles of multimodal learning, along with subspace learning. We also provide an overview of traditional methods used for multiclass multimodal data description and one-class unimodal data description.

2.1. Multimodal learning

The availability of many different modalities can be bliss if it increases the performance of the machine learning model. However, if the data description algorithm fails to make a strong connection between the different available modalities, the performance can be degraded. To ensure better performance of the model by combining data from different modalities, mainly two principles should be ensured, i.e., consensus and complementary principles [7]:

- **Consensus principle** aims at minimizing the disagreement between data available from different modes. Maximizing the agreement will reduce the error rate, and better modeling of data is achieved while combining data from different modalities.
- **Complementary principle** in the context of multimodal learning means that data from each modality may contain some knowledge not contained by the other ones. So it is necessary to exploit information from all the available modes to make an accurate description of data.

The multimodal machine learning techniques can be described by three main properties: two-view vs. multi-view, linear vs. nonlinear, and unsupervised vs. supervised [8]. As the name indicates, in two-view learning, the number of views is limited to two. In multi-view learning, the number of views is not limited. The difference between supervised and unsupervised learning is that, in supervised learning, the information on output labels of the training data is taken into account when training the model, while in unsupervised methods, the labels are not used to model the underlying structure or distribution of the data [9]. Linear techniques for multimodal subspace learning may be too simple to provide a representative model. Hence, kernel methods are proposed to capture non-linear patterns in data.

The multimodal learning techniques have been mainly applied on four applications domains [10]: i.e., audio-visual speech recognition [11], multimedia content indexing and retrieval [12], understanding human multimodal behaviors [13], and language and vision media description [14]. Recently, there has been a rising trend in applying multimodal machine learning algorithms also to other applications. For example, in [15], a multimodal data fusion technique is used for the prediction of soybean yield from an unmanned aerial vehicle.

In multimodal learning, the main goal is to develop a process of fusing information from various modalities. In [16], the fusion strategies are divided into two different categories as model-agnostic and model-based approaches. In model-agnostic approaches, the fusion is either late, early, or hybrid. In early fusion, the data or extracted features are fused together at the very initial phase of modeling. A new feature vector is usually formed by concatenating all the available data from different modes, and the model is trained with the new feature vector. In late fusion, multiple models are trained, and the fusion is done for scores generated by each model for the corresponding modality. The score generated by each model can be a threshold or some probability used in decision making. Hybrid fusion exploits the advantage of both early fusion and late fusion. Model-based approaches for fusion explicitly fuses data during their construction, such as kernel-based approaches, graphical models, and neural networks. In this work, we present a model-based approach for data fusion.

2.2. Subspace learning

In the current era of data science, where high-dimensional multimodal big data are generated every minute in different industries, there is a need to get the essential insights and mine knowledge in this high-dimensional data. Subspace learning aims at representing data in a lower-dimensional space by keeping intact all the information available in the original higher-dimensional space.

Algorithms developed for linear subspace learning find a projection matrix for labeled training data (represented by vectors) satisfying some optimality criteria. Principal Component Analysis (PCA) is one of the first subspace learning methods mentioned in literature. In PCA, a subspace is learned by orthogonally projecting data to a subspace so that the variance of data is maximized. PCA works only with a single mode of data, i.e., all data should be in the same dimension. Another traditional subspace learning method is Linear Discriminant Analysis (LDA), which finds a linear transformation by exploiting the class information.

Analogous to PCA, but used for two-view learning, is canonical-correlation analysis (CCA) [17]. CCA is a classic and conventional method for subspace learning, which aims at relating two sets of data by finding out the pairs of directions which provide a maximum correlation between the two sets. It has recently become one of the popular methods for unsupervised subspace learning because of its generalization capability and has been used extensively for multimodal data fusion and cross-media retrieval [18]. In subspace learning, state-of-the-art results are achieved by methods which have embraced some stimulus from conventional subspace learning methods [19].

As an extension of methods for linear transformation, kernel methods are introduced to describe nonlinear function or decision boundaries. In kernel methods, the data are mapped to a typically higher-dimensional kernel-space using a kernel function where it exhibits linear patterns [20,21]. For example, in [22], kernel-PCA performing a nonlinear form of PCA is proposed.

2.3. One-class classification

In one-class classification, the parameters of the model are estimated using data from the positive class only because data from the other classes are either not available at all or it is too diverse in nature to be modeled statistically [23]. The positive class is also called the target class, and the data from the other classes, which are not available during training, is called negative, or an outlier class. For example, a unimodal biometric system uses a single biometric trait for verification or identification [24].

Support Vector Data Description (SVDD) [25] is among the most widely used one-class classification methods used for anomaly detection and other related applications. SVDD obtains a spherical boundary around target data which can be made flexible by using the kernel trick. The obtained boundary is used to detect outliers during the test, i.e., anything inside the closed boundary is classified as a target class and otherwise as an outlier. The Lagrangian of SVDD is given as follows

$$L = \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^T \mathbf{x}_i - \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \mathbf{x}_i^T \mathbf{x}_j \alpha_j, \tag{1}$$

where $\mathbf{x}_i$ is the input target training instance and maximizing (1) gives a set of α_i corresponding to each instance. The instances with $\alpha_i \geq 0$ define the data description. Other common one-class classification method is One-Class Support Vector Machine (OC-SVM) [26].

Techniques for enhancing the performance of one-class classification methods, mainly extensions of SVDD, can be categorized into four main categories: methods based on data structure, kernel issue, boundary shape, and non-stationary data [27]. As the name indicates, in the data structure category, the main focus is on the structure of data. For example, in [28], a confidence coefficient is associated with each training sample to deal with the uncertainty of data. In kernel issue extensions, the main focus is on reducing the complexity or proposing new kernels for one-class classification. For example, in [29], a new kernel is proposed to improve the accuracy of SVDD for time series classification. Proposing changes in the boundary for enclosing the target data comes under the third category for improving one-class classification accuracy. For example, in [30], the ellipse shape is used for encapsulating target data instead of the traditional sphere used in SVDD. In [31], it is shown that both SVDD and OC-SVM lead to the same solution when exploiting the elliptical shape of the class. The last category of algorithms for improving one-class classifier performance attempts to handle non-stationary data. For example in [32], Incremental-SVDD (I-SVDD) is proposed to handle non-stationary or increasing data. Recently, in [33], an algorithm developed for reducing the effect of uncertain data around the hypersphere of SVDD achieved the state of the art result on many UCI [34] datasets. In this paper, we consider baseline SVDD combined with multimodal subspace learning. However, in the future, the method can be further extended using similar ideas.

In the area of multimodal one-class classification, researchers have mainly focused on fusing the output labels of multiple models trained for each type of feature independently, i.e., without taking into account information from other feature types for one model [35].

3. Multimodal subspace support vector data description

MS-SVDD maps data from high-dimensional feature spaces to a low-dimensional feature space optimized for one-class classification. The optimized subspace is shared by data coming from all modalities. MS-SVDD is an extension of Subspace Support Vector Data Description (S-SVDD), which was proposed for unimodal data in [36]. The main novelty of MS-SVDD is using the multimodal approach for one-class classification. Here, we first derive the linear MS-SVDD. Then we derive two non-linear versions using the kernel trick [20] and the Nonlinear Projection Trick (NPT) [37], respectively.

3.1. Linear MS-SVDD

Let us assume that the items to be modelled are represented by M different modalities. The instances in each modality m, $m = 1, \ldots, M$, are represented by $\mathbf{X}_m = [\mathbf{x}_{m,1}, \mathbf{x}_{m,2}, \ldots \mathbf{x}_{m,N}]$, $\mathbf{x}_{m,i} \in \mathbb{R}^{D_m}$, where N is the total number of instances and D_m is the dimensionality of the feature space in modality m. MS-SVDD tries to find a projection matrix $\mathbf{Q}_m \in \mathbb{R}^{d \times D_m}$ for each modality, which will project the corresponding instances to a lower (d)-dimensional optimized subspace shared by all modalities. Thus, a feature vector $\mathbf{x}_{m,i}$ is projected to a d-dimensional vector $\mathbf{y}_{m,i}$ as

$$\mathbf{y}_{m,i} = \mathbf{Q}_m \mathbf{x}_{m,i}, \forall m \in \{1, \ldots, M\}, \ \forall i \in \{1, \ldots, N\}. \tag{2}$$

To obtain a common description of all the data transformed from their corresponding modalities to the new common subspace, we exploit Support Vector Data Description (SVDD) [25] to form a closed boundary around the target class data in the new subspace. The center and radius of the hypersphere are denoted by $\mathbf{a} \in \mathbb{R}^d$ and R, respectively. Fig. 1 depicts the basic idea of the proposed method.

In order to find a compact hypersphere which encloses all the target data from all the modalities in the new subspace, we minimize

$$F(R, \mathbf{a}) = R^2$$

s.t.

$$\|\mathbf{Q}_m \mathbf{x}_{m,i} - \mathbf{a}\|_2^2 \leq R^2, \forall m \in \{1, \ldots, M\}, \ \forall i \in \{1, \ldots, N\}. \tag{3}$$

By introducing slack variables $\xi_{m,i}$, such that most of the training data from all the modalities in the new common space should lie inside the hypersphere, the above criterion becomes

$$F(R, \mathbf{a}) = R^2 + C \sum_{m=1}^{M}\sum_{i=1}^{N} \xi_{m,i}$$

s.t.

$$\|\mathbf{Q}_m \mathbf{x}_{m,i} - \mathbf{a}\|_2^2 \leq R^2 + \xi_{m,i}, \xi_{m,i} \geq 0, \quad \forall m \in \{1, \ldots, M\}, \forall i \in \{1, \ldots, N\}. \tag{4}$$

The Lagrange function corresponding to (4) can be given as

$$L = R^2 + C \sum_{m=1}^{M}\sum_{i=1}^{N} \xi_{m,i} - \sum_{m=1}^{M}\sum_{i=1}^{N} \gamma_{m,i}\xi_{m,i} - \sum_{m=1}^{M}\sum_{i=1}^{N} \alpha_{m,i}\Big(R^2 + \xi_{m,i} - \mathbf{x}_{m,i}^T \mathbf{Q}_m^T \mathbf{Q}_m \mathbf{x}_{m,i} + 2\mathbf{a}^T \mathbf{Q}_m \mathbf{x}_{m,i} - \mathbf{a}^T \mathbf{a}\Big) \tag{5}$$

The Lagrangian function should be maximized with respect to $\alpha_{m,i} \geq 0$, and $\gamma_{m,i} \geq 0$ and minimized with respect to R, $\mathbf{a}$, $\xi_{m,i}$, and $\mathbf{Q}_m$. By setting the partial derivative to zero, we get

$$\frac{\partial L}{\partial R} = 0 \ \Rightarrow \ \sum_{m=1}^{M}\sum_{i=1}^{N} \alpha_{m,i} = 1 \tag{6}$$

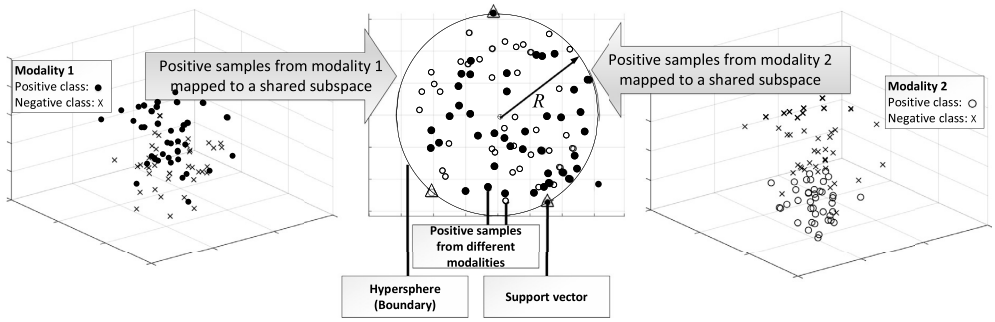

Fig. 1. Depiction of proposed MS-SVDD: Data from two modalities in their corresponding feature space are mapped to a common subspace, where positive class instances are enclosed inside a (hyper)sphere.

$$\frac{\partial L}{\partial \mathbf{a}} = 0 \;\Rightarrow\; \mathbf{a} = \sum_{m=1}^{M}\sum_{i=1}^{N} \alpha_{m,i}\mathbf{Q}_m\mathbf{x}_{m,i} \tag{7}$$

$$\frac{\partial L}{\partial \xi_{m,i}} = 0 \;\Rightarrow\; C - \alpha_{m,i} - \gamma_{m,i} = 0 \tag{8}$$

$$\frac{\partial L}{\partial \mathbf{Q}_m} = 0 \;\Rightarrow\; \mathbf{Q}_m = \Big(\mathbf{a}\sum_{i=1}^{N}\alpha_{m,i}\mathbf{x}_{m,i}^T\Big)\Big(\sum_{i=1}^{N}\alpha_{m,i}\mathbf{x}_{m,i}\mathbf{x}_{m,i}^T\Big)^{-1} \tag{9}$$

It is clear from (6)–(9) that parameters $\boldsymbol{\alpha}$ and $\mathbf{Q}$ are interrelated and cannot be jointly optimized. Hence we apply a two step iterative optimization process where, in each step, we fix one parameter and optimize the other. Substituting (2), (6), (7) and (8) in the Lagrangian function (5), we get

$$L = \sum_{m=1}^{M}\sum_{i=1}^{N}\alpha_{m,i}\mathbf{y}_{m,i}^T\mathbf{y}_{m,i} - \sum_{m=1}^{M}\sum_{i=1}^{N}\sum_{n=1}^{M}\sum_{j=1}^{N}\alpha_{m,i}\mathbf{y}_{m,i}^T\mathbf{y}_{n,j}\alpha_{n,j}. \tag{10}$$

We see that optimizing (10) for α corresponds to the traditional SVDD applied in the subspace. Maximizing (10) for a particular set of data will give us $\alpha_{m,i}$ corresponding each sample. The value of $\alpha_{m,i}$ for corresponding sample defines its position with respect to the hypersphere:

- Samples with $0 < \alpha_{m,i} < C$ define the data description and lie on the boundary of hypersphere, they are refered to as support vectors.
- Samples with $\alpha_{m,i} = C$ are outside the boundary.
- Samples with $\alpha_{m,i} = 0$ lie inside the boundary.

In the second step, we fix α and update $\mathbf{Q}_m$ for each modality. For this step, we add a regularization term ω:

$$L = \sum_{m=1}^{M}\sum_{i=1}^{N}\alpha_{m,i}\mathbf{x}_{m,i}^T\mathbf{Q}_m^T\mathbf{Q}_m\mathbf{x}_{m,i} - \sum_{m=1}^{M}\sum_{i=1}^{N}\sum_{n=1}^{M}\sum_{j=1}^{N}\alpha_{m,i}\mathbf{x}_{m,i}^T\mathbf{Q}_m^T\mathbf{Q}_n\mathbf{x}_{n,j}\alpha_{n,j} + \beta\omega. \tag{11}$$

The regularization term ω expresses the covariance of data from different modalities in the new low-dimensional space, and β is a regularization parameter for controlling the significance of ω. We propose different settings for ω as

$$\omega_0 = 0, \tag{12}$$

$$\omega_1 = \sum_{m=1}^{M}\mathrm{tr}\big(\mathbf{Q}_m\mathbf{X}_m\mathbf{X}_m^T\mathbf{Q}_m^T\big), \tag{13}$$

$$\omega_2 = \sum_{m=1}^{M}\mathrm{tr}\big(\mathbf{Q}_m\mathbf{X}_m\boldsymbol{\alpha}_m\boldsymbol{\alpha}_m^T\mathbf{X}_m^T\mathbf{Q}_m^T\big), \tag{14}$$

$$\omega_3 = \sum_{m=1}^{M}\mathrm{tr}(\mathbf{Q}_m\mathbf{X}_m\boldsymbol{\lambda}_m\boldsymbol{\lambda}_m^T\mathbf{X}_m^T\mathbf{Q}_m^T), \tag{15}$$

$$\omega_4 = \sum_{m=1}^{M}\sum_{n=1}^{M}\mathrm{tr}(\mathbf{Q}_m\mathbf{X}_m\mathbf{X}_n^T\mathbf{Q}_n^T), \tag{16}$$

$$\omega_5 = \sum_{m=1}^{M}\sum_{n=1}^{M}\mathrm{tr}(\mathbf{Q}_m\mathbf{X}_m\boldsymbol{\alpha}_m\boldsymbol{\alpha}_n^T\mathbf{X}_n^T\mathbf{Q}_n^T), \tag{17}$$

$$\omega_6 = \sum_{m=1}^{M}\sum_{n=1}^{M}\mathrm{tr}(\mathbf{Q}_m\mathbf{X}_m\boldsymbol{\lambda}_m\boldsymbol{\lambda}_n^T\mathbf{X}_n^T\mathbf{Q}_n^T), \tag{18}$$

where $\boldsymbol{\alpha}_m \in \mathbb{R}^N$ in (14) and (17) is a vector having the elements $\alpha_{m,1}, \ldots, \alpha_{m,N}$. Thus, $\boldsymbol{\alpha}_m$ has non-zero values for support vectors and outliers. $\boldsymbol{\lambda}_m \in \mathbb{R}^N$ in (15) and (18) is a vector having the elements of $\boldsymbol{\alpha}_m$ that are smaller than C. Values of $\boldsymbol{\alpha}_m$ corresponding to the outliers (i.e., $\alpha_{m,i} = C$) are replaced with zeros in $\boldsymbol{\lambda}_m$. Thus, $\boldsymbol{\lambda}_m$ has non-zero values only for the support vectors. For ω_0, the regularization term becomes obsolete and it is not used in the optimization process. In ω_1, the regularization term only uses representations coming from the respective modality and no representations from the other modalities are used to describe the variance of the positive class. In ω_2, all support vectors, i.e., representations at the hypersphere boundary, and outliers are used to describe the class variance for the update of the corresponding $\mathbf{Q}_m$. In ω_3, only support vectors of the respective modality are used to describe the variance of the class to be modelled. In ω_4, data from all the modalities are used to describe the covariance and regularize the update of $\mathbf{Q}_m$. In ω_5, the instances belonging to the hypersphere boundary and outliers from all modalities are used to describe the covariance. In ω_6, only the support vectors belonging to class boundary from all modalities are used to update $\mathbf{Q}_m$ and describe the covariance of the positive class.

Note that the MS-SVDD formulation reduces to S-SVDD [36] if data from only one modality ($M = 1$) are taken into account for data description. In S-SVDD, a single projection matrix $\mathbf{Q}$ is determined for mapping the data $\mathbf{X}$ from higher-dimensional space to a lower-dimensional space. A regularization term ψ, which expresses the class variance in the low-dimensional space, is added to the Lagrangian function of S-SVDD:

$$\psi = \mathrm{tr}(\mathbf{Q}\mathbf{X}\boldsymbol{\lambda}\boldsymbol{\lambda}^T\mathbf{X}^T\mathbf{Q}^T), \tag{19}$$

where $\boldsymbol{\lambda}$ can take different forms as described in [36]. The regularization terms, ω_0, ω_1, ω_2, and ω_3 for MS-SVDD become equivalent to the regularization terms proposed for S-SVDD when $M = 1$. Hence, MS-SVDD is a more generalized form of S-SVDD, which can form a data description by considering data from multiple modalities.

We update $\mathbf{Q}_m$ by using the gradient of L in (11) with respect to $\mathbf{Q}_m$,

$$\mathbf{Q}_m \leftarrow \mathbf{Q}_m - \eta \Delta L, \tag{20}$$

where η is the learning rate parameter and the gradient of L is calculated as

$$\frac{\partial L}{\partial \mathbf{Q}_m} = 2\sum_{i=1}^{N} \alpha_{m,i} \mathbf{Q}_m \mathbf{x}_{m,i} \mathbf{x}_{m,i}^T - 2\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{n=1}^{M} \mathbf{Q}_n \mathbf{x}_{n,j} \mathbf{x}_{m,i}^T \alpha_{m,i}\alpha_{n,j} + \beta \Delta\omega, \tag{21}$$

where $\Delta\omega$ is the derivative of the regularization term with respect to $\mathbf{Q}_m$

$$\Delta\omega_0 = 0, \tag{22}$$

$$\Delta\omega_1 = 2\mathbf{Q}_m \mathbf{X}_m \mathbf{X}_m^T, \tag{23}$$

$$\Delta\omega_2 = 2\mathbf{Q}_m \mathbf{X}_m \boldsymbol{\alpha}_m \boldsymbol{\alpha}_m^T \mathbf{X}_m^T, \tag{24}$$

$$\Delta\omega_3 = 2\mathbf{Q}_m \mathbf{X}_m \boldsymbol{\lambda}_m \boldsymbol{\lambda}_m^T \mathbf{X}_m^T, \tag{25}$$

$$\Delta\omega_4 = 2\sum_{n=1}^{M} (\mathbf{Q}_n \mathbf{X}_n \mathbf{X}_m^T), \tag{26}$$

$$\Delta\omega_5 = 2\sum_{n=1}^{M} (\mathbf{Q}_n \mathbf{X}_n \boldsymbol{\alpha}_n \boldsymbol{\alpha}_m^T \mathbf{X}_m^T), \tag{27}$$

$$\Delta\omega_6 = 2\sum_{n=1}^{M} (\mathbf{Q}_n \mathbf{X}_n \boldsymbol{\lambda}_n \boldsymbol{\lambda}_m^T \mathbf{X}_m^T). \tag{28}$$

We initialize the $\mathbf{Q}_m$ using PCA. At every iteration, the projection matrix is orthogonalized and normalized so that

$$\mathbf{Q}_m \mathbf{Q}_m^T = \mathbf{I}, \tag{29}$$

where $\mathbf{I}$ is an identity matrix. We use QR decomposition for orthogonalizing and normalizing the projection matrix $\mathbf{Q}_m$. Algorithm 1 describes the overall MS-SVDD algorithm.

Algorithm 1: MS-SVDD optimization.

Inputs : $\mathbf{Z}_m$ for each $m = 1, \ldots, M$, // Input data from all modalities
β, // Regularization parameter for controlling significance of ω
η, // Learning rate parameter
d, // Dimensionality of joint subspace
C, // Regularization parameter in SVDD
M // Total number of modalities

Outputs: $\mathbf{S}_m$ for each $m = 1, \ldots, M$, // Projection matrices for different modalities
R, // Radius of hypersphere
$\boldsymbol{\alpha}$ // Defines the data description

$\mathbf{Z}_m = \mathbf{X}_m$ for linear and NPT case ($\mathbf{K}_m$ for kernel case)
$\mathbf{S}_m = \mathbf{Q}_m$ for linear and NPT case ($\mathbf{W}_m$ for kernel case)

for *m=1:M* **do**
  Initialize $\mathbf{S}_m$ via linear-PCA (kernel-PCA);
end

for $iter = 1 : \max\_iter$ **do**
  For each m, map $\mathbf{Z}_m$ to $\mathbf{Y}_m$ using Eq. (2) (Eq. (31));
  Form Y by combining all $\mathbf{Y}_m$'s;
  Solve SVDD in the subspace to obtain $\boldsymbol{\alpha}$ in Eq. (10);
  for *m=1:M* **do**
    Calculate $\boldsymbol{\Delta L}$ using Eq. (21) (Eq. (31)) ;
    Update $\mathbf{S_m} \leftarrow \mathbf{S_m} - \boldsymbol{\eta \Delta L}$;
    Orthogonalize and normalize $\mathbf{S_m}$ using QR decomposition (eigendecomposition);
  end
end
For each m, compute $\mathbf{Y}_m$ using Eq. (2) (Eq. (31));
Form $\mathbf{Y}$ by combining all $\mathbf{Y}_m$'s;
Solve SVDD to obtain the final data description;

3.2. Non-linear MS-SVDD

For non-linear mapping from the original feature spaces to a new shared feature space, we use two approaches. The first approach is based on the standard kernel trick [20] and the second on the Nonlinear Projection Trick (NPT) [37], which is used as a computationally lighter alternative to the kernel trick.

3.2.1. Non-linear MS-SVDD with standard kernel trick

In the non-linear data description, the original data are mapped to a kernel space $\mathcal{F}$ using a non-linear function $\phi(\cdot)$ such that $\mathbf{x}_{m,i} \in \mathbb{R}^{D_m} \rightarrow \phi(\mathbf{x}_{m,i}) \in \mathcal{F}$. The kernel space dimensionality can possibly be infinite. Then the data are projected from the kernel space to $\mathbb{R}^d$ as

$$\mathbf{y}_{m,i} = \mathbf{Q}_m \phi(\mathbf{x}_{m,i}), \quad \forall i \in \{1, \ldots, N\}. \tag{30}$$

In order to calculate $\mathbf{y}_{m,i}$, we use the so-called kernel trick by expressing the projection matrix $\mathbf{Q}_m$ as a linear combination of the training data representations of the respective modality in the kernel space $\mathcal{F}$, leading to

$$\mathbf{y}_{m,i} = \mathbf{W}_m \boldsymbol{\Phi}_m^T \phi(\mathbf{x}_{m,i}) = \mathbf{W}_m \mathbf{k}_{m,i}, \quad \forall i \in \{1, \ldots, N\}, \tag{31}$$

where $\boldsymbol{\Phi}_m \in \mathbb{R}^{|\mathcal{F}| \times N}$ is a matrix formed in $\mathcal{F}$ containing the training data representations of modality m, $\mathbf{W}_m \in \mathbb{R}^{d \times N}$ is a matrix containing the weights for $\boldsymbol{\Phi}_m$ needed to form $\mathbf{Q}_m$, and $\mathbf{k}_{m,i}$ is the ith column of the Gramian matrix, also called as the kernel matrix, $\mathbf{K}_m \in \mathbb{R}^{N \times N}$, having elements equal to $\mathbf{K}_{m,ij} = \phi(\mathbf{x}_{m,i})^T \phi(\mathbf{x}_{m,j})$. In our experiments, we use the Radial Basis Function (RBF) kernel, given by

$$\mathbf{K}_{m,ij} = \exp\left(\frac{-\|\mathbf{x}_{m,i} - \mathbf{x}_{m,j}\|_2^2}{2\sigma^2}\right), \tag{32}$$

where $\sigma > 0$ is a hyperparameter and determines the width of the kernel.

The augmented version of the Lagrangian function now takes the following form:

$$L = \sum_{m=1}^{M}\sum_{i=1}^{N} \alpha_{m,i} \mathbf{k}_{m,i}^T \mathbf{W}_m^T \mathbf{W}_m \mathbf{k}_{m,i} - \sum_{m=1}^{M}\sum_{i=1}^{N}\sum_{n=1}^{M}\sum_{j=1}^{N} \alpha_{m,i} \mathbf{k}_{m,i}^T \mathbf{W}_m^T \mathbf{W}_n \mathbf{k}_{n,j} \alpha_{n,j} + \beta\omega. \tag{33}$$

The α's are calculated optimizing (10) with $\mathbf{W}_m$'s fixed, i.e., applying SVDD in the subspace. In the second step, the α's are fixed and $\mathbf{W}_m$'s are updated with the gradient descent:

$$\mathbf{W}_m \leftarrow \mathbf{W}_m - \eta \Delta L, \tag{34}$$

where the gradient is calculated as

$$\frac{\partial L}{\partial \mathbf{W}_m} = 2\sum_{i=1}^{N} \alpha_{m,i}\mathbf{W}_m\mathbf{k}_{m,i}\mathbf{k}_{m,i}^T - 2\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{n=1}^{M}\mathbf{W}_n\mathbf{k}_{n,j}\mathbf{k}_{m,i}^T\alpha_{m,i}\alpha_{n,j} + \beta\Delta\omega. \tag{35}$$

The gradient of the regularization term, $\Delta\omega$, now takes the following forms:

$$\Delta\omega_0 = 0, \tag{36}$$

$$\Delta\omega_1 = 2\mathbf{W}_m\mathbf{K}_m\mathbf{K}_m^T, \tag{37}$$

$$\Delta\omega_2 = 2\mathbf{W}_m\mathbf{K}_m\boldsymbol{\alpha}_m\boldsymbol{\alpha}_m^T\mathbf{K}_m^T, \tag{38}$$

$$\Delta\omega_3 = 2\mathbf{W}_m\mathbf{K}_m\boldsymbol{\lambda}_m\boldsymbol{\lambda}_m^T\mathbf{K}_m^T, \tag{39}$$

$$\Delta\omega_4 = 2\sum_{n=1}^{M}(\mathbf{W}_n\mathbf{K}_n\mathbf{K}_m^T), \tag{40}$$

$$\Delta\omega_5 = 2\sum_{n=1}^{M}(\mathbf{W}_n\mathbf{K}_n\boldsymbol{\alpha}_n\boldsymbol{\alpha}_m^T\mathbf{K}_m^T), \tag{41}$$

$$\Delta\omega_6 = 2\sum_{n=1}^{M}(\mathbf{W}_n\mathbf{K}_n\boldsymbol{\lambda}_n\boldsymbol{\lambda}_m^T\mathbf{K}_m^T). \tag{42}$$

We initialize the matrix $\mathbf{W}_m$ for each mode using kernel-PCA. We orthogonalize and normalize $\mathbf{W}_m$ at every iteration so that

$$\mathbf{W}_m\boldsymbol{\Phi}_m^T\boldsymbol{\Phi}_m\mathbf{W}_m^T = \mathbf{I}. \tag{43}$$

We decompose (43) using eigendecomposition as

$$\mathbf{W}_m\boldsymbol{\Phi}_m^T\boldsymbol{\Phi}_m\mathbf{W}_m^T = \mathbf{V}_m\boldsymbol{\Lambda}_m\mathbf{V}_m^T, \tag{44}$$

where $\boldsymbol{\Phi}_m^T\boldsymbol{\Phi}_m$ is $\mathbf{K}_m$, $\boldsymbol{\Lambda}_m$ is a diagonal matrix containing the eigenvalues of $\mathbf{W}_m\boldsymbol{\Phi}_m^T\boldsymbol{\Phi}_m\mathbf{W}_m^T$ and $\mathbf{V}_m$ contains the corresponding eigenvectors. After further simplification, the normalized projection matrix $\hat{\mathbf{W}}_m$ can be computed as

$$\hat{\mathbf{W}}_m = (\boldsymbol{\Lambda}_m^{\frac{1}{2}})^{+}\mathbf{V}_m^T\mathbf{W}_m, \tag{45}$$

where the + sign denotes pseudo-inverse. For notation simplicity, we set $\mathbf{W}_m = \hat{\mathbf{W}}_m$.

3.2.2. Non-linear MS-SVDD with nonlinear projection trick

The non-linear MS-SVDD using the kernel trick requires computing the eigendecomposition (44) at every iteration. This is computationally expensive and, therefore, we propose an alternative non-linear approach using NPT [37]. Here, a non-linear mapping is applied only at the beginning of the process, while the optimization follows the linear MS-SVDD. In the NPT-based MS-SVDD, we first compute kernel matrix $\mathbf{K}_m$ using (32). In the next step, the computed kernel matrix is centralized as

$$\hat{\mathbf{K}}_m = (\mathbf{I} - \mathbf{E}_N)\mathbf{K}_m(\mathbf{I} - \mathbf{E}_N) \tag{46}$$

where $\hat{\mathbf{K}}_m$ is the centralized kernel matrix and $\mathbf{E}_N$ is $N \times N$ matrix defined as

$$\mathbf{E}_N = \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T. \tag{47}$$

$\mathbf{1}_N \in \mathbb{R}^N$is a vector with each element having value of 1. The centralized matrix $\hat{\mathbf{K}}_m$ is decomposed by using eigendecomposition,

$$\hat{\mathbf{K}}_m = \mathbf{U}_m\mathbf{A}_m\mathbf{U}_m^T, \tag{48}$$

where $\mathbf{A}_m$ contains the non-negative eigenvalues of the centered kernel matrix and $\mathbf{U}_m$ contains the corresponding eigenvectors. The data in the reduced dimensional kernel space is obtained as

$$\boldsymbol{\Phi}_m = (\mathbf{A}_m^{\frac{1}{2}})^{+}\mathbf{U}_m^{+}\hat{\mathbf{K}}_m \tag{49}$$

Since we consider NPT as a pure preprocessing step, we continue by considering $\boldsymbol{\Phi}_m$ as our input data, i.e., we set $\mathbf{X}_m = \boldsymbol{\Phi}_m$. Then we follow the linear MS-SVDD. Note that in cases where the number of training samples is high, this pre-processing step can be highly accelerated by following approximations, like the Nyström-based Approximate Kernel Subspace Learning method in [38].

3.3. Test phase

During the test phase, an instance $\mathbf{x}_{m*} \in \mathbb{R}^{D_m}$ (the $*$ in subscript denotes test instance) coming from modality m is projected to the common d-dimensional subspace using (2) for the linear case. For kernel case, first, the kernel vector is computed as

$$\mathbf{k}_{m*} = \boldsymbol{\Phi}_m^T\phi(\mathbf{x}_{m*}) \tag{50}$$

and then projected to the common d-dimensional subspace using (31). For NPT, first the kernel vector $\mathbf{k}_{m^*}$ is computed and then centralized as

$$\hat{\mathbf{k}}_{m*} = (\mathbf{I} - \mathbf{E}_N)[\mathbf{k}_{m*} - \frac{1}{N}\mathbf{K}_m\mathbf{1}_N]. \tag{51}$$

The centralized kernel vector is mapped to

$$\phi_{m*} = (\boldsymbol{\Phi}_m^T)^{+}\hat{\mathbf{k}}_{m*} \tag{52}$$

and then to d-dimensional subspace using (2) (for notation simplicity $\boldsymbol{\phi}_{m^*}$ is considered as $\mathbf{x}_{m^*}$).

The decision to classify the test instance $\mathbf{y}_{m^*}$ as positive or negative is taken on the basis of its distance from the center of hypersphere, i.e.,

$$\|\mathbf{y}_{m*} - \mathbf{a}\|_2^2 = \mathbf{y}_{m*}^T\mathbf{y}_{m*} - 2\sum_{k=1}^{M}\sum_{i=1}^{N}\alpha_{k,i}\mathbf{y}_{m*}^T\mathbf{y}_{k,i} + \sum_{k=1}^{M}\sum_{i=1}^{N}\sum_{n=1}^{M}\sum_{j=1}^{N}\alpha_{k,i}\alpha_{n,j}\mathbf{y}_{k,i}^T\mathbf{y}_{n,j}. \tag{53}$$

The representation $\mathbf{y}_{m^*}$ is assigned to the positive class when $\|\mathbf{y}_{m*} - \mathbf{a}\|_2^2 \leq R^2$ and to the negative class if $\|\mathbf{y}_{m*} - \mathbf{a}\|_2^2 > R^2$, where R^2 is the distance from center $\mathbf{a}$ to any support vector on the boundary,

$$R^2 = \mathbf{v}^T\mathbf{v} - 2\sum_{m=1}^{M}\sum_{i=1}^{N}\alpha_{m,i}\mathbf{y}_{m,i}^T\mathbf{v} + \sum_{m=1}^{M}\sum_{i=1}^{N}\sum_{n=1}^{M}\sum_{j=1}^{N}\alpha_{m,i}\alpha_{n,j}\mathbf{y}_{m,i}^T\mathbf{y}_{n,j}, \tag{54}$$

where $\mathbf{v}$ is any support vector in the training set with corresponding α having value $0 < \alpha < C$. Since the items are represented by M different modalities, the final decision for assigning the item to a particular class (either positive or negative) can be taken using different strategies explained in Section 4.3.

3.4. Complexity analysis

The linear version of the proposed method has the following main steps: 1) Initializing the projection matrices via PCA, 2) mapping data from all modalities to a lower d-dimensional shared space, 3) SVDD for obtaining the α values and final data description for all data points coming from M different modalities, 4)

computing the gradient (ΔL) for each modality, 5) updating the projection matrices and 6) QR decomposition for orthogonalizing and normalizing the projection matrices. We analyze each of these steps and then compute the overall complexity of the algorithm:

1. PCA of a matrix is computed by the eigenvalue decomposition of its covariance matrix, so it involves two steps, i.e., computing the covariance matrix and then the eigenvalue decomposition of the obtained covariance matrix. The complexity of calculating covariance matrix and corresponding eigenvalue decomposition for a single modality is $\mathcal{O}\big(ND_m \times \min(N, D_m)\big)$ and $\mathcal{O}\big(D_m^3\big)$, respectively [39]. The complexity of computing PCA for all modalities is $\mathcal{O}\big(\min(N^2D_1, D_1^2N) + D_1^3) + (\min(N^2D_2, D_2^2N) + D_2^3) + \cdots + (\min(N^2D_M, D_M^2N) + D_M^3\big)$. We denote the sum of dimensions of all modalities as $\Sigma_{\mathcal{D}} = D_1 + D_2 + \cdots + D_M$ and similarly the sum of squared dimensions as $\Sigma_{\mathcal{D}^2} = D_1^2 + D_2^2 + \cdots + D_M^2$ (note that $\Sigma_{\mathcal{D}^2} \neq (\Sigma_{\mathcal{D}})^2$) and sum of cubed dimensions as $\Sigma_{\mathcal{D}^3} = D_1^3 + D_2^3 + \cdots + D_M^3$. Hence, the complexity of initializing the projection matrices via PCA becomes $\mathcal{O}\big(\min(N^2\Sigma_{\mathcal{D}}, \Sigma_{\mathcal{D}^2}N) + \Sigma_{\mathcal{D}^3}\big)$.
2. The complexity of mapping data from the original D_m dimensional space to a lower d-dimensional space is the complexity of multiplying $d \times D_m$ and $D_m \times N$, which has the complexity of $\mathcal{O}\big(dD_mN\big)$. Repeating this for all modalities we get $\mathcal{O}\big(d\Sigma_{\mathcal{D}}N\big)$
3. The complexity of SVDD for N data points is $\mathcal{O}\big(N^3\big)$ [40]. For all data points coming from M different modalities it becomes $\mathcal{O}\big(M^3N^3\big)$.
4. The gradient ΔL to update $\mathbf{Q}_m$ is computed using (21), where the second term has the highest complexity (equally high as regularization terms 4–6). Its complexity is $\mathcal{O}(2dN^2D_m\Sigma_{\mathcal{D}})$. As this step is repeated for all modalities the total complexity becomes $\mathcal{O}(2dN^2{\Sigma_{\mathcal{D}}}^2)$.
5. Updating the projection matrices has $\mathcal{O}\big(d\Sigma_{\mathcal{D}}\big)$ complexity.
6. The complexity of QR decomposition for a single modality is $\mathcal{O}(dD_m{}^2)$ [41]. Thus, the overall complexity of QR decompositions for all the modalities is $\mathcal{O}(d\Sigma_{\mathcal{D}^2})$.

Dropping the relatively lower intensive computational steps and adding the rest, the full complexity of the proposed method reduces to $\mathcal{O}\big(\min(N^2\Sigma_{\mathcal{D}}, \Sigma_{\mathcal{D}^2}N) + \Sigma_{\mathcal{D}^3} + M^3N^3\big)$. Assuming that the total number of samples M^*N is always greater than $\mathcal{D}$ and $M < < N$, the time complexity of (a single iteration of) our proposed algorithm in terms of the big $\mathcal{O}$ notation is $\mathcal{O}(N^3)$. In the testing phase, each representation of a test sample in each modality is projected to the d-dimensional subspace and then its distance is compared to R. This has the total complexity of $\mathcal{O}(d\Sigma_{\mathcal{D}} + Md)$.

For the non-linear version with NPT, the kernel matrix $\mathbf{K}_m$ is first formed which has the complexity of $\mathcal{O}(D_mN^2)$. Then the kernel matrix is centralized and decomposed by using eigendecomposition. Both of these steps have the complexity of $\mathcal{O}(N^3)$. As the data dimensionality in the remaining steps of the proposed method changes from D_m to N, the total complexity of the remaining steps becomes $\mathcal{O}\big(MN^3 + M^3N^3\big)$. Thus, the overall complexity in terms of the big $\mathcal{O}$ notation remains at $\mathcal{O}\big(N^3\big)$ for $M < < N$, while in practice the computational complexity is higher (by a scalar multiplier c) than for the linear version. Also for the non-linear version with the standard kernel trick, the overall complexity remains the same, but the kernel mapping is repeated at every iteration and, thus, the scalar c becomes larger for the overall training process. The testing complexity of the non-linear methods increases to $\mathcal{O}(N\Sigma_{\mathcal{D}} + dMN + Md)$.

4. Experiments

4.1. Datasets and prepossessing

To evaluate the proposed method, we performed different sets of experiments over 5 datasets. Robot Execution Failures dataset, Single Proton Emission Computed Tomography (SPECTF) heart dataset, and Ionosphere dataset were downloaded from UC Irvine (UCI) machine learning repository [34]. Caltech-7 dataset and Handwritten dataset were downloaded from a repository for multi-view learning [42]. The details of datasets and experiments are as follows.

The first set of experiments was performed on the Robot Execution Failures dataset [43]. In Robot Execution Failures dataset, force and torque measurements are collected at regular intervals of time after a task failure is detected. The dataset is divided into five different learning problems (LP) corresponding to different triggering events:

- **LP1:** Failures in approach to grasp position
- **LP2:** Failures in the transfer of a part
- **LP3:** Position of the part after a transfer failure
- **LP4:** Failures in approach to ungrasp position
- **LP5:** Failures in motion with part

The total number of instances and the distribution of the classes are given in Table 1. All instances are given as 15 samples collected at 315 ms regular time intervals for each sensor. For this dataset, we consider all the instances belonging to the normal class as the target class and the remaining classes as the non-target data. Hence, we have two modalities (torque and force measurements), and we consider the dataset as a one-class classification problem.

The second set of experiments was performed SPECTF heart dataset [44]. The SPECTF heart dataset consists of two sets of features corresponding to rest and stress condition SPECTF images of different subjects. The training set consists of 40 examples diagnosed as healthy heart muscle perfusions and 40 diagnosed as pathological perfusions. The test set consists of 15 instances of healthy heart muscle perfusions and 172 from instances diagnosed as pathological perfusions. We convert this to a multimodal one-class classification problem by considering the rest and stress conditions as different modalities and by selecting the healthy heart muscle perfusions as our target class.

The third set of experiments was performed over the Caltech-7 dataset. We used Gabor feature and Wavelet moments as our two different modalities. The dataset contains 1474 total samples from 7 different classes. We selected faces (435 samples) as our target class and the rest of the classes all together (1039 samples) as the outlier class.

We used Ionosphere dataset for the fourth set of experiments. The categories in this dataset are described by two attributes per pulse number resulting from the complex electromagnetic signal, processed by an autocorrelation function. We used the two attributes (real and complex) for each pulse as two different modalities and the attribute "good" as our target class. The total number of samples in this dataset is 351, out of which 225 are from the target class (good), and the rest of 126 samples are from outlier class (bad).

For the fifth set of experiments, we used Handwritten dataset. We considered the samples of numeral 0 as the target. In the Handwritten dataset, the total number of samples is 2000, out of which 200 are from the target class. The rest of the 1800 samples are considered as an outlier class. We used the Zernike moment (ZER) and morphological (MOR) features as our two different modalities.

Table 1
Robot execution failures dataset.

Learning problem	Instances	Classes and distribution
LP1	88	24% normal 19% collision 18% front collision 39% obstruction
LP2	47	43% normal 13% front collision 15% back collision 11% collision to the right 19% collision to the left
LP3	47	43% ok 19% slightly moved 32% moved 06% lost
LP4	117	21% normal 62% collision 18% obstruction
LP5	164	27% normal 16% bottom collision 13% bottom obstruction 29% collision in part 16% collision in tool

4.2. Experimental setup

For the Robot Execution Failures dataset, Ionosphere dataset, Caltech-7 dataset, and Handwritten dataset, we performed our experiments on 70-30% split for training and testing sets. We selected the 70-30% split randomly 5 times, keeping the distribution of classes similar to the original data. To tune the hyperparameters for final testing, we did 5-fold cross-validation on the training set, where the (70%) training data are divided into 5 different sets, and each time one set is used for validation while all the others for training. The process was repeated 5 times until all the sets have been used as validation sets. For SPECTF heart dataset, the train and test sets are given with the dataset. We did 5-fold cross-validation on the training set to optimize the hyperparameters.

For all datasets, the models were trained by using samples from the positive class only, while testing was carried out using all the classes. The hyperparameters were selected from the following ranges:

- $\beta \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}, 10^{4}\}$,
- $C \in \{0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$,
- $\sigma \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$,
- $d \in \{1, 2, 3, 4, 5, 10, 20, 50, 100\}$,
- $\eta = 0.1$.

Here, we restricted the dimension d of the shared subspace as $d < \min\{D_1, \ldots, D_M\}$ for a given dataset, where D_m is the dimensionality of modality m. For competing methods, the features from different modalities were concatenated before training the model. We also report the results of the competing methods by considering data from one modality at a time for training and testing. For competing methods, the hyperparameters were selected from the same ranges as mentioned above.

4.3. Decision strategies

During testing, after the common compact representation of all modalities was formed, each representation (modality) of an instance was mapped to the lower-dimensional subspace via corresponding projection matrix and classified as described in Section 3.3. The following four strategies were used to decide the final class for the instance:

- **Decision strategy 1** (also called the AND gate)**:** The test instance is assigned the target label if the representations from all modalities for that particular instance are classified to the target class and the non-target label otherwise.
- **Decision strategy 2** (also called as the OR gate)**:** The final decision is taken on the basis of the OR gate principle, i.e., if a representation of an instance from any of the modalities is classified to the target class, the overall decision for that particular instance is taken in favor of the target class.
- **Decision strategy 3:** The final classification decision is made on the basis of first modality, i.e., if the representation from the first modality is assigned to a particular class, the overall classification is made following that.
- **Decision strategy 4:** The overall decision is taken on the basis of the label assigned to the representation from the second modality.

It should be noted that for more than two modalities, different decision strategies, such as majority vote, might be more suitable.

4.4. Evaluation criteria

One-class classification models can be evaluated using different metrics. These metrics are decided on the basis of the goals of a given application. For example, in outlier detection, the focus is on detecting negative instances accurately. The most common metrics in one-class classification are true positive rate (*tpr*), and true negative rate (*tnr*). The former, also called as recall, sensitivity, or hit rate, is the proportion of positive instances that is classified by the trained model as positive correctly:

$$tpr = \frac{tp}{p}, \tag{55}$$

where tp is the number of positive samples classified correctly and p is the total number of positive samples in the test set. The latter, *tnr*, also called as specificity, is defined as

$$tnr = \frac{tn}{n}, \tag{56}$$

where tn is the number of negative samples classified correctly and n is the total number of negative samples in the test set. Accuracy (*accu*) is measured as the ratio of the number of correctly classified instances to the total number of instances:

$$accu = \frac{tp + tn}{p + n}. \tag{57}$$

Precision (*pre*) measures the proportion of instances classified positive which really are positive:

$$pre = \frac{tp}{tp + fp}, \tag{58}$$

where fp is the number of false positives. Another useful measure is *F1* measure, which is the harmonic mean of *pre* and *tpr*:

$$F1 = 2 \times \frac{pre \times tpr}{pre + tpr}. \tag{59}$$

Geometric mean (*gm*) is defined as the square root of the product of sensitivity and specificity:

$$gm = \sqrt{tpr \times tnr}. \tag{60}$$

gm has been used by many researchers for imbalanced datasets. Since it takes into consideration both sensitivity and specificity, we opted to finetune hyperparameters based on the *gm* score on the validation data.

4.5. Experimental results and discussion

InTables 2–5, we report the average of different evaluation metrics over the five data splits for Robot Execution Failures dataset, Caltech-7 dataset, Ionosphere dataset, and Handwritten dataset, respectively, for both linear and non-linear versions of the applied

Table 2
Test results for robot execution failures dataset.

	Linear						Non-linear					
	accu	tpr	tnr	pre	F1	gm	accu	tpr	tnr	pre	F1	gm
Proposed method												
MS-SVDD$\omega_2 ds3$	0.97	0.97	0.97	0.93	0.95	**0.97**	0.94	0.98	0.92	0.83	0.90	0.95
MS-SVDD$\omega_5 ds3$	0.97	0.95	0.97	0.93	0.94	0.96	0.94	0.98	0.92	0.83	0.90	0.95
Concatenated features												
S-SVDDψ_1	0.66	0.89	0.57	0.46	0.60	0.71	0.94	0.84	0.98	0.95	0.89	0.91
S-SVDDψ_2	0.70	0.80	0.66	0.58	0.60	0.70	0.92	0.90	0.93	0.84	0.87	0.91
S-SVDDψ_3	0.66	0.78	0.61	0.46	0.56	0.67	0.93	0.93	0.93	0.85	0.89	0.93
S-SVDDψ_4	0.64	0.94	0.52	0.44	0.60	0.70	0.96	0.90	0.98	0.96	0.93	0.94
OC-SVM	0.51	0.47	0.52	0.28	0.35	0.49	0.86	0.49	1.00	1.00	0.65	0.70
SVDD	0.97	0.91	0.99	0.98	0.95	0.95	0.95	0.85	0.99	0.98	0.91	0.92
Force measurements												
S-SVDDψ_1	0.76	0.88	0.71	0.55	0.67	0.79	0.96	0.90	0.98	0.95	0.92	0.94
S-SVDDψ_2	0.77	0.94	0.71	0.56	0.70	0.82	0.96	0.90	0.98	0.95	0.92	0.94
S-SVDDψ_3	0.73	0.70	0.74	0.51	0.58	0.71	0.96	0.91	0.98	0.95	0.93	0.94
S-SVDDψ_4	0.76	0.85	0.72	0.54	0.66	0.78	0.93	0.82	0.98	0.95	0.84	0.88
OC-SVM	0.50	0.53	0.49	0.29	0.37	0.51	0.86	0.49	1.00	1.00	0.65	0.70
SVDD	0.97	0.90	0.99	0.98	0.94	0.95	0.97	0.92	0.99	0.98	0.95	**0.96**
Torque measurements												
S-SVDDψ_1	0.59	0.96	0.44	0.41	0.57	0.65	0.97	0.89	1.00	1.00	0.94	0.94
S-SVDDψ_2	0.61	0.94	0.48	0.42	0.57	0.67	0.71	0.66	0.73	0.51	0.54	0.51
S-SVDDψ_3	0.62	0.92	0.50	0.43	0.58	0.67	0.92	0.76	0.99	0.97	0.82	0.85
S-SVDDψ_4	0.61	0.96	0.48	0.42	0.58	0.68	0.76	0.76	0.76	0.76	0.71	0.66
OC-SVM	0.52	0.59	0.49	0.31	0.40	0.53	0.84	0.58	0.94	0.81	0.66	0.73
SVDD	0.90	0.95	0.88	0.76	0.84	0.91	0.91	0.88	0.92	0.81	0.84	0.90

Table 3
Test results for Caltech-7 dataset.

	Linear						Non-linear					
	accu	tpr	tnr	pre	F1	gm	accu	tpr	tnr	pre	F1	gm
Proposed method												
MS-SVDD$\omega_1 ds1$	0.91	0.96	0.89	0.78	0.86	0.92	0.94	0.98	0.92	0.85	0.91	**0.95**
MS-SVDD$\omega_4 ds1$	0.91	0.95	0.89	0.78	0.86	0.92	0.94	0.95	0.94	0.88	0.91	**0.95**
Concatenated features												
S-SVDDψ_1	0.65	0.96	0.52	0.46	0.62	0.71	0.37	0.35	0.38	0.15	0.20	0.23
S-SVDDψ_2	0.67	0.92	0.57	0.48	0.63	0.72	0.66	0.69	0.64	0.39	0.48	0.53
S-SVDDψ_3	0.71	0.84	0.66	0.59	0.65	0.69	0.90	0.79	0.94	0.86	0.81	0.86
S-SVDDψ_4	0.62	0.96	0.47	0.46	0.61	0.66	0.87	0.61	0.97	0.91	0.72	0.76
OC-SVM	0.22	0.47	0.12	0.18	0.26	0.22	0.86	0.53	1.00	0.99	0.69	0.73
SVDD	0.92	0.94	0.91	0.81	0.87	**0.93**	0.96	0.94	0.97	0.93	0.94	**0.95**
Gabor feature												
S-SVDDψ_1	0.68	0.72	0.67	0.47	0.57	0.69	0.46	0.84	0.31	0.33	0.48	0.50
S-SVDDψ_2	0.68	0.72	0.67	0.47	0.57	0.69	0.54	0.78	0.44	0.46	0.52	0.50
S-SVDDψ_3	0.61	0.74	0.55	0.45	0.52	0.58	0.76	0.68	0.80	0.65	0.63	0.71
S-SVDDψ_4	0.70	0.74	0.68	0.49	0.59	0.71	0.78	0.39	0.94	0.80	0.46	0.55
OC-SVM	0.43	0.53	0.40	0.27	0.36	0.45	0.79	0.55	0.89	0.69	0.61	0.70
SVDD	0.76	0.70	0.78	0.57	0.63	0.74	0.74	0.92	0.67	0.55	0.68	0.78
Wavelet moments												
S-SVDDψ_1	0.70	0.73	0.68	0.50	0.59	0.69	0.54	0.44	0.58	0.22	0.26	0.24
S-SVDDψ_2	0.71	0.73	0.70	0.52	0.60	0.70	0.51	0.93	0.33	0.41	0.55	0.42
S-SVDDψ_3	0.50	0.93	0.33	0.38	0.54	0.50	0.79	0.38	0.96	0.65	0.44	0.51
S-SVDDψ_4	0.56	0.88	0.42	0.40	0.54	0.59	0.61	0.51	0.65	0.50	0.36	0.30
OC-SVM	0.21	0.48	0.10	0.18	0.26	0.21	0.84	0.48	0.99	0.97	0.64	0.69
SVDD	0.91	0.94	0.89	0.79	0.85	0.91	0.94	0.97	0.93	0.85	0.91	**0.95**

methods. In Table 6, we report the results on the test set for the SPECTF heart dataset. In these tables, we only show the best performing versions of the proposed method, along with all competing methods. We compare our results with OC-SVM [26], SVDD [25], and S-SVDD [36]. In S-SVDD, different regularization terms (ψ's) were proposed and, hence, we compare MS-SVDD with all proposed regularization terms of S-SVDD. We use kernel version of the competing methods for non-linear comparisons. In these tables, we report the best performing non-linear version of MS-SVDD for corresponding datasets. To analyze the different regularization terms and decision strategies for the proposed method, we also report the exhaustive results obtained by different settings in the supplementary material in Tables 1–5. The best results in terms of *gm* are reported as in bold formatting.

For the Robot Execution Failures dataset (Table 2), our proposed method outperforms all the competing methods in the linear case. The results achieved by the linear version of the proposed MS-SVDD method are overall best also compared to the non-linear methods. Table 2 shows that using decision strategy 3 with constraint ω_2 (all support vectors and outliers from the corresponding modality considered for the update of the corresponding $\mathbf{Q}_m$) yields the best overall results for the robot dataset. In the

Table 4
Test results for Ionosphere dataset.

	Linear						Non-linear					
	accu	tpr	tnr	pre	F1	gm	accu	tpr	tnr	pre	F1	gm
Proposed method												
MS-SVDD$\omega_2 ds4$	0.87	0.95	0.73	0.87	0.91	0.83	0.76	0.86	0.59	0.79	0.82	0.71
MS-SVDD$\omega_1 ds4$	0.83	0.91	0.69	0.84	0.87	0.79	0.88	0.95	0.74	0.87	0.91	0.84
Concatenated features												
S-SVDDψ_1	0.69	0.88	0.32	0.69	0.77	0.53	0.49	0.37	0.69	0.55	0.39	0.28
S-SVDDψ_2	0.69	0.89	0.31	0.69	0.77	0.51	0.74	0.60	0.98	0.98	0.75	0.77
S-SVDDψ_3	0.58	0.63	0.48	0.66	0.62	0.51	0.72	0.77	0.62	0.83	0.77	0.63
S-SVDDψ_4	0.72	0.98	0.23	0.70	0.82	0.43	0.66	0.61	0.77	0.88	0.67	0.62
OC-SVM	0.38	0.39	0.34	0.52	0.45	0.37	0.66	0.48	0.97	0.97	0.63	0.67
SVDD	0.87	0.93	0.76	0.88	0.90	**0.84**	0.89	0.94	0.78	0.89	0.92	0.86
Real												
S-SVDDψ_1	0.81	0.99	0.50	0.78	0.87	0.69	0.54	0.36	0.86	0.67	0.43	0.46
S-SVDDψ_2	0.80	0.99	0.47	0.78	0.87	0.67	0.62	0.49	0.86	0.87	0.61	0.64
S-SVDDψ_3	0.81	0.99	0.49	0.78	0.87	0.68	0.68	0.63	0.78	0.86	0.70	0.68
S-SVDDψ_4	0.81	0.99	0.50	0.78	0.87	0.70	0.58	0.45	0.83	0.85	0.53	0.56
OC-SVM	0.49	0.52	0.42	0.61	0.56	0.46	0.68	0.56	0.89	0.93	0.67	0.69
SVDD	0.88	0.95	0.74	0.87	0.91	**0.84**	0.89	0.94	0.81	0.90	0.92	**0.87**
Complex												
S-SVDDψ_1	0.50	0.37	0.72	0.70	0.49	0.51	0.43	0.27	0.71	0.52	0.30	0.34
S-SVDDψ_2	0.47	0.35	0.69	0.67	0.46	0.49	0.66	0.56	0.83	0.85	0.68	0.68
S-SVDDψ_3	0.53	0.57	0.46	0.67	0.58	0.39	0.65	0.65	0.65	0.78	0.70	0.63
S-SVDDψ_4	0.50	0.38	0.72	0.70	0.49	0.52	0.63	0.64	0.62	0.76	0.69	0.62
OC-SVM	0.40	0.31	0.57	0.56	0.40	0.42	0.66	0.59	0.78	0.84	0.69	0.67
SVDD	0.77	0.89	0.55	0.79	0.83	0.70	0.79	0.91	0.58	0.80	0.85	0.72

Table 5
Test results for Handwritten dataset.

	Linear						Non-linear					
	accu	tpr	tnr	pre	F1	gm	accu	tpr	tnr	pre	F1	gm
Proposed method												
MS-SVDD$\omega_4 ds4$	0.98	0.99	0.98	0.90	0.93	**0.98**	0.99	0.99	1.00	0.98	0.98	**0.99**
MS-SVDD$\omega_4 ds1$	0.98	0.90	0.99	0.89	0.89	0.94	0.98	0.95	0.99	0.91	0.93	0.97
Concatenated features												
S-SVDDψ_1	0.78	0.92	0.76	0.34	0.49	0.83	0.53	0.40	0.54	0.05	0.09	0.14
S-SVDDψ_2	0.82	0.88	0.81	0.40	0.54	0.84	0.62	0.66	0.61	0.18	0.25	0.44
S-SVDDψ_3	0.82	0.97	0.81	0.39	0.55	0.88	0.63	0.58	0.64	0.20	0.25	0.30
S-SVDDψ_4	0.84	0.92	0.83	0.42	0.56	0.87	0.71	0.39	0.75	0.08	0.13	0.17
OC-SVM	0.50	0.51	0.50	0.12	0.19	0.49	0.95	0.51	1.00	1.00	0.68	0.71
SVDD	0.95	0.93	0.95	0.69	0.79	0.94	0.95	0.92	0.96	0.74	0.81	0.94
ZER												
S-SVDDψ_1	0.55	0.92	0.51	0.18	0.30	0.68	0.59	0.41	0.61	0.06	0.10	0.24
S-SVDDψ_2	0.52	0.88	0.48	0.17	0.28	0.64	0.62	0.78	0.60	0.17	0.27	0.48
S-SVDDψ_3	0.50	0.96	0.45	0.19	0.31	0.63	0.57	0.61	0.57	0.31	0.20	0.37
S-SVDDψ_4	0.64	0.90	0.61	0.21	0.34	0.74	0.55	0.60	0.54	0.09	0.15	0.24
OC-SVM	0.43	0.42	0.43	0.09	0.14	0.41	0.95	0.52	1.00	0.93	0.67	0.72
SVDD	0.88	0.90	0.88	0.47	0.61	0.89	0.92	0.88	0.92	0.56	0.68	0.90
MOR												
S-SVDDψ_1	0.84	0.99	0.82	0.48	0.61	0.90	0.84	0.01	0.93	0.00	0.00	0.03
S-SVDDψ_2	0.92	0.99	0.91	0.66	0.76	0.95	0.58	0.44	0.60	0.43	0.22	0.20
S-SVDDψ_3	0.86	0.99	0.84	0.52	0.64	0.91	0.61	0.70	0.60	0.44	0.42	0.36
S-SVDDψ_4	0.84	0.99	0.82	0.48	0.61	0.90	0.25	0.67	0.20	0.27	0.14	0.04
OC-SVM	0.54	0.45	0.55	0.13	0.18	0.39	0.99	0.87	1.00	1.00	0.93	0.93
SVDD	0.93	0.91	0.93	0.75	0.78	0.92	0.99	0.96	1.00	1.00	0.98	0.98

non-linear case, the best performance for the proposed method is achieved by using the kernel trick with either constraint type ω_2 or ω_5, both with decision strategy 3.

We also notice that the first modality (force measurements) is vital in taking the final decision as in both linear and non-linear cases, the best results are obtained when the decision is taken based on the first modality (decision strategy 3). The importance of the first modality is also evident from the results of the competing methods as the best results are obtained when using force measurements only. The results on the concatenated features are slightly worse, and the results using the torque measurements are clearly worse. Nevertheless, the proposed multimodal approach has managed to boost the results by combining information from both modalities.

For the Caltech-7 dataset, in the linear case, MS-SVDD performs better than all other methods with a single modality. Overall, only SVDD using concatenated features outperforms MS-SVDD and the margin is small. In the non-linear case, MS-SVDD obtains the best results along with SVDD. In terms of *tpr*, MS-SVDD outperforms all the other methods in the non-linear case while maintaining rea-

Table 6
Test results for SPECTF heart dataset.

	Linear						Non-linear					
	accu	tpr	tnr	pre	F1	gm	accu	tpr	tnr	pre	F1	gm
Proposed method												
MS-SVDD$\omega_0 ds1$	0.78	0.80	0.78	0.24	0.37	**0.79**	0.55	0.60	0.55	0.10	0.18	0.57
MS-SVDD$\omega_2 ds1$	0.78	0.80	0.77	0.24	0.36	**0.79**	0.80	0.73	0.80	0.24	0.37	**0.77**
Concatenated features												
S-SVDDψ_1	0.71	0.53	0.73	0.15	0.23	0.62	0.77	0.60	0.78	0.20	0.30	0.69
S-SVDDψ_2	0.69	0.87	0.67	0.19	0.31	0.76	0.77	0.60	0.78	0.20	0.30	0.69
S-SVDDψ_3	0.66	0.93	0.64	0.18	0.31	0.77	0.77	0.60	0.78	0.20	0.30	0.69
S-SVDDψ_4	0.56	0.67	0.55	0.11	0.19	0.60	0.77	0.60	0.78	0.20	0.30	0.69
OC-SVM	0.86	0.27	0.91	0.20	0.23	0.49	0.76	0.73	0.77	0.22	0.33	0.75
SVDD	0.69	0.73	0.69	0.17	0.28	0.71	0.75	0.67	0.76	0.19	0.30	0.71
Rest Mode												
S-SVDDψ_1	0.50	0.73	0.48	0.11	0.19	0.59	0.46	0.87	0.42	0.12	0.20	0.61
S-SVDDψ_2	0.58	0.87	0.55	0.14	0.25	0.69	0.77	0.53	0.79	0.18	0.27	0.65
S-SVDDψ_3	0.40	0.80	0.37	0.10	0.18	0.54	0.79	0.47	0.81	0.18	0.26	0.62
S-SVDDψ_4	0.38	0.87	0.34	0.10	0.18	0.54	0.60	0.87	0.58	0.15	0.26	0.71
OC-SVM	0.76	0.60	0.77	0.19	0.29	0.68	0.61	0.80	0.60	0.15	0.25	0.69
SVDD	0.59	0.73	0.58	0.13	0.22	0.65	0.59	0.73	0.58	0.13	0.22	0.65
Stress Mode												
S-SVDDψ_1	0.53	0.47	0.53	0.08	0.14	0.50	0.68	0.73	0.67	0.16	0.27	0.70
S-SVDDψ_2	0.65	0.80	0.63	0.16	0.27	0.71	0.75	0.53	0.77	0.17	0.26	0.64
S-SVDDψ_3	0.73	0.67	0.73	0.18	0.28	0.70	0.70	0.73	0.70	0.17	0.28	0.72
S-SVDDψ_4	0.55	0.93	0.52	0.14	0.25	0.69	0.75	0.53	0.77	0.17	0.26	0.64
OC-SVM	0.86	0.20	0.91	0.17	0.18	0.43	0.73	0.60	0.74	0.17	0.26	0.67
SVDD	0.76	0.60	0.77	0.19	0.29	0.68	0.78	0.53	0.80	0.19	0.28	0.65

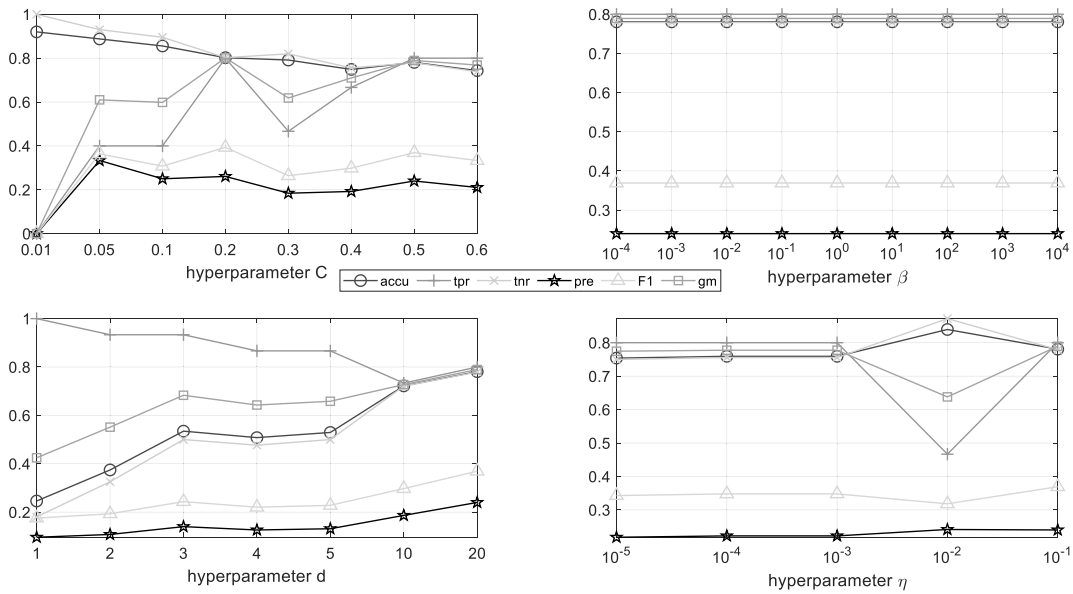

Fig. 2. Hyperparameters sensitivity analysis for $\omega_0 ds1$.

sonably good *tnr*. We also notice that both modalities are vital in taking the final decision as the best performance of MS-SVDD is obtained by decision strategy 1 (AND gate).

For Ionosphere dataset, only SVDD applied on concatenated features or the first modality outperforms MS-SVDD in terms of *gm*. Nevertheless, the performance of MS-SVDD is competitive as shown also by the top results obtained by the other performance metrics such as *F1* measure. In case of MS-SVDD, the second modality (Complex) is found to be more vital for taking the final decision.

For the Handwritten dataset, MS-SVDD outperforms all competing methods in both linear and non-linear cases. It is noticed that decision strategy 4 yields the best results in both linear and non-linear cases for MS-SVDD, i.e., MOR features are more vital than ZER features.

For SPECTF heart dataset, in both linear and non-linear cases, the best results are achieved by MS-SVDD. We note that ω_0 (no constraint used) and ω_2, where all support vectors and outliers are used to describe the class variance for the update of the corresponding $\mathbf{Q}_m$ in decision strategy 1 yield the best overall results.

We compare the results for different variant of MS-SVDD in Tables 1–5 of the supplementary material. Overall in all datasets, NPT is found to be more robust than the kernel version. Linear MS-SVDD is found to perform best over 2 datasets, similar to the NPT version, which performs best on two datasets as well. The kernel MS-SVDD performs best on one out of five datasets as compared to linear and NPT version of MS-SVDD. All the relevant codes (implementation) for the proposed method are available online at [45].

We also carried out a sensitivity analysis of different hyperparameters for linear MS-SVDD over SPECTF heart dataset. To analyze the sensitivity of MS-SVDD for each hyperparameter, we fix the other hyperparameters to their optimal values and record the performance with all the considered hyperparameter values. Fig. 2 shows as an example the results for decision strategy 1 without any constraint. For the other decision strategies and constraints, we show the results in Figures 1–27 in the supplementary material. We note a trend of increase in *tpr* and decrease in *tnr* with the increase of value for hyperparameter C. We also noticed that the performance of trained models are relatively less sensitive to the hyperparameter β as compared to other hyperparameters. For hyperparameter d, initially, there is a noticeable rise in the performance of the trained model; however, after certain value, the change seems to be very small. For hyperparameter η, we notice that precision and *F1* measure stay stable with changing its value.

We also report the numerical training and testing time (in milliseconds) in the supplementary material (Tables 1–10) for all methods over all datasets used in the experiments. In the majority of cases, the proposed method has a higher computational cost than the competing methods, but generally, the difference is in the fractions of a second, which is negligible for datasets used in this work. It is also evident from the numerical results that the time complexity of the proposed method is higher mainly in the training phase, while in the testing phase the difference is negligible. This is as expected based on the complexity analysis in Section 3.4.

5. Conclusion

In this paper, a new multimodal one-class classification method is proposed. The proposed method iteratively transforms data from all the modalities to a new shared subspace optimized for data description in multimodal one-class classification tasks. We derived linear and two different non-linear versions along with a selection of different regularization terms. According to the best of our knowledge, this is the first work in the field of subspace learning for multimodal one-class classification. We conducted experiments comparing the different versions of MS-SVDD and performed comparisons against other one-class classification methods using either concatenated representations or a single modality at a time.

In most cases, linear and NPT version of MS-SVDD outperformed all the competing methods in our experiments. NPT turned out to be more stable than the kernel version. We noticed that the optimal decision strategy depends on the usefulness of different modalities. If a particular modality is more informative than other(s), then it is useful to use that particular modality for making the final decision. Nevertheless, MS-SVDD can improve the results as compared to using a single modality only. If the modalities are more balanced, the AND gate strategy seems to perform better.

MS-SVDD can be interpreted and used in many ways for different one-class multimodal problems. It can be used for anomaly detection and detection of a specific class such as speaker verification and face recognition. In the future, we intend to try different kernels and model-based decision strategies for the proposed method. We also intend to propose changes in the boundary shape (other than spherical) for enclosing the target data in subspace. There is also room for research in other one-class classification techniques for multimodal subspace learning.

Declaration of Competing Interest

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

Acknowledgement

This work was supported by the NSF-Business Finland Center for Visual and Decision Informatics project Co-Botics, jointly sponsored by Tieto Oy Finland and CA Software.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.patcog.2020.107648.

References

[1] Y. Qu, G. Zhang, Z. Zou, Z. Liu, J. Mao, Active multimodal sensor system for target recognition and tracking, Sensors 17 (7) (2017) 1518.
[2] Y. Zhang, B. Song, X. Du, M. Guizani, Vehicle tracking using surveillance with multimodal data fusion, IEEE Trans. Intell. Transp. Syst. 19 (99) (2018) 1–9.
[3] S. Kye, J. Moon, J. Lee, I. Choi, D. Cheon, K. Lee, Multimodal data collection framework for mental stress monitoring, in: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, ACM, 2017, pp. 822–829.
[4] Z. Gu, B. Lang, T. Yue, L. Huang, Learning joint multimodal representation based on multi-fusion deep neural networks, in: International Conference on Neural Information Processing, Springer, 2017, pp. 276–285.
[5] A. Iosifidis, M. Gabbouj, Multi-class support vector machine classifiers using intrinsic and penalty graphs, Pattern Recognit. 55 (2016) 231–246.
[6] S.S. Khan, M.G. Madden, One-class classification: taxonomy of study and review of techniques, Knowl. Eng. Rev. 29 (3) (2014) 345–374.
[7] C. Xu, D. Tao, C. Xu, A survey on multi-view learning, arXiv:1304.5634 (2013).
[8] G. Cao, A. Iosifidis, K. Chen, M. Gabbouj, Generalized multi-view embedding for visual recognition and cross-modal retrieval, IEEE Trans. Cybern. 48 (9) (2018) 2542–2555.
[9] P. Khante, Learning attributes of real-world objects by clustering multimodal sensory data, The University of Texas, 2017 Ph.D. thesis.
[10] T. Baltrušaitis, C. Ahuja, L.-P. Morency, Multimodal machine learning: a survey and taxonomy, IEEE Trans. Pattern Anal. Mach. Intell. 41 (2) (2018) 423–443.
[11] M. Heckmann, Audio-visual word prominence detection from clean and noisy speech, Comput. Speech Lang. 48 (2018) 15–30.
[12] G. Cao, A. Iosifidis, M. Gabbouj, Multi-view nonparametric discriminant analysis for image retrieval and recognition, IEEE Signal Process. Lett. 24 (10) (2017) 1537–1541.
[13] L.-l. Chen, Y. Zhao, P.-f. Ye, J. Zhang, J.-z. Zou, Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers, Expert Syst. Appl. 85 (2017) 279–291.
[14] S. Venugopalan, L.A. Hendricks, M. Rohrbach, R.J. Mooney, T. Darrell, K. Saenko, Captioning images with diverse objects., in: CVPR, vol. 3, 2017, p. 8.
[15] M. Maimaitijiang, V. Sagan, P. Sidike, S. Hartling, F. Esposito, F.B. Fritschi, Soybean yield prediction from UAV using multimodal data fusion and deep learning, Remote Sens. Environ. 237 (2020) 111599.
[16] T. Baltrušaitis, C. Ahuja, L.-P. Morency, Multimodal machine learning: a survey and taxonomy, IEEE Trans. Pattern Anal. Mach. Intell. 41 (2) (2019) 423–443.
[17] H. Hotelling, Relations between two sets of variates, Biometrika 28 (3/4) (1936) 321–377.
[18] J. Benesty, I. Cohen, Canonical correlation analysis, in: Canonical Correlation Analysis in Speech Enhancement, Springer, 2018, pp. 5–14.
[19] P. Xu, Q. Yin, Y. Huang, Y.-Z. Song, Z. Ma, L. Wang, T. Xiang, W.B. Kleijn, J. Guo, Cross-modal subspace learning for fine-grained sketch-based image retrieval, Neurocomputing 278 (2018) 75–86.
[20] B. Scholkopf, A.J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT press, 2001.
[21] M.S. Sadooghi, S.E. Khadem, Improving one class support vector machine novelty detection scheme using nonlinear features, Pattern Recognit. 83 (2018) 14–33.
[22] B. Schölkopf, A. Smola, K.-R. Müller, Kernel principal component analysis, in: International Conference on Artificial Neural Networks, Springer, 1997, pp. 583–588.
[23] T. Kefi-Fatteh, R. Ksantini, M.-B. Kaâniche, A. Bouhoula, A novel incremental one-class support vector machine based on low variance direction, Pattern Recognit. 91 (2019) 308–321.
[24] R. Raghavendra, K.B. Raja, S. Venkatesh, C. Busch, Improved ear verification after surgery-an approach based on collaborative representation of locally competitive features, Pattern Recognit. 83 (2018) 416–429.
[25] D.M. Tax, R.P. Duin, Support vector data description, Mach. Learn. 54 (1) (2004) 45–66.

[26] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, Sv estimation of a distribution's support, Adv. Neural Inf. Process. Syst. 12 (1999).
[27] R. Sadeghi, J. Hamidzadeh, Automatic support vector data description, Soft Comput. 22 (1) (2018) 147–158.
[28] M. El Boujnouni, M. Jedra, N. Zahid, Support vector domain description with a new confidence coefficient, in: Intelligent Systems: Theories and Applications (SITA-14), 2014 9th International Conference on, IEEE, 2014, pp. 1–8.
[29] Y.-S. Jeong, R. Jayaraman, Support vector-based algorithms with weighted dynamic time warping kernel function for time series classification, Knowl. Based Syst. 75 (2015) 184–191.
[30] Y. Forghani, H.S. Yazdi, S. Effati, R.S. Tabrizi, Support vector data description by using hyper-ellipse instead of hyper-sphere, in: Computer and Knowledge Engineering (ICCKE), 2011 1st International eConference on, IEEE, 2011, pp. 22–27.
[31] V. Mygdalis, A. Iosifidis, A. Tefas, I. Pitas, Graph embedded one-class classifiers for media data classification, Pattern Recognit. 60 (2016) 585–595.
[32] D.M. Tax, P. Laskov, Online SVM learning: from classification to data description and back, in: Neural Networks for Signal Processing, 2003. NNSP'03. 2003 IEEE 13th Workshop on, IEEE, 2003, pp. 499–508.
[33] J. Hamidzadeh, N. Namaei, Belief-based chaotic algorithm for support vector data description, Soft Comput. (2018) 1–26.
[34] D. Dheeru, E. Karra Taniskidou, UCI machine learning repository, 2017. http://archive.ics.uci.edu/ml.
[35] Q.D. Tran, P. Liatsis, User-specific fusion using one-class classification for multimodal biometric systems: Boundary methods, in: 2013 Sixth International Conference on Developments in eSystems Engineering, IEEE, 2013, pp. 276–280.
[36] F. Sohrab, J. Raitoharju, M. Gabbouj, A. Iosifidis, Subspace support vector data description, in: 2018 24th International Conference on Pattern Recognition (ICPR), IEEE, 2018, pp. 722–727.
[37] N. Kwak, Nonlinear projection trick in kernel methods: an alternative to the kernel trick, IEEE Trans. Neural Netw. Learn. Syst. 24 (12) (2013) 2113–2119.
[38] A. Iosifidis, M. Gabbouj, Nyström-based approximate kernel subspace learning, Pattern Recognit. 57 (2016) 190–197.
[39] T. Elgamal, M. Hefeeda, Analysis of PCA algorithms in distributed environments, arXiv:1503.05214 (2015).
[40] S. Zheng, Smoothly approximated support vector domain description, Pattern Recognit. 49 (2016) 55–64.
[41] A. Sharma, K.K. Paliwal, S. Imoto, S. Miyano, Principal component analysis using QR decomposition, Int. J. Mach. Learn. Cybern. 4 (2013).
[42] Y. Li, F. Nie, H. Huang, J. Huang, Large-scale multi-view spectral clustering via bipartite graph, in: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015, pp. 2750–2756.
[43] L.S. Lopes, L.M. Camarinha-Matos, Feature transformation strategies for a robot learning problem, in: Feature Extraction, Construction and Selection, Springer, 1998, pp. 375–391.
[44] L.A. Kurgan, K.J. Cios, R. Tadeusiewicz, M. Ogiela, L.S. Goodenday, Knowledge discovery approach to automated cardiac SPECT diagnosis, Artif. Intell. Med. 23 (2) (2001) 149–169.
[45] F. Sohrab, J. Raitoharju, M. Gabbouj, A. Iosifidis, ms-svdd (github repository), 2020. https://github.com/fahadsohrab/mssvdd.git.

Fahad Sohrab is a PhD student in Unit of Computing Sciences, Tampere University, Finland. He received his MS degree in Electronics Engineering from Sabanci University, Istanbul, Turkey in 2016. His research interests include machine learning, pattern recognition, and anomaly detection.

Jenni Raitoharju is a Senior Research Scientist in Programme for Environmental Information at Finnish Environment Institute. Finland. She received her PhD in Information Technology in Tampere University of Technology, Finland in 2017. Her current projects deal with machine learning and pattern recognition in applications such as biomonitoring and autonomous systems.

Alexandros Iosifidis received his PhD degree in Informatics from the Aristotle University of Thessaloniki in 2014. He is an Associate Professor of Machine Learning at Aarhus University, Denmark. His research interests include statistical machine learning and artificial neural networks with applications in Computer Vision and time-series analysis problems.

Moncef Gabbouj received his MS and PhD degrees in electrical engineering from Purdue University, in 1986 and 1989, respectively. Dr. Gabbouj is Professor of Signal Processing at the Department of Computing Sciences, Tampere University, Finland. His research interests include Big Data analytics, multimedia analysis, artificial intelligence, machine learning, and pattern recognition.

Multimodal Subspace Support Vector Data Description
Supplementary Material

Fahad Sohrab, Jenni Raitoharju, Alexandros Iosifidis, Moncef Gabbouj

This document contains the extensive results comparing different variants of the proposed MS-SVDD method over 5 different datasets in Section 1, train and test time in Section 2, and the figures for sensitivity analysis of different hyperparameters over SPECTF heart data set for different regularization terms (ω) and decision strategies (ds) for linear MS-SVDD in Section 3.

1. MS-SVDD results

Table 1: Robot Execution Failures dataset test results for MS-SVDD using different decision strategies and constraints (hyperparameters selected using cross-validation on basis of maximum *gm* score on training set)

	Linear						Kernel						NPT					
	accu	tpr	tnr	pre	F1	gm	accu	tpr	tnr	pre	F1	gm	accu	tpr	tnr	pre	F1	gm
Decision Strategy 1																		
MS-SVDD ω_0	0.94	0.94	0.94	0.86	0.90	0.94	0.58	0.59	0.58	0.44	0.44	0.31	0.93	0.89	0.95	0.88	0.88	0.92
MS-SVDD ω_1	0.97	0.93	0.98	0.96	0.94	0.96	0.47	0.70	0.38	0.45	0.51	0.37	0.91	0.81	0.94	0.85	0.83	0.87
MS-SVDD ω_2	0.93	0.89	0.95	0.88	0.88	0.92	0.85	0.58	0.96	0.86	0.68	0.74	0.94	0.88	0.96	0.91	0.89	0.92
MS-SVDD ω_3	0.93	0.95	0.92	0.84	0.89	0.94	0.72	0.68	0.73	0.70	0.60	0.58	0.94	0.87	0.97	0.93	0.90	0.92
MS-SVDD ω_4	0.97	0.92	0.98	0.95	0.94	0.95	0.92	0.85	0.95	0.88	0.86	0.90	0.94	0.88	0.97	0.92	0.90	0.92
MS-SVDD ω_5	0.91	0.96	0.89	0.78	0.86	0.93	0.77	0.87	0.74	0.70	0.76	0.74	0.94	0.90	0.96	0.91	0.90	0.93
MS-SVDD ω_6	0.94	0.94	0.94	0.86	0.90	0.94	0.89	0.76	0.95	0.68	0.72	0.75	0.96	0.92	0.97	0.92	0.92	0.94
Decision Strategy 2																		
MS-SVDD ω_0	0.89	0.99	0.84	0.72	0.83	0.91	0.72	0.71	0.72	0.52	0.56	0.51	0.89	0.98	0.86	0.73	0.84	0.92
MS-SVDD ω_1	0.89	0.99	0.84	0.72	0.83	0.92	0.47	0.82	0.34	0.46	0.51	0.30	0.88	0.99	0.84	0.72	0.83	0.91
MS-SVDD ω_2	0.88	0.99	0.84	0.71	0.83	0.91	0.90	0.93	0.89	0.80	0.84	0.91	0.88	0.95	0.85	0.73	0.82	0.90
MS-SVDD ω_3	0.90	1.00	0.86	0.74	0.85	0.93	0.78	0.89	0.74	0.76	0.77	0.71	0.91	0.97	0.88	0.76	0.85	0.93
MS-SVDD ω_4	0.87	0.98	0.83	0.69	0.81	0.90	0.82	0.98	0.76	0.62	0.76	0.87	0.85	0.97	0.80	0.65	0.78	0.88
MS-SVDD ω_5	0.88	0.99	0.83	0.70	0.82	0.91	0.90	0.92	0.89	0.79	0.84	0.90	0.89	0.96	0.86	0.73	0.83	0.91
MS-SVDD ω_6	0.88	0.99	0.84	0.71	0.83	0.91	0.75	0.74	0.75	0.56	0.60	0.54	0.88	0.95	0.85	0.72	0.82	0.90
Decision Strategy 3																		
MS-SVDD ω_0	0.97	0.95	0.97	0.93	0.94	0.96	0.71	0.91	0.63	0.57	0.68	0.67	0.95	0.91	0.97	0.92	0.91	**0.94**
MS-SVDD ω_1	0.98	0.96	0.98	0.95	0.96	**0.97**	0.46	0.75	0.35	0.42	0.51	0.36	0.95	0.92	0.96	0.90	0.91	**0.94**
MS-SVDD ω_2	0.97	0.97	0.97	0.93	0.95	**0.97**	0.94	0.98	0.92	0.83	0.90	**0.95**	0.95	0.90	0.96	0.91	0.91	0.93
MS-SVDD ω_3	0.96	0.94	0.97	0.93	0.93	0.96	0.83	0.71	0.87	0.75	0.69	0.77	0.95	0.90	0.97	0.92	0.91	0.93
MS-SVDD ω_4	0.97	0.93	0.98	0.95	0.94	0.96	0.94	0.94	0.95	0.88	0.91	0.94	0.92	0.90	0.92	0.83	0.86	0.91
MS-SVDD ω_5	0.97	0.95	0.97	0.93	0.94	0.96	0.94	0.98	0.92	0.83	0.90	**0.95**	0.95	0.92	0.96	0.91	0.91	**0.94**
MS-SVDD ω_6	0.97	0.94	0.98	0.95	0.94	0.96	0.92	0.98	0.90	0.80	0.88	0.94	0.95	0.92	0.97	0.92	0.92	**0.94**
Decision Strategy 4																		
MS-SVDD ω_0	0.80	0.97	0.73	0.59	0.74	0.84	0.67	0.66	0.67	0.67	0.60	0.58	0.88	0.92	0.86	0.73	0.81	0.89
MS-SVDD ω_1	0.83	0.95	0.78	0.63	0.76	0.86	0.48	0.83	0.34	0.47	0.52	0.32	0.86	0.94	0.82	0.67	0.79	0.88
MS-SVDD ω_2	0.83	0.95	0.78	0.64	0.76	0.86	0.63	0.85	0.54	0.61	0.64	0.49	0.86	0.90	0.85	0.70	0.79	0.87
MS-SVDD ω_3	0.82	0.91	0.78	0.63	0.74	0.84	0.75	0.64	0.80	0.66	0.59	0.51	0.87	0.92	0.85	0.71	0.80	0.88
MS-SVDD ω_4	0.83	0.93	0.79	0.64	0.75	0.86	0.66	0.80	0.60	0.49	0.60	0.61	0.88	0.94	0.85	0.71	0.81	0.90
MS-SVDD ω_5	0.85	0.90	0.83	0.68	0.77	0.86	0.68	0.87	0.61	0.57	0.65	0.65	0.86	0.87	0.86	0.71	0.78	0.86
MS-SVDD ω_6	0.81	0.96	0.76	0.61	0.75	0.85	0.57	0.49	0.60	0.30	0.32	0.17	0.87	0.93	0.85	0.71	0.80	0.89

Table 2: Caltech-7 dataset test results for MS-SVDD using different decision strategies and constraints (hyperparameters selected using cross-validation on basis of maximum *gm* score on training set)

	Linear						Kernel						NPT					
	accu	tpr	tnr	pre	F1	gm	accu	tpr	tnr	pre	F1	gm	accu	tpr	tnr	pre	F1	gm
Decision Strategy 1																		
MS-SVDD ω_0	0.87	0.98	0.83	0.71	0.82	0.90	0.66	0.89	0.57	0.50	0.63	0.68	0.90	0.96	0.88	0.77	0.86	0.92
MS-SVDD ω_1	0.91	0.96	0.89	0.78	0.86	**0.92**	0.65	0.89	0.55	0.59	0.68	0.59	0.94	0.98	0.92	0.85	0.91	**0.95**
MS-SVDD ω_2	0.87	0.98	0.83	0.71	0.82	0.90	0.86	0.94	0.82	0.72	0.81	0.88	0.92	0.95	0.90	0.80	0.87	0.93
MS-SVDD ω_3	0.87	0.98	0.83	0.70	0.82	0.90	0.66	0.89	0.57	0.50	0.63	0.68	0.90	0.97	0.86	0.76	0.85	0.92
MS-SVDD ω_4	0.91	0.95	0.89	0.78	0.86	**0.92**	0.92	0.96	0.91	0.81	0.88	**0.93**	0.94	0.95	0.94	0.88	0.91	**0.95**
MS-SVDD ω_5	0.88	0.98	0.84	0.71	0.83	0.90	0.90	0.94	0.89	0.78	0.85	0.91	0.93	0.96	0.91	0.82	0.88	0.94
MS-SVDD ω_6	0.87	0.98	0.83	0.71	0.82	0.90	0.66	0.71	0.63	0.40	0.49	0.53	0.90	0.98	0.86	0.75	0.85	0.92
Decision Strategy 2																		
MS-SVDD ω_0	0.60	0.97	0.44	0.42	0.59	0.65	0.68	0.78	0.63	0.60	0.60	0.61	0.78	0.76	0.78	0.64	0.62	0.72
MS-SVDD ω_1	0.74	0.97	0.65	0.53	0.69	0.79	0.78	0.96	0.70	0.57	0.71	0.82	0.81	0.97	0.74	0.61	0.75	0.85
MS-SVDD ω_2	0.63	0.97	0.49	0.44	0.61	0.69	0.69	0.68	0.69	0.60	0.52	0.60	0.78	0.76	0.78	0.64	0.62	0.72
MS-SVDD ω_3	0.59	0.97	0.43	0.42	0.58	0.64	0.73	0.75	0.72	0.62	0.61	0.70	0.79	0.94	0.73	0.60	0.73	0.83
MS-SVDD ω_4	0.74	0.97	0.64	0.53	0.68	0.79	0.73	0.97	0.62	0.52	0.68	0.78	0.83	0.95	0.78	0.64	0.77	0.86
MS-SVDD ω_5	0.62	0.97	0.47	0.44	0.60	0.68	0.83	0.79	0.85	0.74	0.73	0.80	0.79	0.94	0.73	0.60	0.73	0.83
MS-SVDD ω_6	0.60	0.97	0.44	0.42	0.59	0.65	0.68	0.86	0.61	0.58	0.62	0.63	0.79	0.94	0.73	0.60	0.73	0.83
Decision Strategy 3																		
MS-SVDD ω_0	0.70	0.81	0.65	0.49	0.61	0.73	0.68	0.65	0.70	0.50	0.53	0.64	0.76	0.83	0.74	0.58	0.68	0.78
MS-SVDD ω_1	0.74	0.83	0.69	0.53	0.65	0.76	0.76	0.84	0.73	0.57	0.68	0.78	0.75	0.87	0.71	0.57	0.69	0.78
MS-SVDD ω_2	0.72	0.85	0.67	0.52	0.64	0.75	0.76	0.81	0.74	0.57	0.67	0.77	0.77	0.86	0.73	0.59	0.69	0.79
MS-SVDD ω_3	0.70	0.82	0.65	0.49	0.61	0.73	0.56	0.76	0.48	0.41	0.52	0.52	0.76	0.83	0.74	0.58	0.68	0.78
MS-SVDD ω_4	0.73	0.83	0.69	0.53	0.64	0.76	0.76	0.89	0.71	0.57	0.69	0.79	0.58	0.94	0.43	0.42	0.57	0.63
MS-SVDD ω_5	0.72	0.82	0.67	0.51	0.63	0.75	0.76	0.84	0.73	0.56	0.67	0.78	0.76	0.89	0.70	0.56	0.68	0.79
MS-SVDD ω_6	0.69	0.81	0.64	0.49	0.61	0.72	0.65	0.71	0.62	0.46	0.55	0.65	0.77	0.82	0.75	0.59	0.68	0.78
Decision Strategy 4																		
MS-SVDD ω_0	0.78	0.95	0.70	0.57	0.72	0.82	0.74	0.57	0.81	0.69	0.51	0.60	0.86	0.97	0.81	0.68	0.80	0.89
MS-SVDD ω_1	0.92	0.89	0.93	0.84	0.86	0.91	0.75	0.97	0.65	0.61	0.73	0.71	0.90	0.99	0.86	0.74	0.85	0.92
MS-SVDD ω_2	0.77	0.93	0.71	0.57	0.71	0.81	0.87	0.92	0.85	0.73	0.81	0.88	0.90	0.94	0.88	0.77	0.85	0.91
MS-SVDD ω_3	0.77	0.96	0.69	0.56	0.71	0.81	0.71	0.60	0.75	0.48	0.47	0.54	0.84	0.98	0.78	0.66	0.79	0.87
MS-SVDD ω_4	0.90	0.91	0.90	0.81	0.85	0.90	0.88	0.97	0.84	0.72	0.83	0.90	0.90	0.96	0.88	0.78	0.86	0.92
MS-SVDD ω_5	0.78	0.93	0.71	0.58	0.71	0.81	0.85	0.97	0.81	0.68	0.80	0.88	0.90	0.95	0.88	0.77	0.85	0.92
MS-SVDD ω_6	0.77	0.95	0.70	0.57	0.71	0.81	0.72	0.39	0.86	0.39	0.34	0.40	0.84	0.97	0.79	0.67	0.79	0.88

Table 3: Ionosphere dataset test results for MS-SVDD using different decision strategies and constraints (hyperparameters selected using cross-validation on basis of maximum *gm* score on training set)

	Linear						Kernel						NPT					
	accu	tpr	tnr	pre	F1	gm	accu	tpr	tnr	pre	F1	gm	accu	tpr	tnr	pre	F1	gm
Decision Strategy 1																		
MS-SVDD ω_0	0.81	0.91	0.62	0.82	0.86	0.75	0.73	0.82	0.56	0.78	0.79	0.67	0.80	0.92	0.58	0.82	0.86	0.65
MS-SVDD ω_1	0.84	0.90	0.73	0.86	0.88	0.81	0.81	0.91	0.64	0.83	0.86	0.75	0.85	0.94	0.69	0.87	0.90	0.71
MS-SVDD ω_2	0.82	0.94	0.60	0.81	0.87	0.75	0.76	0.85	0.60	0.80	0.82	0.71	0.82	0.89	0.70	0.87	0.87	0.69
MS-SVDD ω_3	0.80	0.91	0.60	0.81	0.86	0.73	0.73	0.82	0.56	0.78	0.79	0.67	0.82	0.94	0.61	0.83	0.88	0.66
MS-SVDD ω_4	0.82	0.90	0.69	0.84	0.87	0.79	0.80	0.95	0.53	0.79	0.86	0.69	0.82	0.93	0.61	0.83	0.87	0.67
MS-SVDD ω_5	0.83	0.95	0.63	0.82	0.88	0.77	0.83	0.88	0.74	0.86	0.87	**0.81**	0.81	0.92	0.62	0.83	0.87	0.66
MS-SVDD ω_6	0.81	0.93	0.59	0.81	0.86	0.74	0.73	0.82	0.56	0.78	0.79	0.67	0.84	0.94	0.66	0.86	0.89	0.69
Decision Strategy 2																		
MS-SVDD ω_0	0.86	0.96	0.67	0.84	0.90	0.80	0.58	0.45	0.81	0.82	0.57	0.59	0.71	0.85	0.46	0.75	0.79	0.54
MS-SVDD ω_1	0.86	0.96	0.67	0.84	0.90	0.80	0.64	0.76	0.42	0.72	0.73	0.48	0.67	0.71	0.61	0.80	0.70	0.52
MS-SVDD ω_2	0.87	0.96	0.70	0.86	0.90	0.82	0.58	0.59	0.58	0.83	0.60	0.33	0.71	0.86	0.45	0.75	0.79	0.54
MS-SVDD ω_3	0.86	0.96	0.67	0.84	0.90	0.80	0.54	0.52	0.58	0.73	0.56	0.42	0.71	0.85	0.46	0.75	0.79	0.54
MS-SVDD ω_4	0.85	0.96	0.64	0.83	0.89	0.78	0.74	0.85	0.53	0.77	0.81	0.66	0.70	0.81	0.51	0.76	0.78	0.55
MS-SVDD ω_5	0.87	0.96	0.71	0.86	0.91	0.82	0.60	0.43	0.90	0.94	0.51	0.54	0.71	0.85	0.46	0.75	0.79	0.55
MS-SVDD ω_6	0.86	0.96	0.68	0.85	0.90	0.80	0.63	0.58	0.73	0.83	0.64	0.60	0.71	0.85	0.46	0.75	0.79	0.54
Decision Strategy 3																		
MS-SVDD ω_0	0.83	0.99	0.53	0.79	0.88	0.72	0.70	0.77	0.57	0.78	0.77	0.64	0.85	0.96	0.63	0.83	0.89	0.78
MS-SVDD ω_1	0.82	0.92	0.63	0.83	0.87	0.75	0.80	0.92	0.58	0.81	0.85	0.72	0.84	0.99	0.58	0.82	0.89	0.67
MS-SVDD ω_2	0.78	0.91	0.55	0.81	0.83	0.68	0.78	0.79	0.76	0.86	0.82	0.77	0.87	0.96	0.69	0.85	0.90	0.82
MS-SVDD ω_3	0.78	0.97	0.43	0.76	0.85	0.64	0.74	0.84	0.54	0.78	0.80	0.65	0.80	0.95	0.52	0.80	0.86	0.62
MS-SVDD ω_4	0.81	0.97	0.52	0.79	0.87	0.70	0.82	0.94	0.60	0.82	0.87	0.75	0.84	0.96	0.63	0.83	0.89	0.78
MS-SVDD ω_5	0.68	0.73	0.60	0.78	0.70	0.61	0.80	0.91	0.59	0.81	0.85	0.73	0.85	0.98	0.61	0.82	0.89	0.77
MS-SVDD ω_6	0.80	0.97	0.50	0.78	0.86	0.69	0.74	0.84	0.54	0.78	0.80	0.65	0.83	0.96	0.59	0.82	0.88	0.67
Decision Strategy 4																		
MS-SVDD ω_0	0.85	0.95	0.68	0.85	0.89	0.80	0.62	0.67	0.53	0.70	0.66	0.56	0.71	0.79	0.56	0.79	0.76	0.64
MS-SVDD ω_1	0.83	0.91	0.69	0.84	0.87	0.79	0.66	0.80	0.38	0.70	0.75	0.55	0.88	0.95	0.74	0.87	0.91	**0.84**
MS-SVDD ω_2	0.87	0.95	0.73	0.87	0.91	**0.83**	0.63	0.73	0.45	0.69	0.68	0.50	0.76	0.86	0.59	0.79	0.82	0.71
MS-SVDD ω_3	0.86	0.96	0.69	0.85	0.90	0.81	0.61	0.70	0.44	0.68	0.67	0.51	0.79	0.95	0.50	0.78	0.85	0.69
MS-SVDD ω_4	0.81	0.88	0.68	0.84	0.86	0.78	0.76	0.86	0.58	0.79	0.82	0.71	0.79	0.94	0.52	0.78	0.85	0.70
MS-SVDD ω_5	0.87	0.95	0.73	0.87	0.91	**0.83**	0.69	0.83	0.44	0.73	0.77	0.59	0.72	0.78	0.62	0.80	0.77	0.68
MS-SVDD ω_6	0.86	0.95	0.71	0.86	0.90	0.82	0.62	0.67	0.53	0.70	0.66	0.56	0.77	0.92	0.50	0.77	0.84	0.68

Table 4: Handwritten dataset test results for MS-SVDD using different decision strategies and constraints (hyperparameters selected using cross-validation on basis of maximum *gm* score on training set)

	Linear						Kernel						NPT					
	accu	tpr	tnr	pre	F1	gm	accu	tpr	tnr	pre	F1	gm	accu	tpr	tnr	pre	F1	gm
Decision Strategy 1																		
MS-SVDD ω_0	0.98	0.91	0.99	0.91	0.91	0.95	0.82	0.36	0.87	0.19	0.25	0.35	0.98	0.91	0.99	0.89	0.90	0.95
MS-SVDD ω_1	0.98	0.92	0.99	0.89	0.90	0.95	0.96	0.88	0.96	0.77	0.80	0.92	0.94	0.98	0.94	0.80	0.85	0.96
MS-SVDD ω_2	0.99	0.91	0.99	0.95	0.93	0.95	0.72	0.70	0.72	0.37	0.43	0.54	0.98	0.92	0.98	0.89	0.90	0.95
MS-SVDD ω_3	0.97	0.92	0.97	0.84	0.86	0.94	0.81	0.19	0.88	0.07	0.10	0.17	0.98	0.91	0.99	0.89	0.90	0.95
MS-SVDD ω_4	0.98	0.90	0.99	0.89	0.89	0.94	0.95	0.93	0.95	0.85	0.85	0.94	0.98	0.95	0.99	0.91	0.93	**0.97**
MS-SVDD ω_5	0.96	0.91	0.97	0.80	0.84	0.94	0.95	0.97	0.94	0.70	0.80	0.95	0.97	0.94	0.97	0.84	0.87	0.95
MS-SVDD ω_6	0.96	0.92	0.97	0.82	0.85	0.94	0.75	0.78	0.75	0.23	0.35	0.65	0.98	0.91	0.99	0.89	0.90	0.95
Decision Strategy 2																		
MS-SVDD ω_0	0.85	0.99	0.84	0.46	0.61	0.91	0.33	1.00	0.25	0.15	0.26	0.32	0.60	0.99	0.56	0.51	0.57	0.58
MS-SVDD ω_1	0.91	0.99	0.90	0.60	0.72	0.94	0.74	0.96	0.71	0.50	0.60	0.74	0.46	0.98	0.40	0.45	0.50	0.39
MS-SVDD ω_2	0.90	0.99	0.89	0.63	0.74	0.93	0.65	0.80	0.63	0.31	0.34	0.56	0.60	0.99	0.56	0.51	0.57	0.58
MS-SVDD ω_3	0.86	0.99	0.84	0.45	0.61	0.91	0.59	0.58	0.59	0.24	0.26	0.19	0.60	0.99	0.56	0.51	0.57	0.58
MS-SVDD ω_4	0.94	0.99	0.93	0.65	0.77	0.96	0.95	0.96	0.95	0.76	0.83	0.96	0.81	0.91	0.80	0.79	0.77	0.75
MS-SVDD ω_5	0.85	0.99	0.84	0.49	0.64	0.91	0.83	0.88	0.83	0.49	0.55	0.84	0.60	0.99	0.55	0.46	0.55	0.57
MS-SVDD ω_6	0.85	0.99	0.83	0.44	0.59	0.90	0.38	0.61	0.36	0.09	0.15	0.24	0.60	0.99	0.56	0.51	0.57	0.58
Decision Strategy 3																		
MS-SVDD ω_0	0.84	0.90	0.83	0.38	0.53	0.86	0.33	1.00	0.25	0.15	0.26	0.32	0.89	0.85	0.90	0.49	0.61	0.87
MS-SVDD ω_1	0.85	0.86	0.85	0.40	0.54	0.86	0.63	0.80	0.61	0.26	0.38	0.64	0.96	0.84	0.97	0.79	0.81	0.91
MS-SVDD ω_2	0.84	0.90	0.84	0.39	0.54	0.87	0.65	0.91	0.62	0.30	0.42	0.66	0.91	0.88	0.91	0.54	0.66	0.89
MS-SVDD ω_3	0.84	0.88	0.83	0.38	0.53	0.85	0.31	0.98	0.23	0.14	0.24	0.30	0.89	0.85	0.90	0.49	0.61	0.87
MS-SVDD ω_4	0.81	0.88	0.80	0.35	0.50	0.84	0.84	0.94	0.83	0.39	0.55	0.88	0.94	0.92	0.94	0.64	0.75	0.93
MS-SVDD ω_5	0.82	0.94	0.80	0.35	0.51	0.87	0.73	0.94	0.71	0.29	0.44	0.81	0.88	0.91	0.88	0.46	0.61	0.89
MS-SVDD ω_6	0.83	0.91	0.82	0.38	0.53	0.86	0.34	0.98	0.27	0.15	0.25	0.38	0.89	0.85	0.90	0.49	0.61	0.87
Decision Strategy 4																		
MS-SVDD ω_0	0.89	0.99	0.87	0.55	0.69	0.93	0.39	0.30	0.40	0.06	0.09	0.19	0.94	0.84	0.95	0.77	0.72	0.87
MS-SVDD ω_1	0.93	0.99	0.92	0.62	0.75	0.95	0.97	0.96	0.97	0.85	0.88	0.96	0.94	0.84	0.95	0.77	0.72	0.87
MS-SVDD ω_2	0.92	0.99	0.91	0.76	0.82	0.94	0.67	0.63	0.67	0.18	0.27	0.43	0.94	0.84	0.95	0.77	0.72	0.87
MS-SVDD ω_3	0.88	0.99	0.87	0.51	0.66	0.93	0.59	0.45	0.60	0.26	0.29	0.39	0.94	0.84	0.95	0.77	0.72	0.87
MS-SVDD ω_4	0.98	0.99	0.98	0.90	0.93	**0.98**	1.00	0.99	1.00	0.98	0.98	**0.99**	0.94	0.84	0.95	0.77	0.72	0.87
MS-SVDD ω_5	0.92	0.99	0.91	0.78	0.83	0.94	0.98	0.84	0.99	0.93	0.86	0.90	0.94	0.84	0.95	0.77	0.72	0.87
MS-SVDD ω_6	0.88	0.99	0.87	0.56	0.69	0.92	0.60	0.79	0.58	0.16	0.26	0.53	0.94	0.84	0.95	0.77	0.72	0.87

Table 5: SPECTF heart dataset test results for MS-SVDD using different decision strategies and constraints (hyper-parameters selected using cross-validation on basis of maximum *gm* score on training set)

	Linear						Kernel						NPT					
	accu	tpr	tnr	pre	F1	gm	accu	tpr	tnr	pre	F1	gm	accu	tpr	tnr	pre	F1	gm
Decision Strategy 1																		
MS-SVDD ω_0	0.78	0.80	0.78	0.24	0.37	**0.79**	0.55	0.60	0.55	0.10	0.18	0.57	0.65	0.73	0.65	0.15	0.25	0.69
MS-SVDD ω_1	0.73	0.80	0.73	0.20	0.32	0.76	0.55	0.60	0.55	0.10	0.18	0.57	0.73	0.60	0.74	0.17	0.26	0.67
MS-SVDD ω_2	0.78	0.80	0.77	0.24	0.36	**0.79**	0.80	0.73	0.80	0.24	0.37	**0.77**	0.72	0.73	0.72	0.19	0.30	**0.73**
MS-SVDD ω_3	0.65	0.80	0.64	0.16	0.27	0.72	0.63	0.87	0.60	0.16	0.27	0.72	0.65	0.73	0.65	0.15	0.25	0.69
MS-SVDD ω_4	0.74	0.80	0.74	0.21	0.33	0.77	0.55	0.60	0.55	0.10	0.18	0.57	0.71	0.73	0.70	0.18	0.29	0.72
MS-SVDD ω_5	0.76	0.73	0.76	0.21	0.33	0.75	0.77	0.67	0.78	0.21	0.32	0.72	0.72	0.73	0.72	0.19	0.30	**0.73**
MS-SVDD ω_6	0.67	0.80	0.66	0.17	0.28	0.73	0.63	0.87	0.60	0.16	0.27	0.72	0.72	0.67	0.72	0.17	0.27	0.69
Decision Strategy 2																		
MS-SVDD ω_0	0.69	0.67	0.69	0.16	0.26	0.68	0.80	0.47	0.83	0.19	0.27	0.62	0.74	0.60	0.76	0.18	0.27	0.67
MS-SVDD ω_1	0.58	0.80	0.56	0.14	0.23	0.67	0.49	0.87	0.45	0.12	0.21	0.63	0.76	0.53	0.78	0.18	0.27	0.65
MS-SVDD ω_2	0.37	0.93	0.33	0.11	0.19	0.55	0.60	0.80	0.59	0.14	0.24	0.69	0.75	0.67	0.76	0.20	0.30	0.71
MS-SVDD ω_3	0.65	0.67	0.65	0.14	0.24	0.66	0.80	0.47	0.83	0.19	0.27	0.62	0.74	0.60	0.76	0.18	0.27	0.67
MS-SVDD ω_4	0.57	0.93	0.53	0.15	0.26	0.71	0.75	0.60	0.77	0.18	0.28	0.68	0.76	0.60	0.78	0.19	0.29	0.68
MS-SVDD ω_5	0.60	0.93	0.57	0.16	0.27	0.73	0.60	0.80	0.59	0.14	0.24	0.69	0.75	0.67	0.76	0.20	0.30	0.71
MS-SVDD ω_6	0.49	0.87	0.46	0.12	0.21	0.63	0.80	0.47	0.83	0.19	0.27	0.62	0.74	0.60	0.76	0.18	0.27	0.67
Decision Strategy 3																		
MS-SVDD ω_0	0.59	0.80	0.58	0.14	0.24	0.68	0.66	0.60	0.66	0.13	0.22	0.63	0.80	0.47	0.83	0.19	0.27	0.62
MS-SVDD ω_1	0.61	0.80	0.60	0.15	0.25	0.69	0.76	0.47	0.79	0.16	0.24	0.61	0.80	0.47	0.83	0.19	0.27	0.62
MS-SVDD ω_2	0.53	0.87	0.51	0.13	0.23	0.66	0.47	0.93	0.43	0.13	0.22	0.63	0.80	0.47	0.83	0.19	0.27	0.62
MS-SVDD ω_3	0.58	0.87	0.56	0.15	0.25	0.70	0.66	0.60	0.66	0.13	0.22	0.63	0.80	0.47	0.83	0.19	0.27	0.62
MS-SVDD ω_4	0.55	0.73	0.53	0.12	0.21	0.62	0.78	0.47	0.80	0.17	0.25	0.61	0.80	0.53	0.82	0.21	0.30	0.66
MS-SVDD ω_5	0.58	0.93	0.55	0.15	0.26	0.72	0.62	0.80	0.60	0.15	0.25	0.70	0.80	0.47	0.83	0.19	0.27	0.62
MS-SVDD ω_6	0.56	0.87	0.53	0.14	0.24	0.68	0.66	0.60	0.66	0.13	0.22	0.63	0.80	0.47	0.83	0.19	0.27	0.62
Decision Strategy 4																		
MS-SVDD ω_0	0.68	0.73	0.67	0.16	0.27	0.70	0.72	0.73	0.72	0.18	0.29	0.72	0.72	0.73	0.72	0.18	0.29	0.72
MS-SVDD ω_1	0.72	0.80	0.71	0.19	0.31	0.75	0.52	0.87	0.49	0.13	0.23	0.65	0.63	0.73	0.62	0.14	0.24	0.68
MS-SVDD ω_2	0.73	0.80	0.73	0.20	0.32	0.76	0.60	0.87	0.58	0.15	0.26	0.71	0.60	0.80	0.58	0.14	0.24	0.68
MS-SVDD ω_3	0.75	0.53	0.77	0.17	0.26	0.64	0.60	0.80	0.59	0.14	0.24	0.69	0.66	0.73	0.65	0.15	0.26	0.69
MS-SVDD ω_4	0.73	0.80	0.72	0.20	0.32	0.76	0.74	0.67	0.74	0.19	0.29	0.70	0.65	0.73	0.65	0.15	0.25	0.69
MS-SVDD ω_5	0.74	0.73	0.74	0.20	0.31	0.74	0.75	0.60	0.77	0.18	0.28	0.68	0.74	0.60	0.75	0.17	0.27	0.67
MS-SVDD ω_6	0.76	0.53	0.78	0.17	0.26	0.64	0.72	0.73	0.72	0.18	0.29	0.72	0.72	0.73	0.72	0.18	0.29	0.72

2. Train and test time

The numerical results of the train and test time (Tables 6-10) were computed using Matlab version 9.6.0.1072779 (R2019a) over, Intel64 Family 6 Model 158 Stepping 9 GenuineIntel 2808 Mhz processor.

Table 6: Train and test time (in milliseconds) for Robot Execution Failures dataset

	Linear		Non-linear	
	Train Time	Test Time	Train Time	Test Time
Proposed method				
MS-SVDD $\omega_2 ds3$	1.01E+02	3.43E-01	1.66E+02	3.26E-01
MS-SVDD $\omega_5 ds3$	9.03E+01	3.58E-01	1.65E+02	3.63E-01
Concatenated features				
S-SVDD ψ_1	4.66E+01	1.04E-01	8.52E+01	1.97E-01
S-SVDD ψ_2	4.50E+01	1.39E-01	1.14E+02	3.35E-01
S-SVDD ψ_3	4.79E+01	1.21E-01	8.86E+01	1.67E-01
S-SVDD ψ_4	4.22E+01	8.37E-02	8.94E+01	2.47E-01
OC-SVM	4.65E+00	1.25E+00	1.79E+00	1.50E+00
SVDD	3.73E-01	2.03E-01	4.38E-01	2.24E-01
Force measurements				
S-SVDD ψ_1	3.29E+01	7.91E-02	8.80E+01	2.29E-01
S-SVDD ψ_2	3.60E+01	7.83E-02	8.06E+01	2.12E-01
S-SVDD ψ_3	3.55E+01	1.03E-01	9.46E+01	1.36E-01
S-SVDD ψ_4	3.03E+01	6.59E-02	1.01E+02	2.71E-01
OC-SVM	1.42E+01	5.69E-01	1.15E+00	7.52E-01
SVDD	2.39E-01	9.61E-02	5.70E-01	2.05E-01
Torque measurements				
S-SVDD ψ_1	2.67E+01	5.47E-02	1.22E+02	2.27E-01
S-SVDD ψ_2	4.07E+01	1.29E-01	8.79E+01	1.56E-01
S-SVDD ψ_3	2.97E+01	5.10E-02	8.78E+01	2.12E-01
S-SVDD ψ_4	4.02E+01	9.23E-02	9.03E+01	2.02E-01
OC-SVM	7.17E+00	6.01E-01	9.74E-01	8.20E-01
SVDD	2.71E-01	9.82E-02	3.99E-01	2.05E-01

Table 7: Train and test time (in milliseconds) for Caltec-7 dataset

	Linear		Non-linear	
	Train Time	Test Time	Train Time	Test Time
Proposed method				
MS-SVDD $\omega_1 ds1$	1.81E+02	7.61E-01	5.36E+02	1.29E+00
MS-SVDD $\omega_4 ds1$	1.94E+02	7.95E-01	4.11E+02	1.28E+00
Concatenated features				
S-SVDD ψ_1	1.37E+02	2.46E-01	7.15E+02	4.33E-01
S-SVDD ψ_2	1.54E+02	3.04E-01	7.79E+02	1.04E+00
S-SVDD ψ_3	1.83E+02	2.96E-01	9.54E+02	1.29E+00
S-SVDD ψ_4	1.63E+02	3.06E-01	1.03E+03	1.43E+00
OC-SVM	5.45E+02	1.41E+01	1.47E+01	1.34E+01
SVDD	2.22E+00	1.43E+00	2.13E+00	2.18E+00
Gabor features				
S-SVDD ψ_1	1.71E+02	3.49E-01	7.97E+02	4.14E-01
S-SVDD ψ_2	1.96E+02	3.73E-01	7.66E+02	3.94E-01
S-SVDD ψ_3	2.01E+02	3.86E-01	8.64E+02	8.11E-01
S-SVDD ψ_4	2.00E+02	5.35E-01	8.36E+02	1.54E+00
OC-SVM	8.14E+01	7.21E+00	1.07E+01	9.17E+00
SVDD	4.73E+00	3.20E+00	1.63E+00	1.30E+00
Wavelet moments				
S-SVDD ψ_1	1.62E+02	4.77E-01	7.02E+02	7.31E-01
S-SVDD ψ_2	1.70E+02	3.77E-01	7.33E+02	6.51E-01
S-SVDD ψ_3	1.03E+02	1.47E-01	8.01E+02	1.11E+00
S-SVDD ψ_4	1.33E+02	2.70E-01	8.22E+02	9.00E-01
OC-SVM	1.90E+02	5.56E+00	1.05E+01	7.89E+00
SVDD	1.17E+00	5.55E-01	1.46E+00	8.10E-01

Table 8: Train and test time (in milliseconds) for Ionosphere dataset

	Linear		Non-linear	
	Train Time	Test Time	Train Time	Test Time
Proposed method				
MS-SVDD $\omega_2 ds4$	1.68E+02	3.82E-01	2.11E+02	3.99E-01
MS-SVDD $\omega_1 ds4$	1.79E+02	3.78E-01	2.29E+02	3.99E-01
Concatenated features				
S-SVDD ψ_1	5.65E+01	6.92E-02	1.50E+02	2.29E-01
S-SVDD ψ_2	7.58E+01	1.82E-01	1.43E+02	1.95E-01
S-SVDD ψ_3	8.98E+01	1.56E-01	1.95E+02	3.43E-01
S-SVDD ψ_4	5.10E+01	9.00E-02	2.04E+02	3.54E-01
OC-SVM	5.43E+00	5.81E-01	2.29E+00	1.09E+00
SVDD	4.36E-01	1.13E-01	7.73E-01	2.48E-01
Real				
S-SVDD ψ_1	3.92E+01	8.90E-02	1.49E+02	2.81E-01
S-SVDD ψ_2	3.92E+01	1.07E-01	1.52E+02	2.15E-01
S-SVDD ψ_3	5.37E+01	9.95E-02	1.69E+02	2.55E-01
S-SVDD ψ_4	4.92E+01	1.06E-01	1.83E+02	3.47E-01
OC-SVM	1.70E+00	3.04E-01	2.01E+00	8.21E-01
SVDD	3.67E-01	9.25E-02	5.28E-01	1.38E-01
Complex				
S-SVDD ψ_1	1.00E+02	2.55E-01	1.59E+02	2.87E-01
S-SVDD ψ_2	9.51E+01	1.42E-01	1.54E+02	1.42E-01
S-SVDD ψ_3	1.06E+02	1.65E-01	1.98E+02	4.89E-01
S-SVDD ψ_4	1.02E+02	1.48E-01	2.05E+02	5.08E-01
OC-SVM	1.88E+00	3.09E-01	2.39E+00	8.61E-01
SVDD	3.75E-01	9.73E-02	7.06E-01	2.05E-01

Table 9: Train and test time (in milliseconds) for Handwritten dataset

	Linear		Non-linear	
	Train Time	Test Time	Train Time	Test Time
Proposed method				
MS-SVDD $\omega_4 ds4$	1.18E+02	1.28E+00	1.07E+02	1.14E+00
MS-SVDD $\omega_4 ds1$	1.30E+02	1.39E+00	1.42E+02	1.94E+00
Concatenated features				
S-SVDD ψ_1	4.19E+01	2.27E-01	2.47E+02	6.25E-01
S-SVDD ψ_2	4.09E+01	2.56E-01	2.84E+02	8.02E-01
S-SVDD ψ_3	4.69E+01	1.34E-01	2.66E+02	7.09E-01
S-SVDD ψ_4	5.06E+01	2.44E-01	2.67E+02	7.02E-01
OC-SVM	1.04E+02	6.01E+00	2.89E+00	6.22E+00
SVDD	4.24E-01	5.09E-01	5.32E-01	8.47E-01
ZER				
S-SVDD ψ_1	4.20E+01	1.39E-01	2.42E+02	3.57E-01
S-SVDD ψ_2	4.41E+01	1.54E-01	2.93E+02	1.44E+00
S-SVDD ψ_3	6.21E+01	3.45E-01	2.88E+02	5.51E-01
S-SVDD ψ_4	6.73E+01	4.14E-01	2.83E+02	7.22E-01
OC-SVM	9.63E+01	5.17E+00	2.31E+00	5.65E+00
SVDD	5.21E-01	6.69E-01	8.43E-01	1.36E+00
MOR				
S-SVDD ψ_1	3.44E+01	9.09E-02	2.40E+02	3.96E-01
S-SVDD ψ_2	2.64E+01	1.12E-01	3.01E+02	7.59E-01
S-SVDD ψ_3	3.48E+01	1.05E-01	2.58E+02	4.90E-01
S-SVDD ψ_4	2.93E+01	1.46E-01	2.52E+02	3.43E-01
OC-SVM	4.09E-01	5.72E-01	1.31E+00	3.48E+00
SVDD	2.76E-01	2.06E-01	5.39E-01	4.33E-01

Table 10: Train and test time (in milliseconds) for SPECTF heart dataset

	Linear		Non-linear	
	Train Time	Test Time	Train Time	Test Time
Proposed method				
MS-SVDD $\omega_0 ds1$	4.16E+01	5.44E-01	5.20E+01	5.41E-01
MS-SVDD $\omega_2 ds1$	4.60E+01	5.82E-01	5.07E+01	2.29E-01
Concatenated features				
S-SVDD ψ_1	6.27E+01	7.59E-02	6.19E+01	3.35E-01
S-SVDD ψ_2	5.55E+01	1.42E-01	3.51E+01	1.72E-01
S-SVDD ψ_3	3.21E+01	1.39E-01	6.34E+01	2.91E-01
S-SVDD ψ_4	4.02E+01	7.66E-02	3.89E+01	2.36E-01
OC-SVM	9.03E-01	6.78E-01	1.38E+00	1.31E+00
SVDD	5.70E-01	2.66E-01	4.46E-01	3.30E-01
Rest Mode				
S-SVDD ψ_1	2.79E+01	6.45E-02	3.90E+01	3.31E-01
S-SVDD ψ_2	1.96E+01	7.84E-02	4.56E+01	1.40E-01
S-SVDD ψ_3	2.66E+01	6.42E-02	6.08E+01	1.27E-01
S-SVDD ψ_4	3.29E+01	7.99E-02	4.17E+01	3.34E-01
OC-SVM	6.63E-01	2.03E-01	9.58E-01	9.93E-01
SVDD	4.08E-01	2.03E-01	3.06E-01	1.55E-01
Stress Mode				
S-SVDD ψ_1	3.02E+01	6.45E-02	3.40E+01	1.39E-01
S-SVDD ψ_2	2.78E+01	6.24E-02	3.97E+01	4.21E-01
S-SVDD ψ_3	2.68E+01	5.54E-02	5.56E+01	2.90E-01
S-SVDD ψ_4	3.70E+01	1.72E-01	4.10E+01	3.36E-01
OC-SVM	9.89E-01	1.79E-01	6.78E-01	6.59E-01
SVDD	2.70E-01	9.01E-02	7.54E-01	4.37E-01

3. Sensitivity analysis

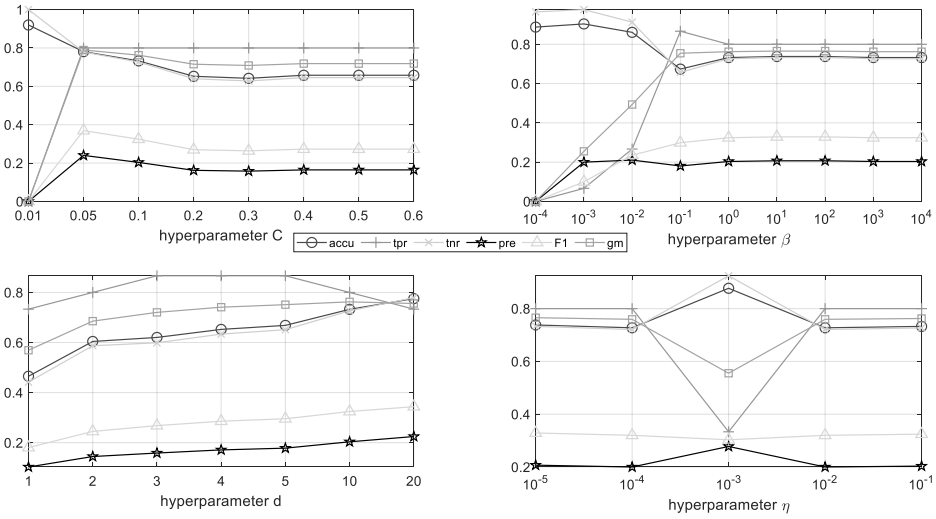

Figure 1: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_1 ds1$ on SPECTF heart dataset

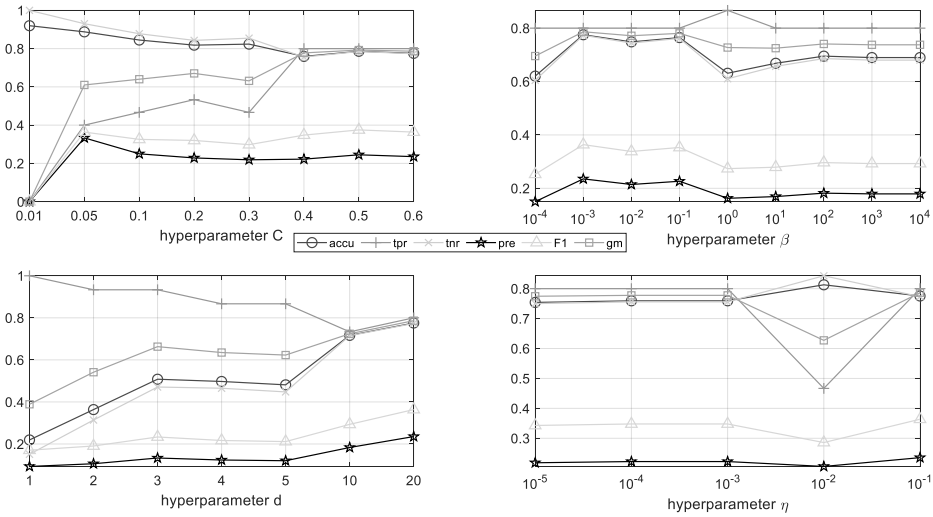

Figure 2: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_2 ds1$ on SPECTF heart dataset

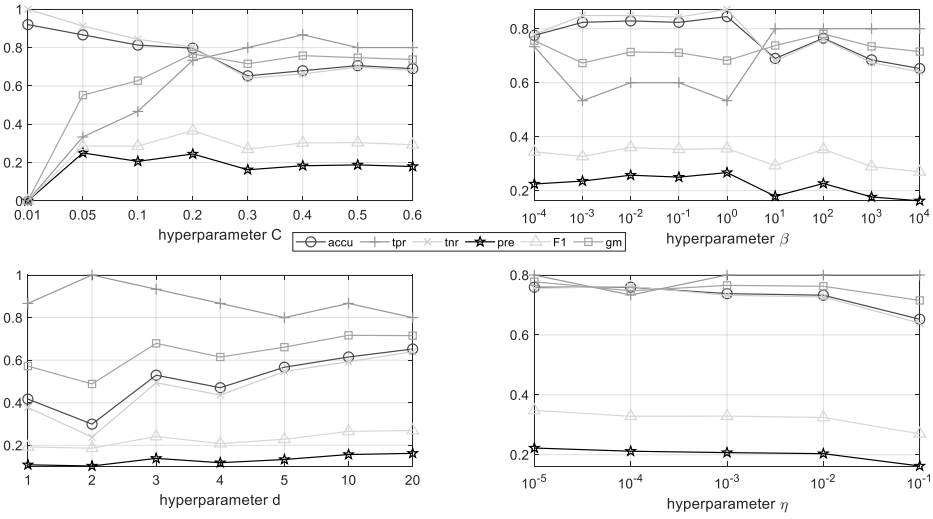

Figure 3: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_3 ds1$ on SPECTF heart dataset

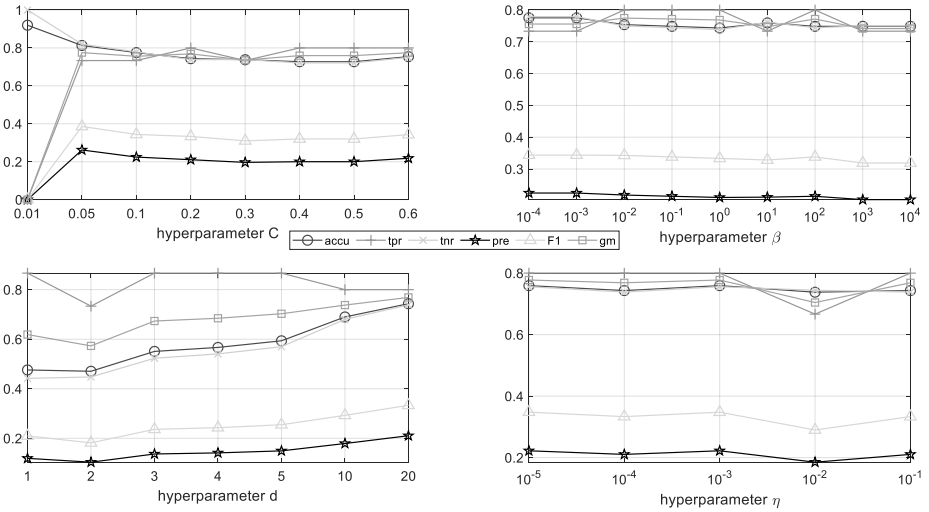

Figure 4: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_4 ds1$ on SPECTF heart dataset

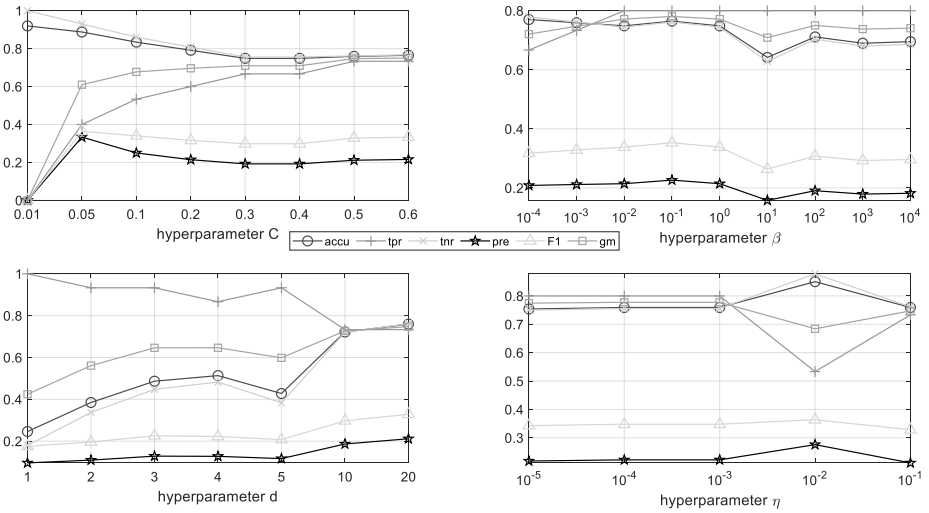

Figure 5: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_5 ds1$ on SPECTF heart dataset

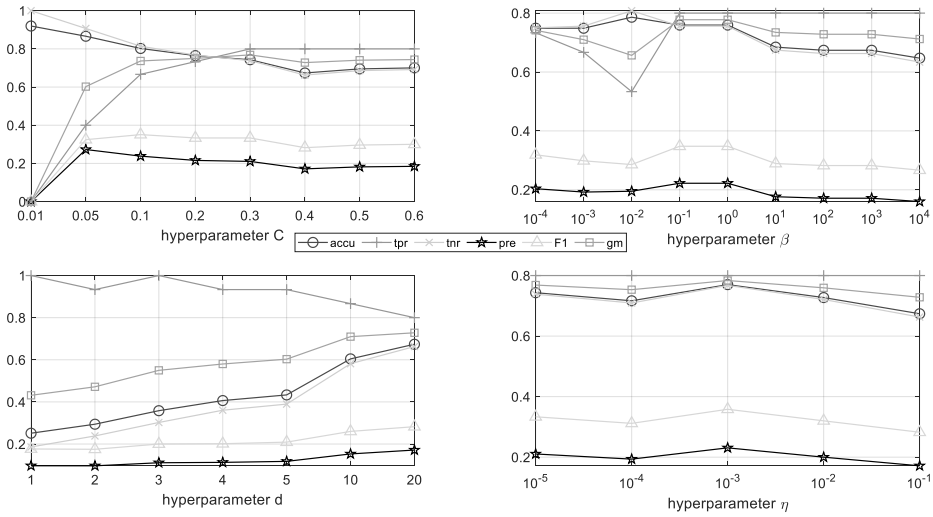

Figure 6: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_6 ds1$ on SPECTF heart dataset

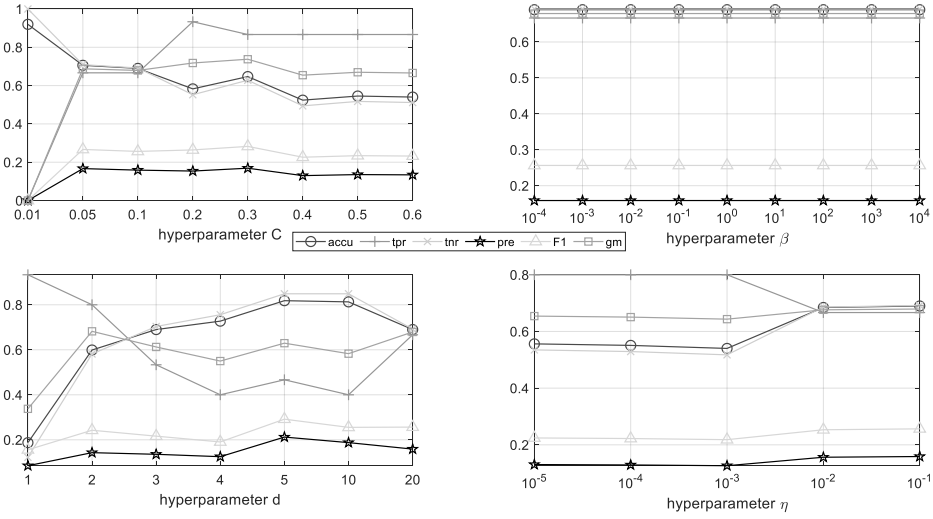

Figure 7: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_0 ds2$ on SPECTF heart dataset

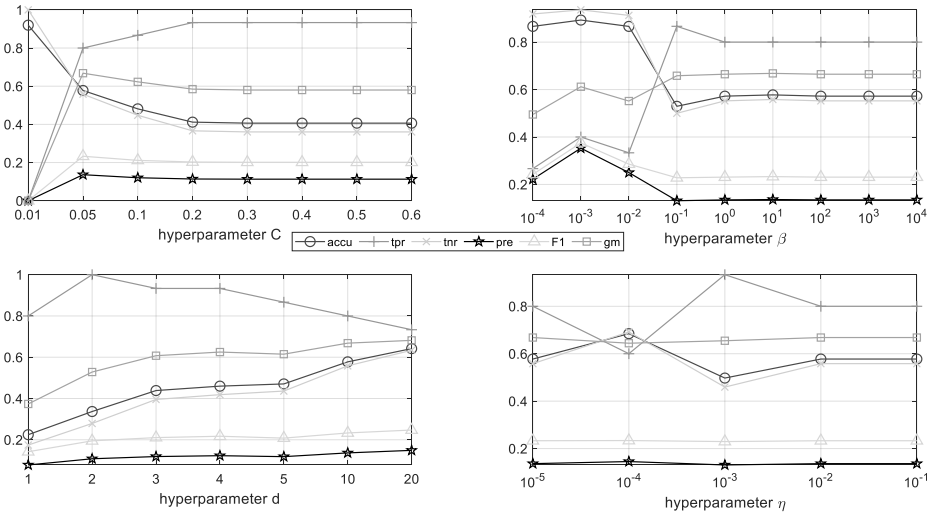

Figure 8: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_1 ds2$ on SPECTF heart dataset

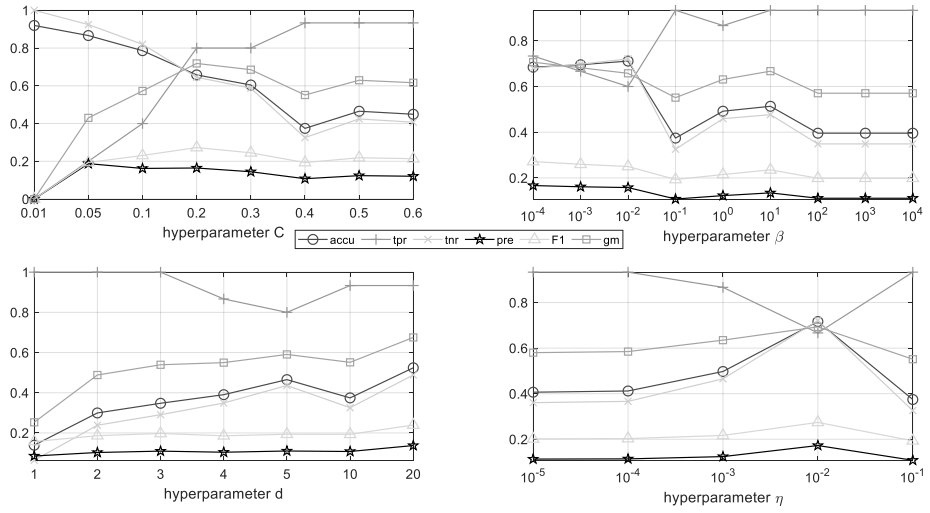

Figure 9: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_2 ds2$ on SPECTF heart dataset

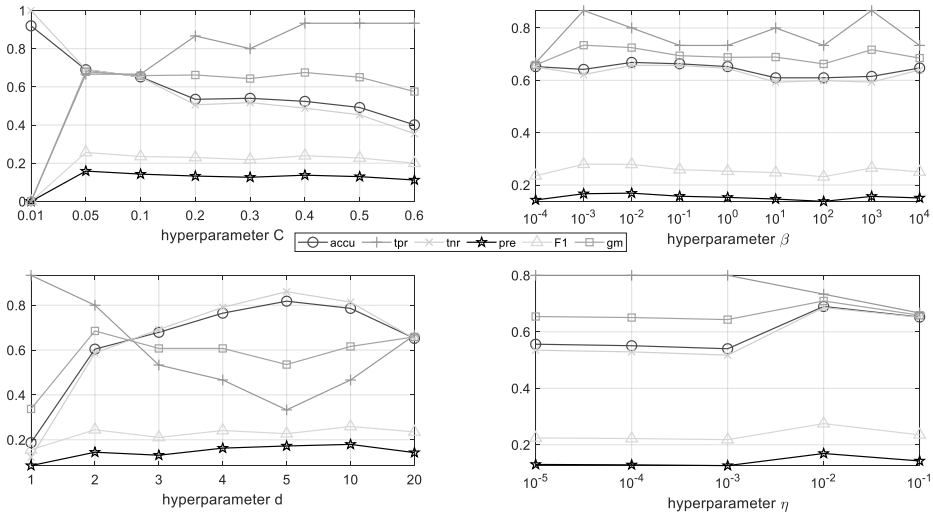

Figure 10: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_3 ds2$ on SPECTF heart dataset

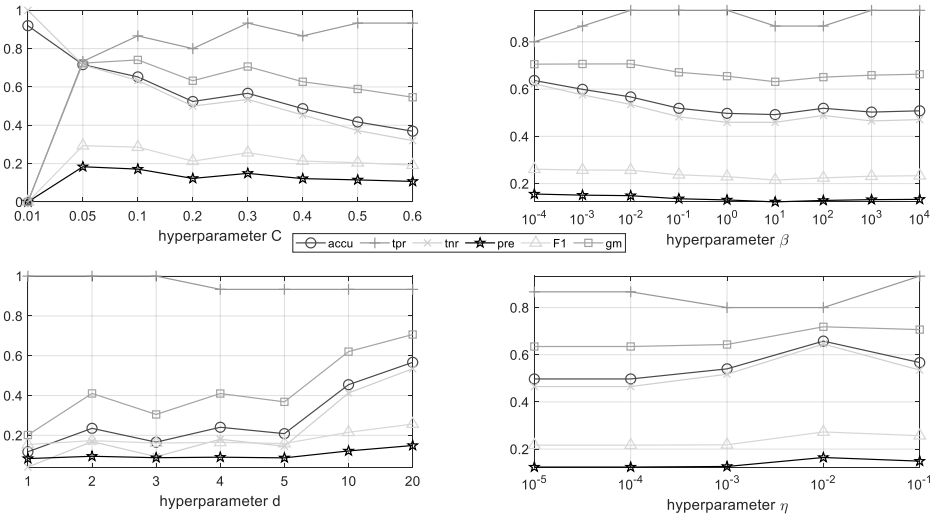

Figure 11: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_4 ds2$ on SPECTF heart dataset

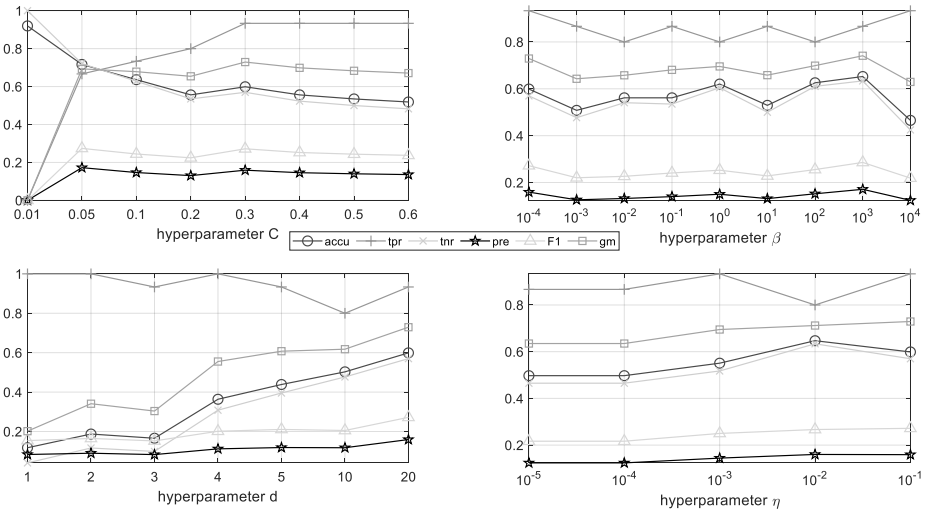

Figure 12: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_5 ds2$ on SPECTF heart dataset

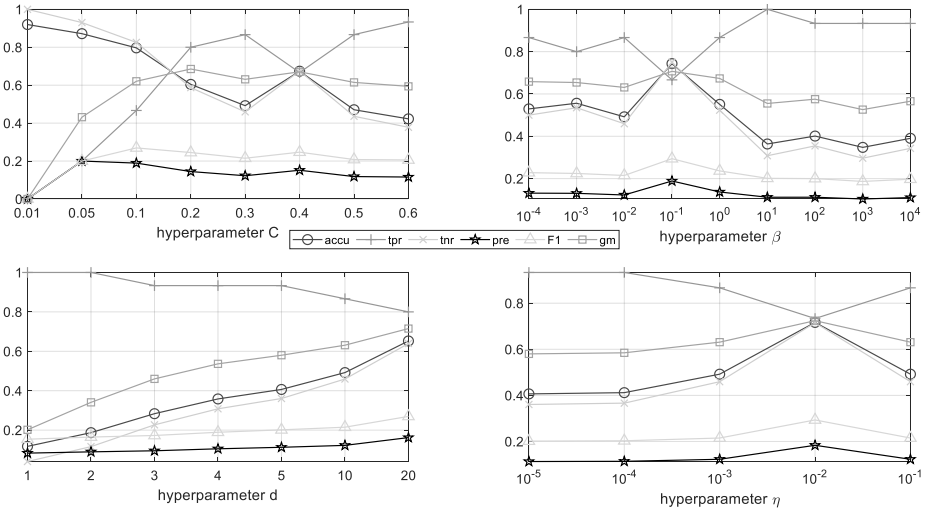

Figure 13: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_6 ds2$ on SPECTF heart dataset

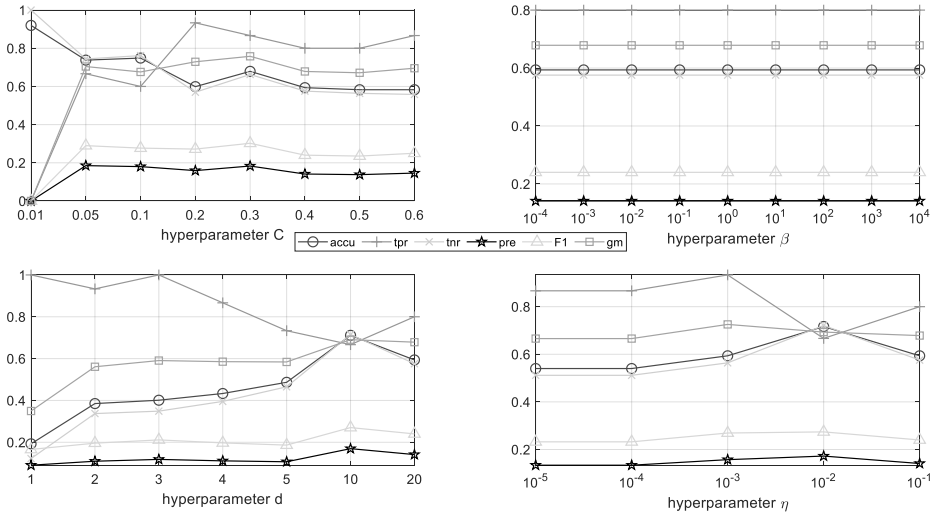

Figure 14: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_0 ds3$ on SPECTF heart dataset

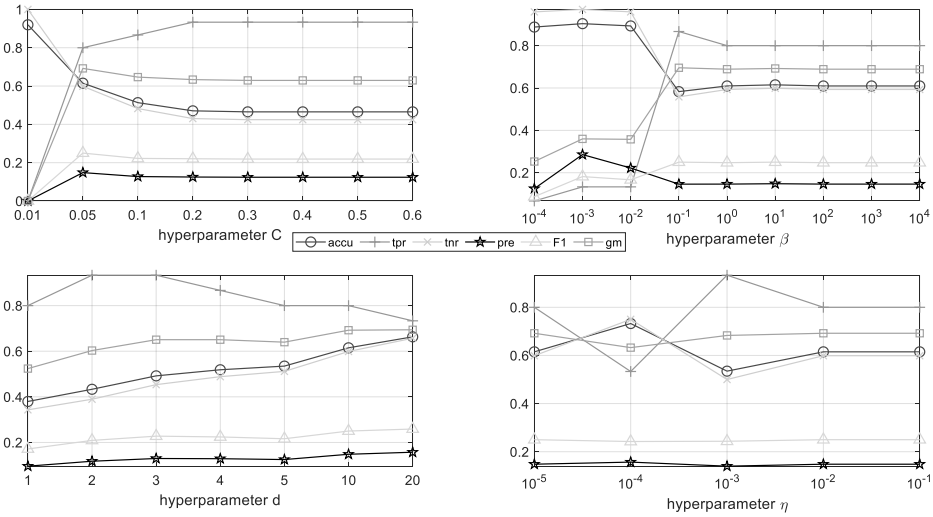

Figure 15: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_1 ds3$ on SPECTF heart dataset

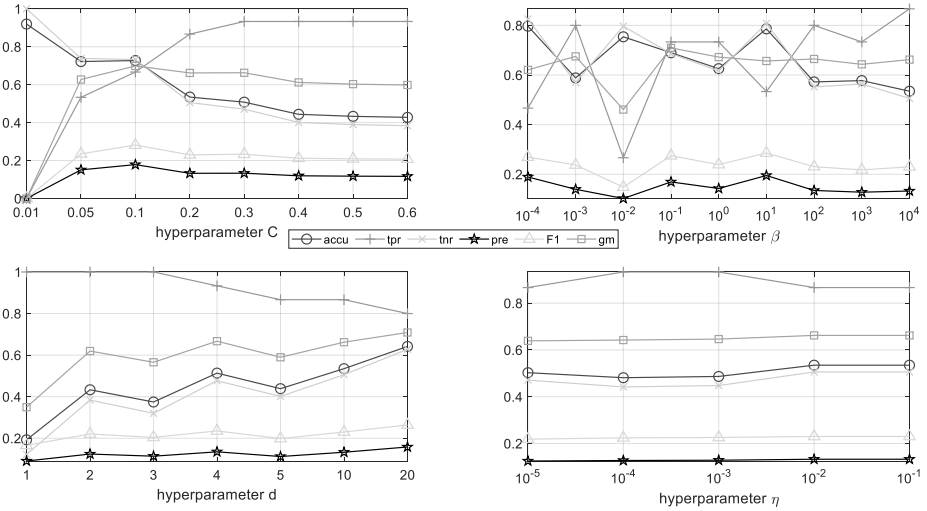

Figure 16: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_2 ds3$ on SPECTF heart dataset

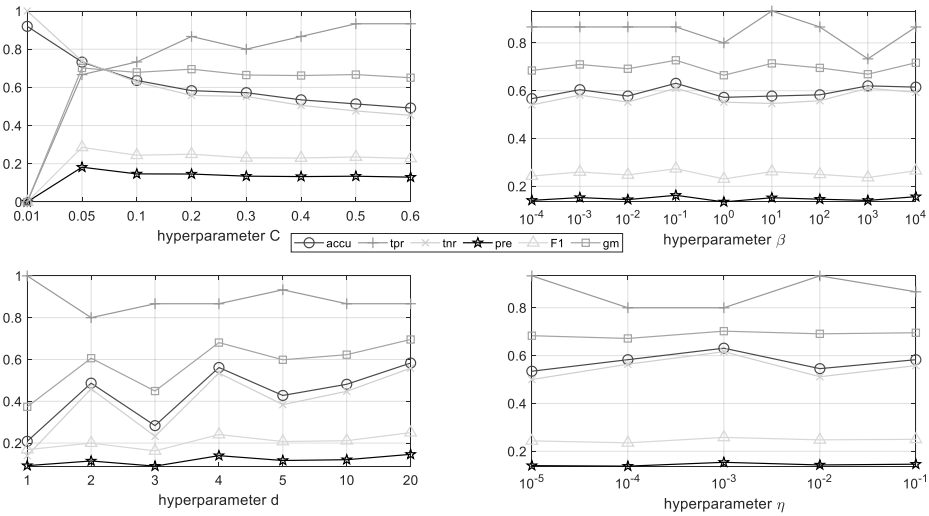

Figure 17: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_3 ds3$ on SPECTF heart dataset

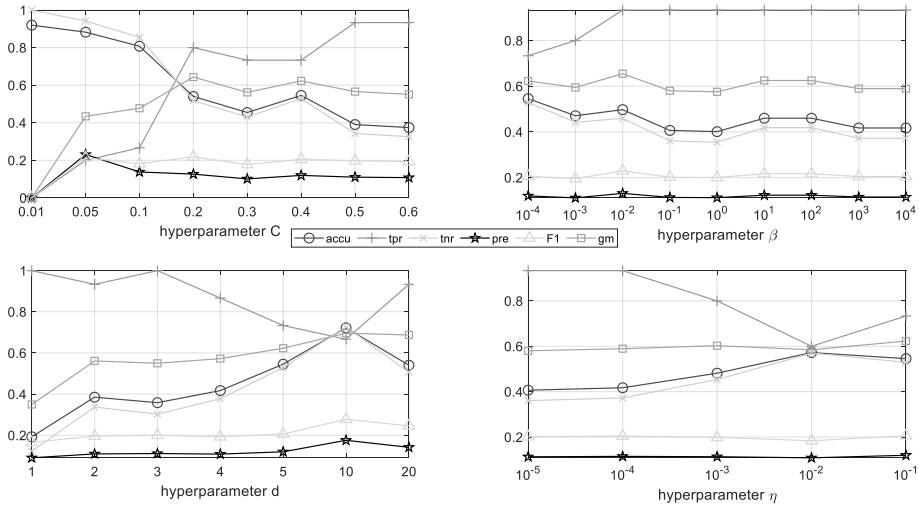

Figure 18: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_4 ds3$ on SPECTF heart dataset

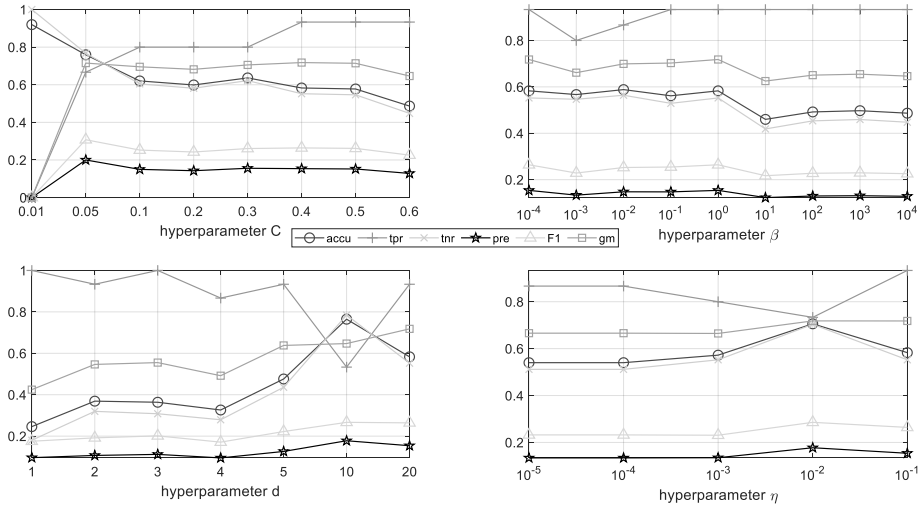

Figure 19: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_5 ds3$ on SPECTF heart dataset

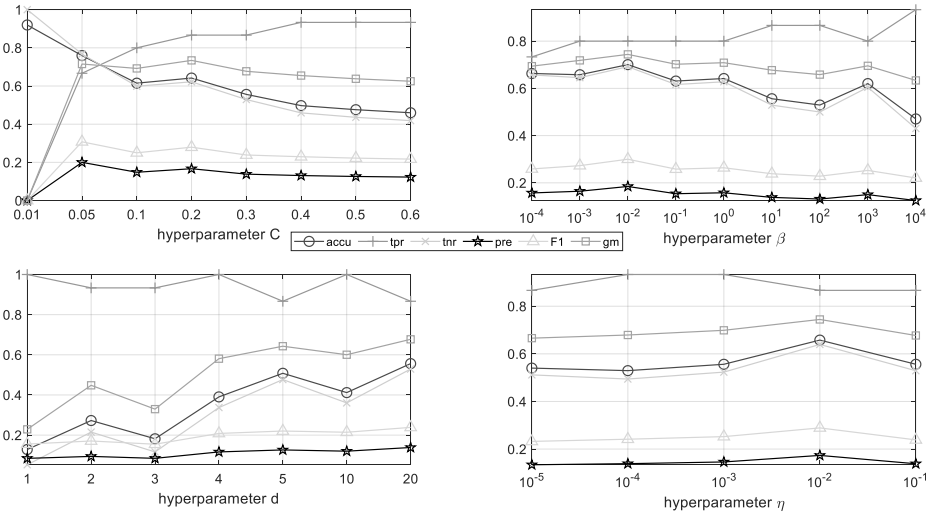

Figure 20: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_6 ds3$ on SPECTF heart dataset

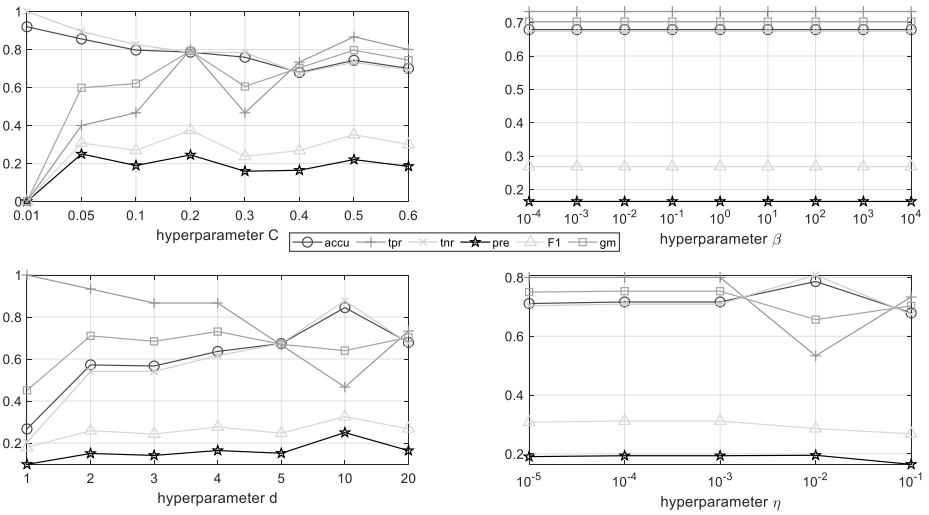

Figure 21: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_0 ds4$ on SPECTF heart dataset

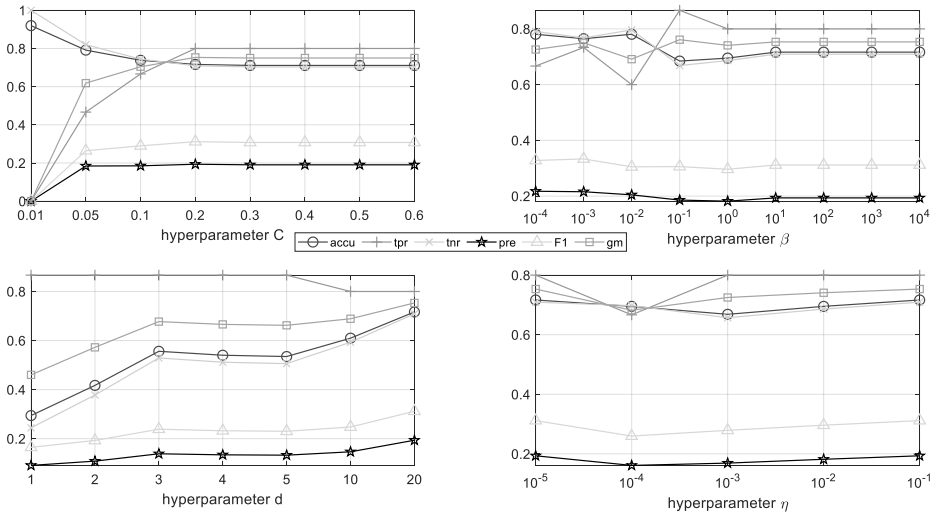

Figure 22: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_1 ds4$ on SPECTF heart dataset

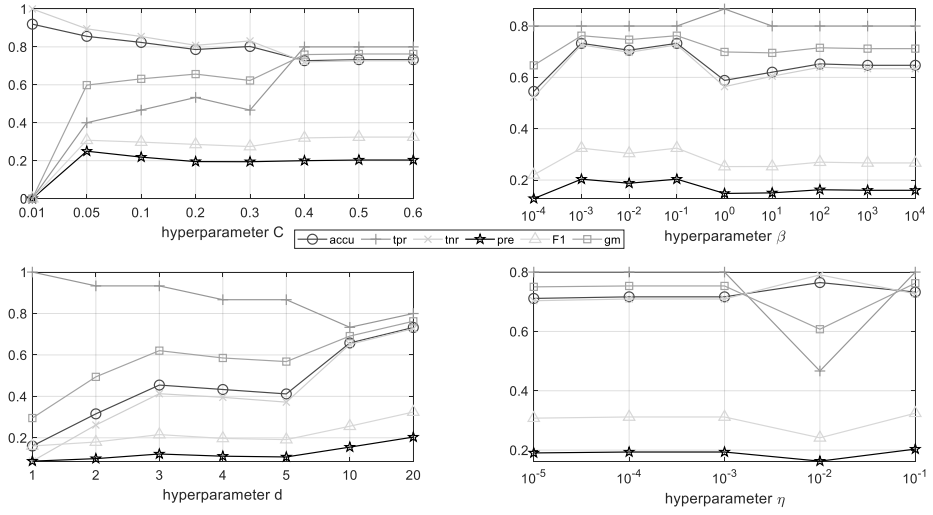

Figure 23: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_2 ds4$ on SPECTF heart dataset

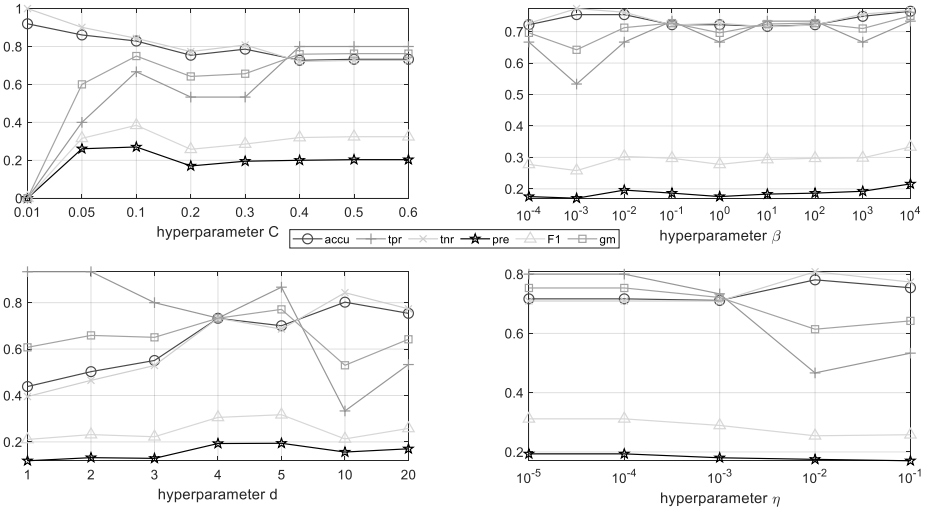

Figure 24: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_3 ds4$ on SPECTF heart dataset

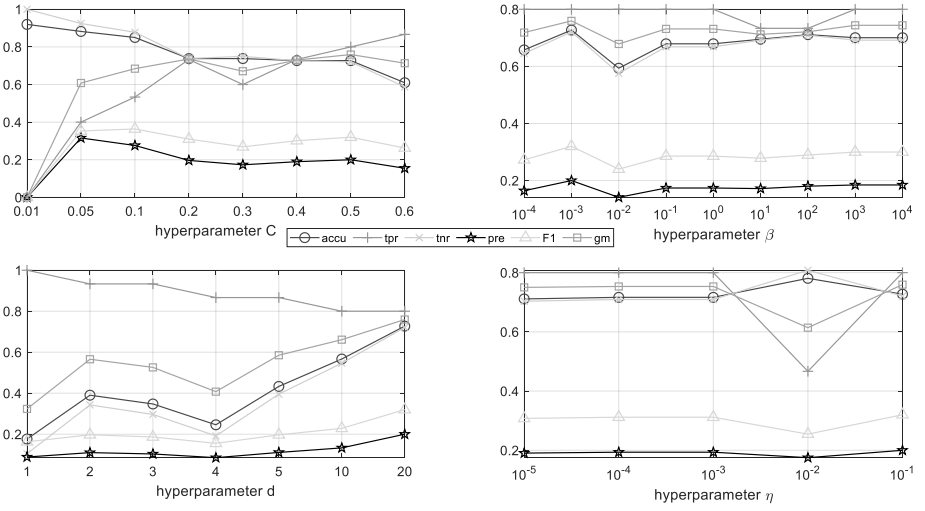

Figure 25: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_4 ds4$ on SPECTF heart dataset

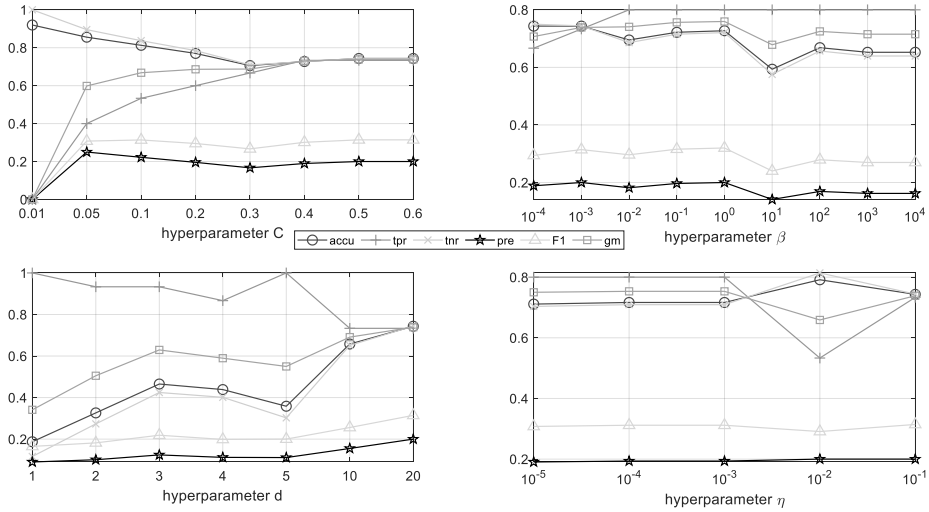

Figure 26: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_5 ds4$ on SPECTF heart dataset

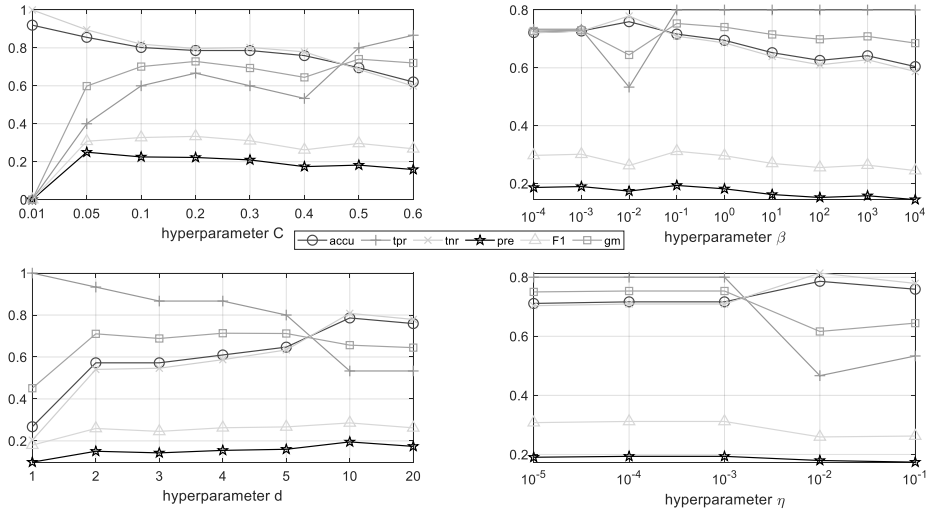

Figure 27: Hyperparameter sensitivity analysis for linear MS-SVDD $\omega_6 ds4$ on SPECTF heart dataset

PUBLICATION IV

Graph-Embedded Subspace Support Vector Data Description

F. Sohrab, A. Iosifidis, M. Gabbouj and J. Raitoharju

arXiv:2104.14370.

Graph-Embedded Subspace Support Vector Data Description

Fahad Sohrab[a,*], Alexandros Iosifidis[b], Moncef Gabbouj[a], Jenni Raitoharju[c]

[a]*Faculty of Information Technology and Communication Sciences, Tampere University, FI-33720 Tampere, Finland*
[b]*Department of Electrical and Computer Engineering, Aarhus University, DK-8200 Aarhus, Denmark*
[c]*Programme for Environmental Information, Finnish Environment Institute, FI-40500 Jyväskylä, Finland*

Abstract

In this paper, we propose a novel subspace learning framework for one-class classification. The proposed framework presents the problem in the form of graph embedding. It includes the previously proposed subspace one-class techniques as its special cases and provides further insight on what these techniques actually optimize. The framework allows to incorporate other meaningful optimization goals via the graph preserving criterion and reveals a spectral solution and a spectral regression-based solution as alternatives to the previously used gradient-based technique. We combine the subspace learning framework iteratively with Support Vector Data Description applied in the subspace to formulate Graph-Embedded Subspace Support Vector Data Description. We experimentally analyzed the performance of newly proposed different variants. We demonstrate improved performance against the baselines and the recently proposed subspace learning methods for one-class classification.

Keywords: One-Class Classification, Support Vector Data Description, Subspace Learning, Spectral Regression

1. Introduction

Dimensionality reduction has been an important and active research area in the field of machine learning and data science. The aim is to enhance the performance of a specific application by transforming the data from its original feature space to a lower-dimensional subspace. Dimensionality reduction has been used effectively as a tool in applications ranging from traditional data analysis and classification to many modern applications such as video analytics, recommendation system design, and detecting anomalies in computer and social networks [1].

The three main application domains of dimensionality reduction algorithms are feature matching, model interpretation, and data representation [2]. In feature matching, the aim is to find the similarity between two or more objects via a distance metric such as the Euclidean distance [3]. The model interpretation is enhanced by reducing the number of variables in the subspace by dimensionality reduction methods [4]. In data representation applications, dimensionality reduction methods are used to better represent the data in a lower dimensional space for the task at hand [5].

*corresponding author

Email addresses: fahad.sohrab@tuni.fi (Fahad Sohrab), ai@ece.au.dk (Alexandros Iosifidis), moncef.gabbouj@tuni.fi (Moncef Gabbouj), jenni.raitoharju@syke.fi (Jenni Raitoharju)

April 10, 2022

The approaches used for dimensionality reduction can be either supervised or unsupervised. In supervised learning, the algorithm relies mainly on the structure of data, and the mapping function is inferred from a set of labeled training samples. For example, Fisher's Linear Discriminant Analysis (LDA) is an example of a supervised method that exhibits good discrimination qualities. LDA maximizes the between-class scatter and minimizes the within-class scatter. In unsupervised learning, the algorithm does not leverage the information of pre-existing labels. For example, Principal Component Analysis (PCA) is a well-known unsupervised method for dimensionality reduction. PCA extracts the dominant features of a high-dimensional data and represents it by a small number of orthogonal basis vectors, i.e., the principal components. Numerous extensions and applications of PCA and LDA have been proposed in the literature [6, 7], and it has been shown that LDA can outperform PCA when the training data set is large [8]. However, for large-scale datasets, the computation and memory problems, particularly for the eigen-decomposition step of LDA, can be cumbersome. The spectral regression-based technique was proposed in [9] for speeding up the eigen-decomposition step of LDA. The spectral regression-based technique consolidates spectral graph analysis and regression to provide an efficient solution to LDA.

In general, the supervised dimensionality reduction approaches work better than unsupervised algorithms if sufficient data are available [2]. However, in real case scenarios, the labeled data may be scarce, noisy, or expensive to collect. In such situations, semi-supervised learning algorithms are preferred [10]. Semi-supervised learning mitigates the necessity for labeled data by allowing a model to leverage unlabeled data. Semi-supervised algorithms can extend the learning strategies of either supervised or unsupervised learning algorithms. If the data are available from only one class during the training, one-class classification algorithms are used to determine the predictive model [11]. In one-class classification, the decision function is inferred using training data from a single class only [12]. The class used to obtain the data description is referred to as the positive class, while all other classes are referred to as the negative class.

One-class classifiers have been extensively studied and improved for several technology-driven applications [13]. One-class classification techniques are found suitable for a specific target class detection in applications such as document classification [14], disease diagnosis [15], fraud detection [16], rare species identification [17], intrusion detection [18], or novelty detection [19]. Figure 1 depicts the basic idea of one-class classification.

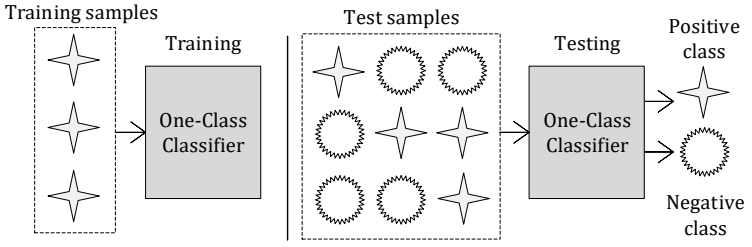

Figure 1: In one-class classification, a data model is learned by using samples of a positive class only. During inference, the model is used to detect objects also from the negative class.

Most one-class classification techniques operate in the original feature space and suffer from

the curse of dimensionality [20]. In this paper, we propose a general subspace learning framework for one-class classification. We pose the subspace learning for one-class classification as a graph embedding problem. We show that the previously proposed subspace one-class techniques can be reformulated through the proposed framework, while the framework brings more insight into their optimization process. The framework also allows to integrate other data relations to the optimization process and highlights the similarities to other subspace learning techniques. The framework motivates a novel spectral solution as well as a spectral regression-based solution as alternatives to the previously used gradient-based approach. Finally we integrate the subspace learning framework with the Support Vector Data Description (SVDD) applied in the subspace into an iterative Graph-Embedded Subspace Support Vector Data Description (GESSVDD) method.

The rest of the paper is organized as follows. In Section 2, we review the related work. In Section 3, we formulate the proposed framework, describe the full GESSVDD algorithm, and discuss the new insights obtained from the framework. Details of the experiments and the results are provided in Section 4. We finally deduce the conclusions in Section 5.

2. Related work and background

In this work, we focus on support vector (SV)-based one-class classification methods, which form a decision boundary represented by so-called support vectors by solving an optimization problem. The support vectors are selected from the training data points to define the boundary maximizing the considered criterion uniquely. One-class Support Vector Machine (OCSVM) [21] and SVDD [22] are classic examples of SV-based one-class classification methods. OCSVM constructs a hyperplane that separates the positive class by maximizing the distance of the hyperplane from the origin. In SVDD, a hypersphere with minimum volume is formed around the positive class. Numerous extensions of OCSVM and SVDD have been proposed in the literature [23, 24]. Traditionally, the SV-based one-class classification models data in the initially given feature space, but we have recently proposed one-class classification algorithms operating in an optimized lower-dimensional subspace [25, 26].

2.1. Support Vector Data Description

SVDD [22] finds a hyperspherical boundary around the positive class data in the original feature space by minimizing the volume of the hypersphere. Let us denote the training samples to be encapsulated inside a closed boundary by a matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N], \mathbf{x}_i \in \mathbb{R}^D$, where N is total number of samples and D is the dimensionality of data. The optimization problem of SVDD is formulated as follows:

$$\begin{aligned} \min \quad & F(R, \mathbf{a}) = R^2 + C \sum_{i=1}^{N} \xi_i \\ \text{s.t.} \quad & \|\mathbf{x}_i - \mathbf{a}\|_2^2 \leq R^2 + \xi_i, \\ & \xi_i \geq 0, \ \ \forall i \in \{1, \ldots, N\}, \end{aligned} \tag{1}$$

where R is the radius and $\mathbf{a} \in \mathbb{R}^D$ is the center of the hypersphere. The slack variables $\xi_i,\ i = 1, \ldots, N$ are introduced to allow the possibility of data being outliers and the hyperparameter

$C > 0$ controls the trade-off between the volume of the hypersphere and the amount of data outside the hypersphere. The Lagrangian of SVDD can be given as

$$L = \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^\intercal \mathbf{x}_i - \sum_{i}^{N} \sum_{j}^{N} \alpha_i \alpha_j \mathbf{x}_i^\intercal \mathbf{x}_j, \tag{2}$$

subject to the constraint that $0 \leq \alpha_i \leq C$ [22]. Maximizing (2) gives a set of α_i values corresponding to each data points. The data points with $0 < \alpha_i < C$ are called *support vectors* and define the data description. A test sample $\mathbf{x}_*$ is classified to the positive class if the distance of the test sample from the center of the hypersphere is smaller than or equal to the radius:

$$\|\mathbf{x}_* - \mathbf{a}\|_2 \leq R, \tag{3}$$

where R is the distance from the center of hypersphere to any sample with $0 < \alpha_i < C$.

2.2. Subspace Support Vector Data Description

SSVDD [26] optimizes a data mapping to a lower-dimensional subspace along with data description in the subspace. The optimization function is as follows:

$$\begin{aligned} \min \quad & F(R, \mathbf{a}) = R^2 + C \sum_{i=1}^{N} \xi_i \\ \text{s.t.} \quad & \|\mathbf{Q}\mathbf{x}_i - \mathbf{a}\|_2^2 \leq R^2 + \xi_i, \\ & \xi_i \geq 0, \ \ \forall i \in \{1, \ldots, N\}, \end{aligned} \tag{4}$$

where $\mathbf{Q} \in \mathbb{R}^{d \times D}$ is the projection matrix for mapping the data from original *D*-dimensional feature space to an optimized lower *d*-dimensional space. In SSVDD, an iterative process is followed: at each iteration, a set of α_i values is obtained by solving SVDD in the subspace, and then an augmented Lagrangian is optimized to update the projection matrix. The augmented Lagrangian is given as follows:

$$L = \sum_{i=1}^{N} \alpha_i \mathbf{x}_i^\intercal \mathbf{Q}^\intercal \mathbf{Q} \mathbf{x}_i - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \mathbf{x}_i^\intercal \mathbf{Q}^\intercal \mathbf{Q} \mathbf{x}_j \alpha_j + \beta \psi, \tag{5}$$

where ψ is an optional regularization term expressing the class variance in the lower *d*-dimensional space and β is the regularization parameter which controls the weight of ψ. The regularization term ψ has the following form:

$$\psi = \text{Tr}(\mathbf{Q}\mathbf{X}\boldsymbol{\lambda}\boldsymbol{\lambda}^\intercal \mathbf{X}^\intercal \mathbf{Q}^\intercal), \tag{6}$$

where Tr is the trace operator and different values of $\boldsymbol{\lambda}$ lead to different variants of SSVDD. The projection matrix $\mathbf{Q}$ is updated by using the gradient of (5), i.e.,

$$\mathbf{Q} \leftarrow \mathbf{Q} - \eta \Delta L, \tag{7}$$

where η is the learning rate parameter. The projection matrix is orthogonalized after every update.

Recently, Ellipsoidal Subspace Support Vector Data Description (ESSVDD) was proposed in [25]. ESSVDD considers the covariance of the data in the subspace and the optimization problem is given as

$$\begin{aligned} \min \quad & R^2 + C\sum_{i=1}^{N}\xi_i \\ \text{s.t.} \quad & (\mathbf{Q}\mathbf{x}_i - \mathbf{a})^\intercal \mathbf{E}^{-1}(\mathbf{Q}\mathbf{x}_i - \mathbf{a}) \leq R^2 + \xi_i, \\ & \xi_i \geq 0, \forall i \in \{1, \ldots, N\}, \end{aligned} \tag{8}$$

where

$$\mathbf{E} = \mathbf{Q}\mathbf{X}\mathbf{X}^\intercal\mathbf{Q}^\intercal \tag{9}$$

is a covariance matrix of the data in *d*-dimensional subspace. The rest of the ESSVDD solution follows the main principles of SSVDD explained above, while including the covariance matrix yields are more generalized solutions compared to SSVDD.

2.3. Graph embedding

Let $\mathcal{G} = \{\mathbf{X}, \mathbf{A}\}$ be an undirected weighted graph, where the data points in $\mathbf{X}$ are the graph nodes and $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the graph weight matrix that can measure different relations between the data points. The Laplacian matrix $\mathbf{L}$ of the graph and the diagonal degree matrix $\mathbf{D}$ are defined as follows:

$$\mathbf{L} = \mathbf{D} - \mathbf{A}, \quad [\mathbf{D}]_{ii} = \sum_{j\neq i}[\mathbf{A}]_{ij}, \forall i \in \{1, \ldots, N\}. \tag{10}$$

Graph embedding [27] was proposed as a general framework for encapsulating several subspace learning algorithms under the graph preserving criterion

$$\begin{aligned} \mathbf{Q}^* &= \underset{\text{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L_p}\mathbf{X}^\intercal\mathbf{Q}^\intercal)=m}{\arg\min} \sum_{i\neq j}(\mathbf{Q}\mathbf{x}_i - \mathbf{Q}\mathbf{x}_j)^2\mathbf{A}_{ij} \\ &= \underset{\text{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L}_p\mathbf{X}^\intercal\mathbf{Q}^\intercal)=m}{\arg\min} \text{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L}\mathbf{X}^\intercal\mathbf{Q}^\intercal), \\ &= \arg\min \frac{\text{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L}\mathbf{X}^\intercal\mathbf{Q}^\intercal)}{\text{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L}_p\mathbf{X}^\intercal\mathbf{Q}^\intercal)}, \end{aligned} \tag{11}$$

where $\mathbf{L}$ and $\mathbf{L}_p$ are the graph Laplacian matrices of the *intrinsic* and *penalty* graphs that correspond to data relations to be preserved or penalized, respectively. With different formulations of $\mathbf{L}$ and $\mathbf{L}_p$, (11) can represent different subspace learning algorithms. If there are no data-dependent penalty criteria to consider, the constraint $\text{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L_p}\mathbf{X}^\intercal\mathbf{Q}^\intercal) = m$ can be replaced with the orthogonality constraint $\text{Tr}(\mathbf{Q}\mathbf{Q}^\intercal) = m$.

The solution to the *trace ratio* optimization in (11) is typically approximated by the corresponding *ratio trace* problem

$$\mathbf{Q}^* = \arg\min \ \ \mathrm{Tr}\left((\mathbf{Q}\mathbf{X}\mathbf{L}_p\mathbf{X}^\intercal\mathbf{Q}^\intercal)^{-1}\mathbf{Q}\mathbf{X}\mathbf{L}\mathbf{X}^\intercal\mathbf{Q}^\intercal\right). \tag{12}$$

The solution to (12) can be obtained by solving the generalized eigenvalue value problem

$$\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{q} = \lambda\mathbf{X}\mathbf{L}_p\mathbf{X}^T\mathbf{q} \tag{13}$$

and keeping the eigenvectors corresponding to the d smallest non-zero eigenvalues as the rows of $\mathbf{Q}$.

The total scatter, within-class, and between-classes matrices commonly used in subspace learning can be expressed in the graph embedding framework as follows:

$$\mathbf{S}_t = \mathbf{X}\Big(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\intercal\Big)\mathbf{X}^\intercal = \mathbf{X}\mathbf{L}_t\mathbf{X}^\intercal \tag{14}$$

$$\mathbf{S}_w = \mathbf{X}\Big(\mathbf{I} - \sum_{c=1}^{\mathcal{C}}\frac{1}{N_c}\mathbf{1}_c\mathbf{1}_c^\intercal\Big)\mathbf{X}^\intercal = \mathbf{X}\mathbf{L}_w\mathbf{X}^\intercal \tag{15}$$

$$\mathbf{S}_b = \mathbf{X}\Big(\sum_{c=1}^{\mathcal{C}}N_c(\frac{1}{N_c}\mathbf{1}_c - \frac{1}{N}\mathbf{1})(\frac{1}{N_c}\mathbf{1}_c - \frac{1}{N}\mathbf{1})^\intercal\Big)\mathbf{X}^\intercal = \mathbf{X}\mathbf{L}_b\mathbf{X}^\intercal \tag{16}$$

where $\mathbf{I}$ is an identity matrix, $\mathbf{1}$ is a vector of ones, N_c is the total number of instances belonging to class c and $\mathbf{1}_c$ represents a vector with ones corresponding to instances which belongs to class c and zeros elsewhere. For centered data $\mathbf{S}_t$ reduces to $\mathbf{S}_t = \mathbf{X}\mathbf{X}^\intercal$. Using these Laplacians, LDA can be expressed in the graph embedding framework by setting $\mathbf{L} = \mathbf{L}_w$ and $\mathbf{L}_p = \mathbf{L}_b$ in (11). In a similar manner, PCA can be expressed in the graph embedding framework by setting $\mathbf{L} = \frac{1}{N}\mathbf{L}_t$, and replacing the constraint $\mathrm{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L_p}\mathbf{X}^\intercal\mathbf{Q}^\intercal) = m$ with the orthogonality constraint $\mathrm{Tr}(\mathbf{Q}\mathbf{Q}^\intercal) = m$. Since PCA seeks the projection directions with maximal variances, the criterion is maximized in the case of PCA.

Graph-Embedded Support Vector Data Description [23] was proposed to solve the following optimization problem

$$\begin{aligned} \min \quad & R^2 + C\sum_{i=1}^{N}\xi_i \\ \text{s.t.} \quad & (\mathbf{x}_i - \mathbf{a})^\intercal\mathbf{S}_x^{-1}(\mathbf{x}_i - \mathbf{a}) \le R^2 + \xi_i, \\ & \xi_i \ge 0, \forall i \in \{1, \ldots, N\}, \end{aligned} \tag{17}$$

where $\mathbf{S}_x = \mathbf{X}\mathbf{L}_x\mathbf{X}^\intercal$ and $\mathbf{L}_x$ is the graph Laplacian of any graph expressing geometric data relationship.

2.4. Spectral regression

Spectral regression [28] is an alternative way to solve the generalized eigen-decomposition in (13). If $\mathbf{X}^\intercal\mathbf{q} = \mathbf{t}$, and $\mathbf{t}$ and λ are an eigenvector and eigenvalue solving the eigenproblem

$$\mathbf{L}\mathbf{t} = \lambda\mathbf{L}_p\mathbf{t}, \tag{18}$$

$\mathbf{q}$ is the eigenvector of (13) with the same eigenvalue, because $\mathbf{X}\mathbf{L}\mathbf{X}^\intercal\mathbf{q} = \mathbf{X}\mathbf{L}\mathbf{t} = \lambda\mathbf{X}\mathbf{L}_p\mathbf{t} = \lambda\mathbf{X}\mathbf{L}_p\mathbf{X}^\intercal\mathbf{q}$. In order to find $\mathbf{Q}$, first the target vectors $\mathbf{t}$ can be obtained from (18) and then vectors $\mathbf{q}$ satisfying $\mathbf{X}^\intercal\mathbf{q} = \mathbf{t}$ found. An exact solution may not exists but it can be estimated using regularized least squares also known as ridge regression [29]:

$$\begin{aligned}\mathbf{q} &= \arg\min\left(\|\mathbf{X}^\intercal\mathbf{q} - \mathbf{t}\|^2 + \eta\|\mathbf{q}\|^2\right) \\ &= (\mathbf{X}\mathbf{X}^\intercal + \epsilon\mathbf{I})^{-1}\mathbf{X}\mathbf{t},\end{aligned} \tag{19}$$

where ϵ is a tiny constant. The above technique combines the spectral analysis and the regression, hence the approach is named as spectral regression. The main benefit of spectral regression approach is that most graph Laplacian are sparse and, thus, the approach bypasses the need of computing the eigen-decomposition of dense matrices. The least squares problem can be solved efficiently and, in some cases [9] it is also possible to compute the target vectors $\mathbf{t}$ directly without using eigen-decomposition at all, which makes the process much faster.

3. Graph Embedded Subspace Support Vector Data Description

In subspace one-class classification, the aim is to determine a projection matrix $\mathbf{Q} \in \mathbb{R}^{d\times D}$ for mapping data $\mathbf{X} \in \mathbb{R}^{D\times N}$ from the *D*-dimensional original feature space to a lower *d*-dimensional subspace optimized for one-class classification. In this work, we assume that the data has been centered by setting $\mathbf{X} \leftarrow \mathbf{X} - \boldsymbol{\mu}$, where $\boldsymbol{\mu}$ represents the mean of the training data. The mapped data in the subspace is represented by

$$\mathbf{y}_i = \mathbf{Q}\mathbf{x}_i, \;\; i = 1, \ldots, N. \tag{20}$$

After the transformation, the data is encapsulated inside a closed boundary to obtain an optimized data description in the subspace. In order to obtain a generalized solution, we consider the following optimization criterion:

$$\begin{aligned}\min \quad & R^2 + C\sum_{i=1}^{N}\xi_i \\ \text{s.t.} \quad & (\mathbf{Q}\mathbf{x}_i - \mathbf{a})^\intercal\mathbf{S}_Q^{-1}(\mathbf{Q}\mathbf{x}_i - \mathbf{a}) \le R^2 + \xi_i, \\ & \xi_i \ge 0, \forall i \in \{1, \ldots, N\},\end{aligned} \tag{21}$$

where the matrix $\mathbf{S}_Q$ encodes geometric data relationships in the subspace as

$$\mathbf{S}_Q = \mathbf{Q}\mathbf{X}\mathbf{L}_x\mathbf{X}^\intercal\mathbf{Q}^\intercal = \mathbf{Q}\mathbf{S}_x\mathbf{Q}^\intercal, \tag{22}$$

where $\mathbf{L}_x$ is a graph Laplacian. It can take different forms depending on the graph type used. By defining a new vector $\mathbf{u} = \mathbf{S}_Q^{-\frac{1}{2}}\mathbf{a}$, (21) can be written as

$$\begin{aligned} \min \quad & R^2 + C\sum_{i=1}^{N}\xi_i \\ \text{s.t.} \quad & \|\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i - \mathbf{u}\|_2^2 \leq R^2 + \xi_i, \\ & \xi_i \geq 0, \forall i \in \{1,\ldots,N\}. \end{aligned} \tag{23}$$

This shows that we can consider $\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}$ as a new projection matrix to a subspace, where SVDD is to be applied. We denote the mapped input vectors as $\mathbf{z}_i = \mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i$.

The constraints in (23) can be incorporated into a corresponding dual objective function by using Lagrange multipliers:

$$\begin{aligned} L = R^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\big(R^2 + \xi_i - \\ (\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i)^\intercal\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i + 2\mathbf{u}^\intercal\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i - \mathbf{u}^\intercal\mathbf{u}\big) - \sum_{i=1}^{N}\gamma_i\xi_i, \end{aligned} \tag{24}$$

where $\alpha_i \geq 0$ and $\gamma_i \geq 0$ are the Lagrange multipliers. The Lagrangian (24) should be minimized with respect to R, $\mathbf{u}$, and ξ_i and maximized with respect to Lagrange multipliers α_i and γ_i. By setting partial derivative to zero, we get

$$\frac{\partial L}{\partial R} = 0 \;\Rightarrow\; \sum_{i=1}^{N}\alpha_i = 1, \tag{25}$$

$$\frac{\partial L}{\partial \mathbf{u}} = 0 \;\Rightarrow\; \mathbf{u} = \sum_{i=1}^{N}\alpha_i\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i, \tag{26}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \;\Rightarrow\; C - \alpha_i - \xi_i = 0. \tag{27}$$

By substituting (25)-(27) into (24), we get

$$\begin{aligned} L &= \sum_{i=1}^{N}\alpha_i\mathbf{x}_i^\intercal\mathbf{Q}^\intercal\mathbf{S}_Q^{-1}\mathbf{Q}\mathbf{x}_i - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\mathbf{x}_i^\intercal\mathbf{Q}^\intercal\mathbf{S}_Q^{-1}\mathbf{Q}\mathbf{x}_j\alpha_j \\ &= \sum_{i=1}^{N}\alpha_i\mathbf{z}_i^\intercal\mathbf{z}_i - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\mathbf{z}_i^\intercal\mathbf{z}_j. \end{aligned} \tag{28}$$

Maximizing (28) corresponds to solving SVDD in the new subspace and will give us α_i values for

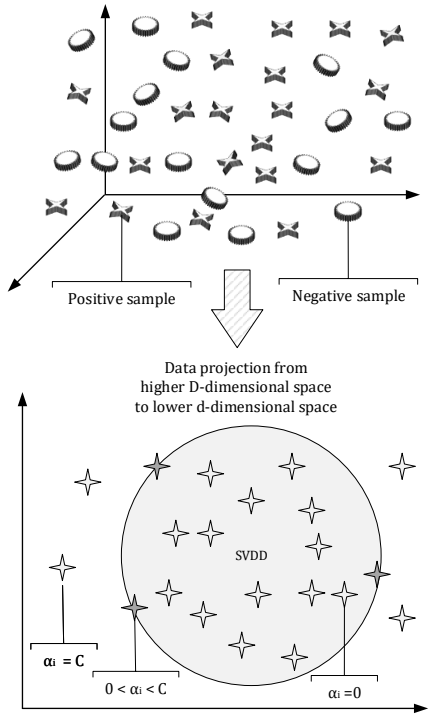

Figure 2: Depiction of data projection to a lower d-dimensional space optimized for one-class classification with corresponding α_i values

all instances, which will define their position in the data description. The samples in the subspace corresponding to values $0 < \alpha_i < C$ will lie on the boundary, while those outside the boundary will correspond to values $\alpha_i = C$. For the samples inside the closed boundary, the corresponding values of α_i will be equal to zero:

$$\|\mathbf{z}_i - \mathbf{u}\|_2 < R \rightarrow \alpha_i = 0, \gamma_i = 0, \tag{29}$$
$$\|\mathbf{z}_i - \mathbf{u}\|_2 = R \rightarrow 0 < \alpha_i < C, \gamma_i = 0, \tag{30}$$
$$\|\mathbf{z}_i - \mathbf{u}\|_2 > R \rightarrow \alpha_i = C, \gamma_i > 0. \tag{31}$$

Figure 2 depicts the idea of projecting data into an optimized subspace along with the positions of instances according to α values. The negative class samples are not considered in the process; hence, it is not guaranteed that they will be outside the obtained closed boundary.

The Lagrangian in (28) can be written in a trace form as

$$\begin{aligned} L &= \text{Tr}(\mathbf{S}_Q^{-1}\mathbf{Q}\mathbf{X}\mathbb{A}\mathbf{X}^\intercal\mathbf{Q}^\intercal) - \text{Tr}(\mathbf{S}_Q^{-1}\mathbf{Q}\mathbf{X}\boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal\mathbf{X}^\intercal\mathbf{Q}^\intercal) \\ &= \text{Tr}((\mathbf{Q}\mathbf{X}\mathbf{L}_x\mathbf{X}^\intercal\mathbf{Q}^\intercal)^{-1}\mathbf{Q}\mathbf{X}(\mathbb{A} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal)\mathbf{X}^\intercal\mathbf{Q}^\intercal), \end{aligned} \tag{32}$$

where the matrix $\mathbb{A} \in \mathbb{R}^{N \times N}$ contains α_i values in its diagonal and zeros elsewhere, $\boldsymbol{\alpha}$ is a vector of α_i values. Now by defining the matrices

$$\mathbf{L}_\alpha = \mathbb{A} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal \tag{33}$$
$$\mathbf{S}_\alpha = \mathbf{X}\mathbf{L}_\alpha\mathbf{X}^\intercal, \tag{34}$$

we can simplify (32) to

$$L = \text{Tr}\big((\mathbf{Q}\mathbf{S}_x\mathbf{Q}^\intercal)^{-1}\mathbf{Q}\mathbf{S}_\alpha\mathbf{Q}^\intercal\big). \tag{35}$$

We note that (35) is in a ratio trace form that resembles the trace ratio in (11). As mentioned, the trace ratio in (11) is typically approximated by the corresponding ratio trace to be able to solve the optimization using eigen-decomposition. We also note that $\mathbf{L}_\alpha$ is a graph Laplacian (see Section 3.2). Thus, we have presented the subspace learning for SVDD in the general graph embedding framework for subspace learning with its own fixed intrinsic graph $\mathbf{L}_\alpha$. Different graphs $\mathbf{L}_x$ create different variants and can be selected to enforce different constraints for the data. We will get back to different insights offered by the new framework in Section 3.2, but first we will introduce the full Graph-Embedded Subspace Support Vector Data Description (GESSVDD) algorithm.

3.1. GESSVDD algorithm

We can directly see from (35) that it can be minimized/maximized by solving the generalized eigenproblem in (13) and keeping the eigenvectors corresponding to the smallest/largest non-zero eigenvalues as projection vectors. We can also formulate a spectral regression-based solution as explained in Section 2.4. While earlier subspace SVDD variants [26, 25, 30] have only used gradient-based solution, we now have three alternatives: 1) gradient-based, 2) spectral, and 3) spectral regression-based updates. Furthermore, we can pick any desired graph as $\mathbf{L}_x$ and we note that it can be meaningful to also maximize (27) (see further discussion in Section 3.2). With this we can give the main GESSVDD algorithm in Algorithm 1 and the three update options in Sub-algorithms 1-3. The gradient of (32) used in the gradient-based update can be obtained using identity 126 in [31].

3.1.1. Non-linear data description

To obtain a non-linear mapping with the proposed method, we employ a non-linear projection trick (NPT) [32]. NPT is equivalent to applying the well-known kernel trick, while allows using the linear variant of the method. In NPT, the data $\mathbf{X}$ is mapped from the original *D*-dimensional space to $\mathbf{\Phi}$ in *F*-dimensional space as follows: The kernel matrix is obtained as

$$\mathbf{K}_{ij} = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2\sigma^2}\right), \tag{36}$$

where σ is a hyperparameter scaling the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$. The kernel matrix is centered as

$$\hat{\mathbf{K}} = \big(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\intercal\big)\mathbf{K}\big(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\intercal\big), \tag{37}$$

Algorithm 1: GESSVDD optimization

Input : $\mathbf{X}$, // Input data
$\mathbf{L}_x$ // Selected Laplacian
η, // Learning rate parameter
d, // Dimensionality of subspace
C, // Regularization parameter in SVDD
min or max // Either minimize or maximize the criterion

Output: $\mathbf{Q}$ // Projection matrix
R, // Radius of hypersphere
$\boldsymbol{\alpha}$ // Defines the data description

Initialize $\mathbf{Q}$ via PCA; // Select d-vectors corresponding to d largest eigenvalues.
Compute $\mathbf{S}_x = \mathbf{X}\mathbf{L}_x\mathbf{X}^\intercal$;

for $iter = 1 : max\_iter$ **do**
  Calculate $\mathbf{S}_{inv} = \mathbf{S}_Q^{-1} = (\mathbf{Q}\mathbf{S}_x\mathbf{Q}^\intercal)^{-1}$;
  Project data to subspace $\mathbf{z}_i = \mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i = (\mathbf{S}_{inv})^{\frac{1}{2}}\mathbf{Q}\mathbf{x}_i$;
  Calculate α values by maximizing $L = \sum_{i=1}^{N}\alpha_i\mathbf{z}_i^\intercal\mathbf{z}_i - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\mathbf{z}_i^\intercal\mathbf{z}_j$;
  Compute $\mathbf{L}_\alpha = \mathbb{A} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal$;
  if *gradient-based update*
  Call Sub-algorithm 1 to obtain $\mathbf{Q}$;
  elseif *spectral update*
  Call Sub-algorithm 2 to obtain $\mathbf{Q}$;
  elseif *spectral regression-based update*:
  Call Sub-algorithm 3 to obtain $\mathbf{Q}$;
  endif
  Orthogonalize $\mathbf{Q}$ using QR decomposition;
end
Project data to subspace $\mathbf{z}_i = \mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i$;
Calculate α values by maximizing $L = \sum_{i=1}^{N}\alpha_i\mathbf{z}_i^\intercal\mathbf{z}_i - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\mathbf{z}_i^\intercal\mathbf{z}_j$;
Compute center of data description in the subspace as $\mathbf{u} = \sum_{i=1}^{N}\alpha_i\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_i$;
Identify any support vector $\mathbf{s}$ having $0 < \alpha_s < C$;
Compute radius $R = \sqrt{(\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{s})^\intercal\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{s} - 2(\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{s})^\intercal\mathbf{u} + \mathbf{u}^\intercal\mathbf{u}}$;

Sub-algorithm 1: Gradient-based update

Input : $\mathbf{Q}$, $\mathbf{X}$, $\mathbf{S}_x$, $\mathbf{S}_{inv}$ $\mathbf{L}_\alpha$, η, *min/max* //Input from Algorithm 1
Output: $\mathbf{Q}$ //Return output to Algorithm 1

Compute $\mathbf{S}_\alpha = \mathbf{X}\mathbf{L}_\alpha\mathbf{X}^\intercal$;
Compute $\Delta L = 2\mathbf{S}_{inv}\mathbf{Q}\mathbf{S}_\alpha - 2\mathbf{S}_{inv}\mathbf{Q}\mathbf{S}_\alpha\mathbf{Q}^\intercal\mathbf{S}_{inv}\mathbf{Q}\mathbf{S}_x^\intercal$;
if *minimization*
Update $\mathbf{Q} \leftarrow \mathbf{Q} - \eta\Delta L$;
elseif *maximization*
Update $\mathbf{Q} \leftarrow \mathbf{Q} + \eta\Delta L$;

Sub-algorithm 2: Spectral update

Input : $\mathbf{X}$, $\mathbf{S}_x$, $\mathbf{L}_\alpha$, *min/max* //Input from Algorithm 1
Output: $\mathbf{Q}$ //Return output to Algorithm 1

Compute $\mathbf{S}_\alpha = \mathbf{X}\mathbf{L}_\alpha\mathbf{X}^\intercal$;
Solve generalized eigenvalue problem $\mathbf{S}_\alpha\mathbf{q} = \upsilon\mathbf{S}_x\mathbf{q}$;
if *minimization*
Select the eigenvectors corresponding to d smallest positive eigenvalues as rows of $\mathbf{Q}$;
elseif *maximization*
Select the eigenvectors corresponding to d largest eigenvalues as rows of $\mathbf{Q}$;

Sub-algorithm 3: Spectral regression-based update

Input : $\mathbf{X}$, $\mathbf{L}$, $\mathbf{L}_\alpha$, *min/max* //Input from Algorithm 1
Output: $\mathbf{Q}$ //Return output to Algorithm 1

Solve generalized eigenvalue problem: $\mathbf{L}_\alpha\mathbf{t} = \upsilon\mathbf{L}_x\mathbf{t}$;
if *minimization* then
Select the eigenvectors corresponding to d smallest positive eigenvalues as columns of $\mathbf{T}$;
elseif *maximization* then
Select the eigenvectors corresponding to d largest eigenvalues to as columns of $\mathbf{T}$;
Obtain $\mathbf{Q} = \mathbf{T}^\intercal\mathbf{X}^\intercal(\mathbf{X}\mathbf{X}^\intercal + \eta\mathbf{I})^{-1}$;

The centered kernel matrix $\hat{\mathbf{K}}$ is decomposed by using eigen-decomposition:

$$\hat{\mathbf{K}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\intercal, \tag{38}$$

where $\mathbf{\Lambda}$ contains the non-negative eigenvalues of $\hat{\mathbf{K}}$ in its diagonal and the columns of $\mathbf{U}$ contain the corresponding eigenvectors. Finally, the data representation $\mathbf{\Phi}$ is obtained as

$$\mathbf{\Phi} = \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}^\intercal. \tag{39}$$

Now we consider the obtained data transformation $\mathbf{\Phi}$ as the input to the linear algorithm, which is equivalent to applying the kernel method on $\mathbf{X}$.

3.1.2. Test phase

During testing, a test instance $\mathbf{x}_*$ is first mapped to an optimized d-dimensional space as

$$\mathbf{z}_* = \mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{x}_*. \tag{40}$$

The distance of the test instance to the center of the data description in the subspace is calculated. The test instance is classified as a positive instance if the distance is equal to or smaller than the radius:

$$\|\mathbf{z}_* - \mathbf{u}\|_2^2 \leq R^2, \tag{41}$$

where $\mathbf{u}$ is obtained by solving (26), and R^2 is calculated as

$$R^2 = (\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{s})^\intercal\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{s} - 2(\mathbf{S}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{s})^\intercal\mathbf{u} + \mathbf{u}^\intercal\mathbf{u}, \tag{42}$$

and $\mathbf{s}$ is any support vector with $0 < \alpha_s < C$. Otherwise, the test instance is classified as a negative instance.

In the non-linear approach, we first find the kernel vector

$$\mathbf{k}_* = \mathbf{\Phi}^\intercal\phi(\mathbf{x}_*). \tag{43}$$

The kernel vector is centered as

$$\hat{\mathbf{k}}_* = (\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\intercal)[\mathbf{k}_* - \frac{1}{N}\mathbf{K}\mathbf{1}]. \tag{44}$$

Finally, the NPT representation of the test instance is obtained as

$$\boldsymbol{\phi}_* = (\mathbf{\Phi}^T)^+\hat{\mathbf{k}}_*, \tag{45}$$

where $(.)^+$ is a pseudo-inverse. Now $\boldsymbol{\phi}_*$ is classified similar to the linear case, which is equivalent to applying a kernel method on $\mathbf{x}_*$.

3.1.3. Different variants

While any suitable graph can be used as $\mathbf{L}_x$, we list here some reasonable choices, which are also used in our experiments. In the first option, GESSVDD-0, we have no data-dependent constraint, but $\mathbf{S}_x$ in (35) is replaced by an identity matrix $\mathbf{I}$, which corresponds to the orthogonality constraint. In the second option GESSVDD-I, we use $\mathbf{L}_x = \mathbf{I}$. The third option GESSVDD-PCA uses the PCA graph: $\mathbf{S}_x = \frac{1}{N}\mathbf{S}_t$.

While we only have samples from the positive class, it may include several clusters. To consider this option, we cluster the positive training samples using k-means and then define options GESSVDD-Sw and GESSVDD-Sb with $\mathbf{S}_x = \mathbf{S}_w$ and $\mathbf{S}_x = \mathbf{S}_b$, respectively. Here, $\mathbf{S}_w$ and $\mathbf{S}_b$ are

solved as in (15) and (16), but c now refers to a cluster, not a class.

We also exploit the local geometric information by employing k-Nearest Neighbor (kNN) and setting

$$\mathbf{S}_x = \mathbf{S}_{kNN} = \mathbf{X}\Big(\mathbf{D}_{kNN} - \mathbf{A}_{kNN}\Big)\mathbf{X}^\intercal = \mathbf{X}\mathbf{L}_{kNN}\mathbf{X}^\intercal, \tag{46}$$

where $[\mathbf{A}]_{ij} = 1$, if $\mathbf{x}_i \in \mathcal{N}_j$ or $\mathbf{x}_j \in \mathcal{N}_i$ and 0, otherwise. $\mathcal{N}_i$ denotes the nearest neighbors of $\mathbf{x}_i$. This gives our last option denoted as GESSVDD-kNN.

Each of these options using different $\mathbf{S}_x$ can be solved using one of the update choices: gradient-based (GR), spectral ($\mathcal{S}$), or spectral regression-based (SR). Furthermore, in each case it is possible to either minimize or maximize the criterion in (35). To refer all these variants, we denote them as GESSVDD-0-GR-min, GESSVDD-0-GR-max, GESSVDD-0-$\mathcal{S}$-min and so on.

3.2. Framework analysis

Now we will get back to our main result, the general subspace learning framework for SVDD expressed as follows (repeated from (32)):

$$\mathrm{Tr}((\mathbf{Q}\mathbf{X}\mathbf{L}_x\mathbf{X}^\intercal\mathbf{Q}^\intercal)^{-1}\mathbf{Q}\mathbf{X}(\mathbb{A} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal)\mathbf{X}^\intercal\mathbf{Q}^\intercal), \tag{47}$$

where $\mathbf{L}_x$ can be used to enforce local/global data relations relevant for the task. Let us consider a graph with a weight matrix $[\mathbf{A}_\alpha]_{ij} = \alpha_i\alpha_j \ \forall i \neq j$ and $[\mathbf{A}_\alpha]_{ii} = 0$. With the constraint $\sum_{i=1}^N \alpha_i = 1$ (25), we get $[\mathbf{D}_\alpha]_{ii} = \sum_{j\neq i}[\mathbf{A}_\alpha]_{ij} = \sum_{j=1}^N \alpha_j\alpha_i - \alpha_i^2 = \alpha_i - \alpha_i^2$ and $\mathbf{L}_\alpha = \mathbf{D}_\alpha - \mathbf{A}_\alpha = \mathrm{diag}(\boldsymbol{\alpha}) - \boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal = \mathbb{A} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal$. This shows that $\mathbb{A} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal$ is a graph Laplacian of a graph that connects the samples i and j with a weight $\alpha_i\alpha_j$. As α_i values are zero for any samples inside the hypersphere, the resulting graph has only connections between the support vectors and outliers.

We also see that the graph of $\mathbf{L}_\alpha$ has a strong similarity with the PCA graph. PCA maximizes the variance of the samples to their center $\boldsymbol{\mu} = \frac{1}{N}\sum_{i=1}^N \mathbf{x}_i$, i.e.,

$$\begin{aligned}
\mathbf{S}_{pca} &= \frac{1}{N}\sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\intercal \\
&= \frac{1}{N}\sum_{i=1}^N (\mathbf{x}_i\mathbf{x}_i^\intercal - 2\mathbf{x}_i\boldsymbol{\mu}^\intercal + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal) \\
&= \frac{1}{N}\sum_{i=1}^N (\mathbf{x}_i\mathbf{x}_i^\intercal) - 2\boldsymbol{\mu}\boldsymbol{\mu}^\intercal + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal = \frac{1}{N}\sum_{i=1}^N (\mathbf{x}_i\mathbf{x}_i^\intercal) - \boldsymbol{\mu}\boldsymbol{\mu}^\intercal \\
&= \frac{1}{N}\mathbf{X}\mathbf{X}^\intercal - \frac{1}{N^2}\mathbf{X}\mathbf{1}\mathbf{1}^\intercal\mathbf{X}^\intercal = \frac{1}{N}\mathbf{X}(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\intercal)\mathbf{X}^\intercal \\
&= \mathbf{X}\mathbf{L}_{pca}\mathbf{X}^\intercal,
\end{aligned} \tag{48}$$

where $\mathbf{L}_{pca} = \mathbf{D}_{pca} - \mathbf{A}_{pca}$ and $[\mathbf{A}_{pca}]_{ij} = 1/N^2 \ \ \forall i \neq j$ and $[\mathbf{A}_{pca}]_{ii} = 0$. With an analogous derivation using the constraint $\sum_{i=1}^N \alpha_i = 1$, we see that $\mathbf{L}_\alpha$ represents the weighted variance of

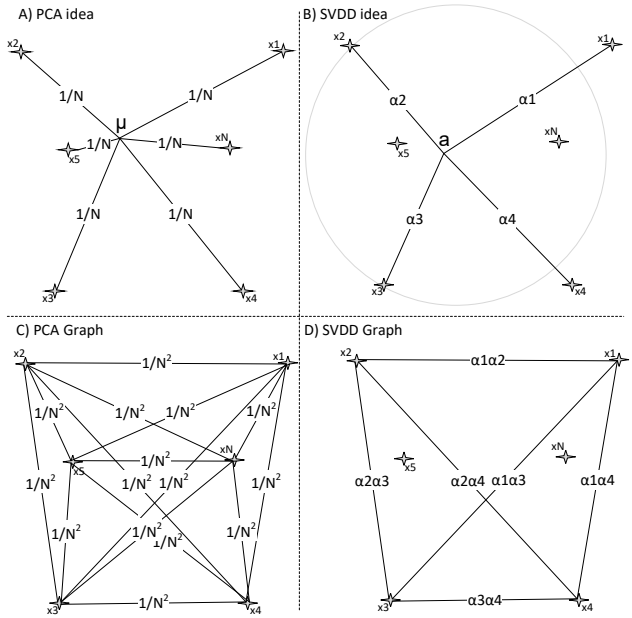

Figure 3: A) PCA considers the (unweighted) variance of all the points from the center $\boldsymbol{\mu}$. B) SVDD considers weighted variance of support vectors and outliers from the SVDD center $\mathbf{a}$. C) PCA graph is fully-connected with equal weights. D) SVDD graph is sparse (only the support vectors and outliers are connected) and has varying weights.

the support vectors and outliers to the center of SVDD defined as $\mathbf{a} = \sum_{i=1}^{N} \alpha_i \mathbf{x}_i$:

$$\begin{aligned}
\mathbf{S}_\alpha &= \sum_{i=1}^{N} (\mathbf{x}_i - \mathbf{a})(\mathbf{x}_i - \mathbf{a})^\intercal \alpha_i \\
&= \sum_{i=1}^{N} (\alpha_i \mathbf{x}_i \mathbf{x}_i^\intercal - 2\alpha_i \mathbf{x}_i \mathbf{a}^\intercal + \alpha_i \mathbf{a}\mathbf{a}^\intercal) \\
&= \sum_{i=1}^{N} (\alpha_i \mathbf{x}_i \mathbf{x}_i^\intercal) - 2\mathbf{a}\mathbf{a}^\intercal + \mathbf{a}\mathbf{a}^\intercal = \sum_{i=1}^{N} (\alpha_i \mathbf{x}_i \mathbf{x}_i^\intercal) - \mathbf{a}\mathbf{a}^\intercal \\
&= \mathbf{X}\mathrm{diag}(\boldsymbol{\alpha})\mathbf{X}^\intercal - \mathbf{X}\boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal\mathbf{X}^\intercal = \mathbf{X}(\mathbb{A} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal)\mathbf{X}^\intercal \\
&= \mathbf{X}\mathbf{L}_\alpha\mathbf{X}^\intercal.
\end{aligned} \tag{49}$$

The main idea of PCA and SVDD along with graphs $\mathbf{L}_{pca}$ and $\mathbf{L}_\alpha$ are illustrated in Figure 3.

By approximating the ratio trace in (47) with the corresponding trace ratio, we obtain a general

subspace learning graph embedding framework with the graph preserving criterion

$$\begin{aligned}\mathbf{Q}^* &= \underset{\mathrm{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L}_\mathbf{x}\mathbf{X}^\intercal\mathbf{Q}^\intercal)=m}{\arg\min} \sum_{i\neq j}(\mathbf{Q}\mathbf{x}_i - \mathbf{Q}\mathbf{x}_j)^2\alpha_i\alpha_j \\ &= \arg\min \frac{\mathrm{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L}_\alpha\mathbf{X}^\intercal\mathbf{Q}^\intercal)}{\mathrm{Tr}(\mathbf{Q}\mathbf{X}\mathbf{L}_x\mathbf{X}^\intercal\mathbf{Q}^\intercal)}.\end{aligned} \tag{50}$$

The criteria minimized in the previously proposed SSVDD [26, 30] and ESSVDD [25] are special cases of the proposed framework and correspond to variants GESSVDD-0-GR-min and GESSVDD-I-GR-min. We conclude that SSVDD minimizes the weighted variance of the support vectors and outliers, while having an orthogonality constraint. ESSVDD also minimizes the weighted variance of the support vectors and outliers, while simultaneously maximizing the total scatter of the centered inputs.

Previously, SSVDD and ESSVDD used the gradient-based update of the projection vector. It should be noted that while the gradient-based approach moves only a single step toward the optimum of (28), the spectral and spectral regression-based updates proposed in Section 3.1 directly jump to the optimum. This may help the overall iterative GESSVDD process converge faster, but it may also introduce some instability, because the objectives of the iteration steps may be contradictory.

To summarize, the new framework in (47) places subspace learning for SVDD in the general graph embedding framework with a fixed data-dependent SVDD graph $\mathbf{L}_\alpha$, which resembles PCA on the support vectors and outliers, and an additional constraint graph $\mathbf{L}_x$, which allows to incorporate other meaningful data relationships to the subspace learning step. When the overall objective function in (47) is minimized, $\mathbf{L}_\alpha$ represents data relationships to be minimized and $\mathbf{L}_x$ represents data relationships to be maximized. In the earlier works, the overall objective function has been minimized via gradient-descent. However, the new framework hints that it can also make sense to reverse the objective and maximize instead of minimizing. Also this approach has been previously followed in the literature in [33], where kernel PCA was successfully applied for novelty detection.

Intuitively, the original minimization of $\mathbf{L}_\alpha$ focuses on dimensions where the target class samples are the most similar, which indeed may help to discriminate the class from (unseen) other classes. On the other hand, from the similarity to PCA, we understand that these dimensions may be the dimensions that are not providing useful information in general (the corresponding PCA would discard them). Therefore, it is necessary to combine the criterion on $\mathbf{L}_\alpha$ with another criterion so that the combination can help to preserve the overall variance and minimize intra-class similarity simultaneously. In general, it may not be clear which criterion to minimize and which to maximize, but when considering the intra-cluster based graphs $\mathbf{L}_w$ and $\mathbf{L}_b$, an intuitive assumption is that within-cluster scatter $\mathbf{L}_w$ should be minimized (i.e., (47) maximized), while the between-cluster scatter $\mathbf{L}_b$ is more reasonable to be maximized (i.e., (47) minimized).

3.3. Complexity analysis

The proposed GESSVDD comprises three solutions: 1) gradient-based, 2) spectral, and 3) spectral regression-based updates. We first carry out the complexity analysis of the main algorithm

(1), which contains the shared steps for all the updates, and then proceed to the steps different in each solution update. The following steps contribute to the overall complexity of the Algorithm 1:

1. Initializing of the projection matrix $\mathbf{Q}$ via PCA comprises two steps, i.e., computing the covariance matrix and then the eigenvalue decomposition. The complexity of these steps is $\mathcal{O}\big(ND \times min(N, D)\big)$ and $\mathcal{O}\big(D^3\big)$, respectively.
2. Computing $\mathbf{S}_x = \mathbf{X}\mathbf{L}_x\mathbf{X}^\intercal$ for a given $\mathbf{L}_x$ has the complexity of $\mathcal{O}(DN^2 + ND^2)$.
3. Computing $\mathbf{S}_Q = \mathbf{Q}\mathbf{S}_x\mathbf{Q}^\intercal$ has the complexity of $\mathcal{O}(dD^2 + d^2D)$. Since, $D > d$, the complexity becomes $\mathcal{O}(dD^2)$.
4. Computing $\mathbf{S}_{inv}$ and the square-root of the matrix $\mathbf{S}_Q$ have the complexity of $\mathcal{O}(N^3)$.
5. SVDD has the complexity of $\mathcal{O}\big(N^3\big)$ for N data points [34].
6. The complexity of QR decomposition is $\mathcal{O}(dD^2)$ [35].

Dropping relatively lower computational costs and adding the rest, the complexity becomes $\mathcal{O}\big(N^3 + D^3\big)$. The total number of samples is assumed to be always greater than the dimensionality; hence the complexity becomes $\mathcal{O}\big(N^3\big)$. The complexity of each Sub-algorithm 1, 2, and 3 is $\mathcal{O}\big(N^3\big)$. We provide the details of complexity analysis of Sub-algorithms 1, 2, and 3 in Sections *1.1. Complexity analysis of gradient-based update*, *1.2. Complexity analysis of spectral-based update*, and *1.3. Complexity analysis of spectral regression-based update* respectively in the supplementary material. Adding the complexity of each Sub-algorithm to the main algorithm, the overall complexity remains at $\mathcal{O}(N^3)$, which is the same as for the original SVDD [34]. Moreover, in the non-linear case, the steps involved in NPT have the complexity of $\mathcal{O}(N^3)$; thus, the complexity in terms of the big $\mathcal{O}$ notation still stays as $\mathcal{O}(N^3)$.

4. Experiments

4.1. Datasets and experimental setup

To evaluate the proposed method's performance, we used nine different datasets. The datasets used in the experiments are Seeds, Qualitative bankruptcy, Somerville happiness, Liver, Iris, Ionosphere, Sonar, Heart (from UCI[1] machine learning repository) and MNIST [36] with original dimensionality D of 7, 6, 6, 6, 4, 34, 60, 13, and 784 respectively. MNIST has 10 classes, Seeds and Iris datasets are ternary, while the rest of the datasets are binary.

In Seeds dataset, the classes are named as Kama (S-K), Rosa (S-R), and Canadian (S-C) with 70 samples from each class. In Qualitative bankruptcy, the class labels are bankruptcy (QB-B) and non-bankruptcy (QB-N) with 107 and 143 samples, respectively. The Somerville happiness dataset contains 77 samples from the happy (SH-H) category and 66 from the unhappy (SH-U) category. Liver contain 145 samples from Disorder Present (DP) category and 200 samples from Disorder Absent (DA) category. Iris dataset contains 50 samples from each category of Setosa (I-S), Versicolor (S-VC), and Virginica (S-V). The Ionosphere dataset contains samples categorized as Bad (I-B) and Good (I-G). It contains 126 and 225 samples from bad and good categories, respectively. Sonar dataset has Rock (S-R) and Mines (S-M) as its two classes with 97 samples

[1] http://archive.ics.uci.edu/ml

from Rock and 111 samples from Mines category. Heart dataset contain 139 samples from disease present and 164 samples from disease absent categories, respectively.

MNIST dataset contains 5923, 6742, 5958, 6131, 5842, 5421, 5918, 6265, 5851, 5949 samples in the training set for classes 0-9, respectively. In the test set, it contains 980, 1135, 1032, 1010, 982, 892, 958, 1028, 974, and 1009 from corresponding classes (0-9). In our experiments, we select 10% of the data from MNIST while keeping the representation of each class in train and test set similar to the original train and test split in the dataset.

We manually created a corrupted version of the heart dataset to report the impact of noise. We added the noise in the manner described in [37]. The corrupted data were created by adding pseudo-random values drawn from the standard normal distribution to the features. We bound the range of added noise for the corresponding attribute to the maximum and minimum value of each feature of the target class in the training set.

We converted these datasets into one-class classification datasets by considering a single class at a time as the positive class and the rest as the negative class. For MNIST, the train and test sets are given, so we used the original train and test splits for the experiments. We divided the rest of the datasets into train and test sets by considering 70% of data as training data and the remaining 30% as test data. We selected the 70-30 splits randomly by keeping the representation of each class similar to the original dataset. We performed the 70-30% selection five times; hence we repeated the experiment 5 times for a single scenario where each class is considered a positive class. Note that at this point, both the training and test sets contained samples from both positive and negative classes. We did not use the negative samples in the training set in optimizing the models but only to select the hyperparameters by using five-fold cross-validation within the training set. To this end, four of the folds (only positive items) at a time were used for optimizing the model, and the fifth fold (both positive and negative items) was used to evaluate the performance. Finally, we used the best-performing hyperparameter values to optimize the model with the entire training set (only positive items) and reported the performance over the test set. We used a similar setup for all the competing methods. During the five-fold cross-validation over the training set, we found the best hyperparameters from the following values: $C \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$, $\sigma \in \{10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$, $d \in \{1, 2, 3, 4, 5, 10, 20\}$, $\eta \in \{10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$. The number of iterations for all the iterative methods was set to 5.

As our evaluation metrics, we report Geometric Mean $Gmean$, True Positive Rate (TPR), True Negative Rate (TNR), False Positive Rate (FPR), and False Negative Rate (FNR), where $TPR = \frac{TP}{P}$, $TNR = \frac{TN}{\mathcal{N}}$, $FPR = \frac{FP}{\mathcal{N}}$, and $FNR = \frac{FN}{P}$. TP, TN, FP, FN, P, $\mathcal{N}$ denote true positives, true negatives, false positives, false negatives, and number of positive samples, and number of negative samples, respectively. We use $Gmean$ as the main performance metric as it takes into account both TPR and TNR. We also report the standard deviations over the five data splittings.

For the proposed method, we consider all the variants introduced in Section 3.1.3: GESSVDD-0, GESSVDD-I, GESSVDD-PCA, GESSVDD-Sw, GESSVDD-Sb, and GESSVDD-kNN. For each, we consider all the alternative solutions (GR-gradient-based, $\mathcal{S}$-spectral, SR-spectral regression-based). The criterion in (35) is maximized and minimized in a separate set of experiments respectively for each variant and alternative solution. In order to construct the Laplacians $\mathbf{L}_w$ and $\mathbf{L}_b$, the number of clusters $\mathcal{C}$ was fixed to 5. Moreover, the numbers of neighbours for defining $\mathbf{L}_{kNN}$ was

also fixed to 5.

We also carried out sensitivity analysis for the model for the range of hyperparameters. We followed the approach mentioned in [30] for sensitivity analysis. In order to analyze the sensitivity of the model for the corresponding hyperparameter, we fix other hyperparameters to their optimal values found over the training set and record the performance with all the hyperparameter values considered in the given range.

To evaluate whether the observed differences between different methods are statistically significant, we follow the recommendations of [38]. We perform Wilcoxon Sign-Ranks test over the average results for the nine datasets to evaluate the pair-wise differences between the methods. The test ranks the differences between each pair of classifiers ignoring the signs and uses the ranks to determine value T as described, e.g., in [38]. Finally, the T value is compared to a critical value which depends on the number of datasets. In our experiments, we used 9 datasets, which means that the null hypothesis can be rejected at 0.05 significance level level if $T \leq 5$.

4.2. Experimental results and discussion

We report the results of the best performing linear and non-linear variants among the proposed variants compared against the previously proposed SSVDD [26] and ESSVDD [25], and the competing methods GESVM [23], GESVDD [23], OCSVM [21], SVDD [22], and ESVDD for all datasets in Table 1. In each experiment, a single class is used as a target class and the rest of the data as outliers. The average performance over each dataset is reported in the average (Av.) column. We report the average test results of different variants of the proposed framework over the five splittings of the Seeds, Qualitative bankruptcy, Somerville happiness, Iris, Ionosphere, and Sonar datasets in Table 2 for the non-linear data description, while the results over MNIST, Liver, and Heart datasets are reported in Section S2 of the supplementary material along with the results of all variants in the proposed framework in case of linear data description. We also provide TPR, TNR, FPR, and FNR results in S3 of the supplementary material. The corresponding standard deviations of $Gmean$ over five splits are provided in Section S4 and Section S5 for linear and non-linear cases, respectively, in the supplementary material. Implementations of the proposed framework are available online in GitHub[2].

From the experimental results comparing different variants of GESSVDD, we observe that in both linear and non-linear methods, the gradient-based solution performs better than the spectral and spectral regression-based solutions in the majority of the cases. The spectral approaches are typically more unstable over iterations as discussed in Section 3.2. When comparing the minimization/maximization, we see that our claim that $\mathbf{L}_w$ should be used with maximization and $\mathbf{L}_b$ with minimization seems to be valid in most cases. Overall, minimization typically leads to better results. Moreover, the performance of kNN graph is better than that of other variants for both min and max cases and for both linear and non-linear methods.

Overall in linear methods, it is noted that employing the kNN graph for encoding geometric information in the subspace yields better results also compared to the competing methods in the majority of the cases. Linear GESSVDD-kNN-GR-min variant performs best over 5 and second-best over 2 out of 9 datasets. For non-linear methods, the different variants of GESSVDD have

[2]https://github.com/fahadsohrab/gessvdd (Codes will be made public upon final acceptance of the manuscript)

Table 1: *Gmean* results for linear and non-linear data description over different datasets, selected variants from the proposed framework vs. other one-class classification methods

Dataset	Seeds				Qualitative bankruptcy			Somerville happiness			Liver		
Target class	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.	DP	DA	Av.
Linear													
GESSVDD-kNN-GR-min	0.83	**0.94**	**0.95**	**0.91**	0.80	**0.46**	**0.63**	0.44	0.44	0.44	0.45	0.38	0.42
GESSVDD-I-GR-min (ESSVDD)	**0.87**	0.92	0.90	0.90	0.90	0.12	0.51	**0.51**	0.39	0.45	0.35	0.38	0.37
GESSVDD-0-GR-min (SSVDD)	0.85	0.93	**0.95**	**0.91**	0.90	0.17	0.53	0.49	0.43	0.46	0.32	0.34	0.33
ESVDD	0.79	0.87	0.87	0.84	**0.96**	0.19	0.58	0.42	0.41	0.41	0.35	**0.40**	0.38
SVDD	0.85	0.92	0.94	0.90	0.94	0.00	0.47	0.41	0.36	0.39	**0.50**	0.39	**0.45**
OCSVM	0.48	0.69	0.45	0.54	0.37	0.41	0.39	0.45	**0.53**	**0.49**	0.40	0.36	0.38
Non-Linear													
GESSVDD-kNN-SR-max	0.86	0.92	**0.96**	0.91	0.81	**0.71**	**0.76**	0.47	0.47	0.47	0.41	0.42	0.41
GESSVDD-I-GR-min (ESSVDD)	0.83	0.91	0.90	0.88	0.92	0.28	0.60	**0.59**	0.39	0.49	0.40	0.49	0.45
GESSVDD-0-GR-min (SSVDD)	**0.87**	**0.94**	0.94	**0.92**	0.94	0.46	0.70	0.47	0.35	0.41	0.37	0.39	0.38
ESVDD	0.81	0.88	0.87	0.85	0.00	0.00	0.00	0.00	0.31	0.16	0.43	0.54	0.49
SVDD	0.85	0.91	0.95	0.90	0.33	0.28	0.31	0.40	0.32	0.36	0.49	0.40	0.45
OCSVM	0.47	0.60	0.45	0.51	0.36	0.58	0.47	0.47	0.49	0.48	0.27	0.08	0.17
GESVDD-PCA	0.85	0.93	0.93	0.90	0.94	0.28	0.61	0.50	0.48	0.49	**0.51**	0.49	0.50
GESVDD-Sw	0.82	0.93	0.93	0.89	0.94	0.28	0.61	0.49	**0.50**	0.49	**0.51**	0.52	0.51
GESVDD-kNN	0.84	0.92	0.94	0.90	0.84	0.31	0.57	0.50	0.45	0.47	**0.51**	0.52	0.52
GESVM-PCA	0.85	0.90	0.93	0.89	**0.95**	0.26	0.60	0.52	0.48	**0.50**	0.50	**0.55**	0.52
GESVM-Sw	0.85	0.90	0.91	0.89	0.93	0.20	0.57	0.55	0.41	0.48	0.50	0.51	0.51
GESVM-kNN	0.84	0.90	0.90	0.88	0.92	0.20	0.56	0.55	0.51	0.53	**0.51**	0.55	**0.53**

Dataset	Iris				Ionosphere			Sonar			Heart		
Target class	I-S	I-VC	S-V	Av.	I-B	I-G	Av.	S-R	S-M	Av.	DP	DA	Av.
Linear													
GESSVDD-kNN-GR-min	**0.97**	0.89	**0.91**	**0.92**	0.42	**0.92**	**0.67**	**0.54**	0.57	**0.56**	0.54	0.61	0.58
GESSVDD-I-GR-min (ESSVDD)	0.93	0.82	0.89	0.88	0.36	0.90	0.63	0.52	**0.58**	0.55	0.53	**0.69**	**0.61**
GESSVDD-0-GR-min (SSVDD)	0.96	**0.91**	0.90	**0.92**	0.12	0.78	0.45	0.51	0.55	0.53	**0.59**	0.62	**0.61**
ESVDD	0.89	0.85	0.86	0.87	0.33	0.88	0.61	0.00	0.03	0.02	0.56	0.62	0.59
SVDD	0.92	0.90	0.89	0.91	0.02	0.86	0.44	0.52	0.56	0.54	0.46	0.35	0.41
OCSVM	0.58	0.50	0.46	0.51	**0.49**	0.51	0.50	0.48	0.45	0.46	0.57	0.63	0.60
Non-Linear													
GESSVDD-kNN-SR-max	**0.94**	0.87	0.83	0.88	**0.67**	0.86	0.76	0.52	0.47	0.49	0.42	0.43	0.42
GESSVDD-I-GR-min (ESSVDD)	**0.94**	0.88	0.89	0.90	0.64	0.89	**0.77**	0.54	0.55	0.54	0.38	0.37	0.37
GESSVDD-0-GR-min (SSVDD)	**0.94**	0.92	**0.90**	**0.92**	0.40	0.89	0.65	0.48	0.47	0.47	0.53	0.49	0.51
ESVDD	0.68	0.84	0.83	0.78	0.37	0.88	0.63	0.55	0.52	0.53	0.34	0.27	0.31
SVDD	0.92	0.92	0.88	0.90	0.21	0.85	0.53	0.53	0.59	0.56	0.53	0.55	0.54
OCSVM	0.56	0.26	0.55	0.46	0.52	0.47	0.49	0.47	0.55	0.51	0.20	0.23	0.21
GESVDD-PCA	0.83	0.92	0.89	0.88	0.38	0.88	0.63	**0.55**	**0.60**	0.57	0.68	**0.74**	**0.71**
GESVDD-Sw	0.89	0.87	0.90	0.89	0.36	0.90	0.63	0.53	0.54	0.54	0.68	0.73	0.70
GESVDD-kNN	0.83	0.91	0.89	0.88	0.34	0.89	0.62	0.54	**0.60**	**0.57**	**0.70**	0.72	0.71
GESVM-PCA	0.90	0.90	**0.90**	0.90	0.38	**0.91**	0.64	0.52	0.61	0.57	0.66	0.71	0.68
GESVM-Sw	0.89	**0.93**	0.88	0.90	0.45	0.90	0.67	0.54	0.59	0.57	0.67	0.70	0.68
GESVM-kNN	0.89	0.89	0.89	0.89	0.41	0.88	0.65	0.54	0.58	0.56	0.67	0.72	0.70

Dataset	MNIST											***Wilcoxon test***
Target class	0	1	2	3	4	5	6	7	8	9	Av.	*T*
Linear												
GESSVDD-kNN-GR-min	0.40	0.84	0.33	0.47	0.60	0.38	0.69	0.53	0.51	0.58	0.53	-
GESSVDD-I-GR-min (ESSVDD)	0.38	0.83	0.31	0.46	0.47	0.34	0.62	0.65	0.40	0.50	0.50	6.5
GESSVDD-0-GR-min (SSVDD)	0.41	0.81	0.29	0.39	0.45	0.31	0.57	0.52	0.40	0.44	0.46	9.0
ESVDD	0.00	0.81	0.00	0.00	0.00	0.00	0.06	0.22	0.00	0.16	0.13	**1.0**
SVDD	0.47	0.55	**0.51**	0.50	0.51	**0.52**	0.45	0.57	0.49	0.51	0.51	**5.0**
OCSVM	**0.57**	**0.92**	0.47	**0.52**	**0.64**	0.41	**0.73**	**0.74**	**0.53**	**0.63**	**0.62**	8.0
Non-Linear												
GESSVDD-kNN-SR-max	0.38	0.53	0.16	0.34	0.49	0.46	0.48	0.43	0.31	0.50	0.41	-
GESSVDD-I-GR-min (ESSVDD)	0.36	0.34	0.18	0.09	0.19	0.52	0.46	0.43	0.36	0.21	0.31	17.5
GESSVDD-0-GR-min (SSVDD)	0.60	0.34	0.48	0.39	0.43	0.49	0.43	0.35	0.44	0.17	0.41	15.5
ESVDD	0.54	0.19	0.34	0.14	0.39	0.52	0.42	0.32	0.17	0.36	0.34	**5.0**
SVDD	0.15	0.05	0.63	0.14	0.12	0.17	0.11	0.13	0.13	0.13	0.18	15.0
OCSVM	0.59	0.69	0.56	0.46	0.61	0.64	0.66	0.56	0.53	0.66	0.60	6.0
GESVDD-PCA	**0.92**	**0.96**	**0.75**	0.74	0.84	**0.73**	0.86	0.86	0.73	0.85	0.82	-15.5
GESVDD-Sw	**0.92**	**0.96**	**0.75**	0.74	0.84	0.72	0.00	0.85	0.71	0.85	0.74	-15.5
GESVDD-kNN	0.91	**0.96**	**0.75**	0.74	0.84	0.72	0.86	0.86	0.73	0.85	0.82	-17.5
GESVM-PCA	0.90	0.95	**0.75**	0.74	**0.87**	0.71	0.89	**0.86**	**0.76**	**0.86**	**0.83**	-14.5
GESVM-Sw	0.90	0.95	**0.75**	0.74	0.85	0.66	0.87	0.84	0.75	0.85	0.82	-14.5
GESVM-kNN	0.90	0.95	0.74	**0.76**	**0.87**	0.71	**0.89**	0.85	0.73	0.85	0.82	-14.0

a more varying performance suggesting that finding a suitable graph for the task at hand may be more important. For comparisons, we report the results of a single variant GESSVDD-kNN-SR-max in the non-linear section of Table 1. It performs best over the Qualitative Bankruptcy and second-best over Seeds and Ionosphere datasets. For MNIST, we see that some other methods out-

Table 2: *Gmean* results for non-linear data description in the proposed framework

Dataset	Seeds				Qualitative bankruptcy			Somerville happiness		
Target class	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.75	0.91	0.88	0.85	0.58	0.47	0.52	0.37	0.33	0.35
GESSVDD-Sb-GR-max	0.83	0.72	0.89	0.81	0.82	0.36	0.59	0.41	0.40	0.41
GESSVDD-Sb-SR-max	0.77	0.81	0.93	0.83	0.56	0.29	0.43	0.44	0.30	0.37
GESSVDD-Sb-$\mathcal{S}$-min	0.79	0.90	0.84	0.84	0.61	0.42	0.52	0.43	0.39	0.41
GESSVDD-Sb-GR-min	0.83	0.61	0.91	0.78	0.80	0.50	0.65	0.50	0.42	0.46
GESSVDD-Sb-SR-min	0.72	0.86	0.92	0.83	0.56	0.47	0.52	0.45	0.30	0.37
GESSVDD-Sw-$\mathcal{S}$-max	0.85	0.89	0.92	0.89	0.84	0.25	0.55	0.53	0.37	0.45
GESSVDD-Sw-GR-max	0.88	0.86	0.91	0.88	0.81	0.12	0.46	0.47	0.43	0.45
GESSVDD-Sw-SR-max	0.82	0.89	0.89	0.87	0.85	0.53	0.69	0.51	0.36	0.44
GESSVDD-Sw-$\mathcal{S}$-min	0.78	0.89	0.92	0.86	0.86	0.33	0.60	0.55	**0.50**	**0.52**
GESSVDD-Sw-GR-min	**0.89**	0.94	0.92	**0.92**	0.77	0.03	0.40	0.49	0.42	0.45
GESSVDD-Sw-SR-min	0.81	0.87	0.91	0.87	0.93	0.71	**0.82**	0.55	0.44	0.49
GESSVDD-kNN-$\mathcal{S}$-max	0.87	0.90	0.90	0.89	0.73	0.62	0.68	0.51	0.34	0.42
GESSVDD-kNN-GR-max	0.82	0.91	0.89	0.87	0.88	0.28	0.58	0.55	0.41	0.48
GESSVDD-kNN-SR-max	0.86	0.92	**0.96**	0.91	0.81	0.71	0.76	0.47	0.47	0.47
GESSVDD-kNN-$\mathcal{S}$-min	0.87	0.88	0.94	0.89	0.80	**0.78**	0.79	0.49	0.43	0.46
GESSVDD-kNN-GR-min	0.84	0.94	0.91	0.90	0.85	0.38	0.61	**0.60**	0.39	0.49
GESSVDD-kNN-SR-min	0.87	0.89	0.94	0.90	0.76	0.75	0.76	0.46	0.38	0.42
GESSVDD-PCA-$\mathcal{S}$-max	0.83	0.91	0.94	0.89	0.60	0.67	0.63	0.51	0.46	0.48
GESSVDD-PCA-GR-max	0.78	0.90	0.90	0.86	0.90	0.14	0.52	0.56	0.37	0.46
GESSVDD-PCA-SR-max	0.83	0.74	0.94	0.84	0.90	0.40	0.65	0.48	0.46	0.47
GESSVDD-PCA-$\mathcal{S}$-min	0.85	0.94	0.94	0.91	0.85	0.48	0.67	0.55	0.38	0.47
GESSVDD-PCA-GR-min	0.86	0.86	0.94	0.89	0.93	0.17	0.55	0.53	0.40	0.46
GESSVDD-PCA-SR-min	0.84	0.90	0.94	0.89	0.93	0.61	0.77	0.51	0.42	0.47
GESSVDD-I-$\mathcal{S}$-max	0.85	0.93	0.76	0.84	0.83	0.39	0.61	0.43	0.37	0.40
GESSVDD-I-GR-max	0.83	0.91	0.91	0.88	0.91	0.13	0.52	0.52	0.41	0.47
GESSVDD-I-SR-max	0.85	0.93	0.94	0.91	0.86	0.54	0.70	0.49	0.44	0.46
GESSVDD-I-$\mathcal{S}$-min	0.85	0.94	0.93	0.91	0.85	0.42	0.64	0.47	0.42	0.44
GESSVDD-I-GR-min (ESSVDD)	0.83	0.91	0.90	0.88	0.92	0.28	0.60	0.59	0.39	0.49
GESSVDD-I-SR-min	0.85	**0.95**	0.93	0.91	0.84	0.49	0.67	0.54	0.38	0.46
GESSVDD-0-$\mathcal{S}$-max	0.85	0.92	0.93	0.90	0.70	0.44	0.57	0.39	0.44	0.41
GESSVDD-0-GR-max	0.85	0.93	0.94	0.90	0.93	0.47	0.70	0.37	0.46	0.42
GESSVDD-0-$\mathcal{S}$-min	0.86	0.91	0.92	0.89	0.70	0.47	0.59	0.38	0.38	0.38
GESSVDD-0-GR-min (SSVDD)	0.87	0.94	0.94	**0.92**	**0.94**	0.46	0.70	0.47	0.35	0.41

Dataset	Iris				Ionosphere			Sonar		
Target class	I-S	S-VC	S-V	Av.	I-B	I-G	Av.	S-R	S-M	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.94	0.88	0.77	0.86	0.57	0.63	0.60	0.42	0.38	0.40
GESSVDD-Sb-GR-max	0.95	0.83	0.90	0.89	0.42	0.87	0.64	0.50	0.40	0.45
GESSVDD-Sb-SR-max	0.80	0.85	0.85	0.83	0.53	0.42	0.47	0.40	0.39	0.40
GESSVDD-Sb-$\mathcal{S}$-min	0.93	0.90	0.87	0.90	0.46	0.52	0.49	0.42	0.47	0.45
GESSVDD-Sb-GR-min	0.95	0.86	0.90	0.90	0.30	0.87	0.59	0.50	0.49	0.50
GESSVDD-Sb-SR-min	0.82	0.82	0.74	0.79	0.51	0.42	0.46	0.40	0.34	0.37
GESSVDD-Sw-$\mathcal{S}$-max	0.85	0.91	0.84	0.87	0.49	0.86	0.67	0.54	0.38	0.46
GESSVDD-Sw-GR-max	0.94	0.90	0.89	0.91	0.48	0.89	0.68	0.52	0.46	0.49
GESSVDD-Sw-SR-max	0.97	0.86	0.85	0.89	0.35	0.87	0.61	0.37	0.50	0.44
GESSVDD-Sw-$\mathcal{S}$-min	0.74	0.89	0.85	0.83	0.46	0.86	0.66	0.51	0.50	0.51
GESSVDD-Sw-GR-min	0.95	0.86	0.83	0.88	0.61	0.87	0.74	0.50	0.44	0.47
GESSVDD-Sw-SR-min	0.97	0.86	0.86	0.89	0.32	0.87	0.59	0.37	0.41	0.39
GESSVDD-kNN-$\mathcal{S}$-max	0.94	0.88	0.83	0.88	**0.67**	0.87	0.77	0.51	0.48	0.50
GESSVDD-kNN-GR-max	0.94	0.92	0.88	0.91	0.65	0.90	**0.78**	0.54	0.49	0.51
GESSVDD-kNN-SR-max	0.94	0.87	0.83	0.88	**0.67**	0.86	0.76	0.52	0.47	0.49
GESSVDD-kNN-$\mathcal{S}$-min	0.94	0.88	0.83	0.89	0.38	0.87	0.62	**0.58**	0.43	0.50
GESSVDD-kNN-GR-min	0.92	0.88	**0.91**	0.90	0.59	**0.92**	0.75	0.54	**0.59**	**0.57**
GESSVDD-kNN-SR-min	0.94	0.87	0.82	0.88	0.49	0.86	0.67	0.54	0.56	0.55
GESSVDD-PCA-$\mathcal{S}$-max	0.94	**0.95**	0.87	**0.92**	0.37	0.87	0.62	0.42	0.36	0.39
GESSVDD-PCA-GR-max	0.92	0.87	0.85	0.88	0.57	0.89	0.73	0.54	0.43	0.48
GESSVDD-PCA-SR-max	0.94	0.92	0.88	0.91	0.34	0.84	0.59	0.29	0.52	0.41
GESSVDD-PCA-$\mathcal{S}$-min	0.94	0.92	0.80	0.89	0.37	0.87	0.62	0.50	0.51	0.50
GESSVDD-PCA-GR-min	0.92	0.73	0.85	0.83	0.59	0.91	0.75	0.53	0.53	0.53
GESSVDD-PCA-SR-min	0.94	0.92	0.83	0.90	0.30	0.87	0.58	0.56	0.49	0.53
GESSVDD-I-$\mathcal{S}$-max	**0.98**	0.92	0.84	**0.92**	0.25	0.88	0.56	0.48	0.36	0.42
GESSVDD-I-GR-max	0.95	0.86	0.88	0.90	0.63	0.88	0.76	0.56	0.54	0.55
GESSVDD-I-SR-max	**0.98**	0.93	0.84	**0.92**	0.24	0.88	0.56	0.49	0.51	0.50
GESSVDD-I-$\mathcal{S}$-min	0.95	0.93	0.84	0.91	0.22	0.88	0.55	0.50	0.52	0.51
GESSVDD-I-GR-min (ESSVDD)	0.94	0.88	0.89	0.90	0.64	0.89	0.77	0.54	0.55	0.54
GESSVDD-I-SR-min	0.95	0.93	0.85	0.91	0.29	0.88	0.58	0.53	0.54	0.53
GESSVDD-0-$\mathcal{S}$-max	0.94	0.91	0.90	0.91	0.61	0.63	0.62	0.47	0.58	0.52
GESSVDD-0-GR-max	0.96	0.91	0.88	**0.92**	0.44	0.82	0.63	0.49	0.54	0.51
GESSVDD-0-$\mathcal{S}$-min	0.95	0.88	0.90	0.91	0.41	0.73	0.57	0.51	0.40	0.45
GESSVDD-0-GR-min (SSVDD)	0.94	0.92	0.90	**0.92**	0.40	0.89	0.65	0.48	0.47	0.47

perform the proposed variants by a clear margin. As the maximum dimensionality allowed for our proposed methods in our experiments is 20, whereas the original dimensionality of MNIST data

Table 3: *Gmean* results for linear and non-linear data description over manually created corrupted versions of heart dataset, selected variants from the proposed framework vs. other one-class classification methods

Dataset	Heart Clean train set Corrupted test set			Heart Corrupted train set Clean test set			Heart Corrupted train set Corrupted test set		
Target class	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.
Linear									
GESSVDD-Sb-GR-max	0.17	0.22	0.20	0.24	**0.36**	0.30	0.36	0.30	0.33
GESSVDD-kNN-GR-min	0.09	0.00	0.04	0.32	0.08	0.20	0.35	0.44	0.39
GESSVDD-I-GR-min (ESSVDD)	0.00	0.00	0.00	0.00	0.00	0.00	0.35	0.39	0.37
GESSVDD-0-GR-min (SSVDD)	0.00	0.17	0.09	0.00	0.00	0.00	0.36	0.40	0.38
ESVDD	0.00	0.00	0.00	0.00	0.00	0.00	0.43	0.49	0.46
SVDD	**0.38**	**0.41**	**0.39**	**0.42**	0.27	**0.35**	**0.48**	**0.51**	**0.49**
OCSVM	0.00	0.00	0.00	0.00	0.00	0.00	0.34	0.41	0.38
Non-Linear									
GESSVDD-Sb-GR-max	**0.35**	0.23	**0.29**	**0.33**	0.31	**0.32**	0.27	0.45	0.36
GESSVDD-kNN-SR-max	0.04	0.03	0.03	0.00	0.00	0.00	0.47	0.40	0.44
GESSVDD-I-GR-min (ESSVDD)	0.37	0.15	0.26	0.00	0.11	0.06	0.38	0.46	0.42
GESSVDD-0-GR-min (SSVDD)	0.12	**0.24**	0.18	0.00	0.09	0.04	0.37	0.47	0.42
ESVDD	0.00	0.00	0.00	0.00	0.03	0.02	0.49	0.52	0.51
SVDD	0.07	0.15	0.11	0.22	0.07	0.15	0.47	0.45	0.46
OCSVM	0.00	0.00	0.00	0.00	0.07	0.03	**0.50**	0.49	0.50
GESVDD-PCA	0.00	0.00	0.00	0.00	0.00	0.00	0.51	**0.53**	**0.52**
GESVDD-Sw	0.00	0.00	0.00	0.00	0.00	0.00	0.53	0.52	**0.52**
GESVDD-kNN	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.52	0.50
GESVM-PCA	0.00	0.00	0.00	0.00	0.00	0.00	0.49	0.49	0.49
GESVM-Sw	0.00	0.00	0.00	0.00	0.05	0.03	**0.50**	0.48	0.49
GESVM-kNN	0.00	0.00	0.00	0.00	0.03	0.02	0.49	0.50	0.49

is 784, we can conclude that the reduction in dimensionality is likely too dramatic for preserving the significant information.

We applied Wilcoxon Sign-Ranks separately for linear and non-linear methods. We compared all other linear methods in Table 1 against our proposed GESSVDD-kNN-GR-min variant and all other non-linear methods Table 1 against our proposed GESSVDD-kNN-SR-max variant. We give the T-values in Table 1 and bold the values if they show that the difference between the methods is statistically significant at 0.05 significance level. Negative values indicate that the other method was performing better than our proposed variant GESSVDD-kNN-GR-min or GESSVDD-kNN-SR-max. We see that GESSVDD-kNN-GR-min outperforms ESVDD and SVDD in a statistically significant manner for linear data description and GESSVDD-kNN-SR-max outperforms ESVDD in a statistically significant manner for non-linear data description. All other differences are statistically insignificant. However, it should be noted that for individual datasets the differences in both ways can be still significant due to different reasons, such as our proposed variant failing with MNIST due to the drastic dimensionality reduction, and it cannot be concluded that the selection of the method is insignificant.

In evaluating the effect of added noise on the features of the heart dataset, it can be noticed that GESSVDD-Sb-GR-max performs second-best in the linear case when only the train or test set is corrupted. In the non-linear case of adding noise to either train or test set, GESSVDD-Sb-GR-max performs best on average. While the competing methods perform better than the proposed methods when both train and test datasets are corrupted, the competing methods underperform severely if the only train or test set is corrupted. There is not a single case where the proposed method would severely underperform. We report the performance of the selected variants of our method along with the competing methods in Table 3. We provide the $Gmean$ results for all proposed variants over the Heart dataset and its manually created corrupted versions in Section S2 and TPR, TNR FPR, and FNR results in Section S3 of the supplementary material.

We carried out a sensitivity analysis of different hyperparameters. Figure 4 shows the sensitiv-

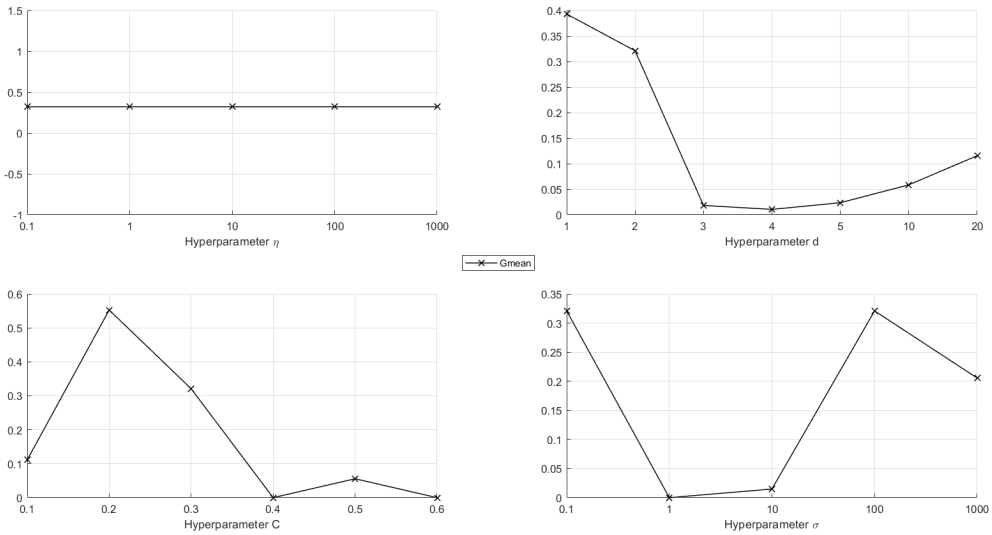

Figure 4: Sensitivity Analysis for non-linear GESSVDD-kNN-SR-max trained over MNIST dataset with target class 0

ity plot for non-linear GESSVDD-kNN-SR-max trained over MNIST dataset with target class 0. For all other variants, we provide the plots of sensitivity analysis in Section S6 of the supplementary material. We observe that the performance of GESSVDD-kNN-SR-max is not sensitive to the hyperparameter η. In the case of increasing the value of hyperparameter d, a sudden drop and then a steady rise in the performance is observed over the range of values. We also notice the poor performance of the model at higher values of hyperparameter C; moreover, a varying performance is noticed at different values for hyperparameter σ.

5. Conclusion

In this paper, we formulated subspace learning for one-class classification in the graph embedding framework and discussed the novel insights obtained from this formulation. In particular, we showed that subspace learning for SVDD applies a weighted PCA over the support vectors and outliers to define the projection matrix and we discussed how this information can be combined with other data relationships in the optimization process via an adaptable graph. We also formulated a novel Graph-Embedded Subspace Support Vector Data Description with gradient-based, spectral, and spectral regression-based solutions and different adaptable graphs. We reported the experimental results over nine different datasets by considering each class of a dataset as a target class at a time. The results showed that the proposed framework with the kNN graph as the adaptable graph had the best overall performance, while the gradient-based solution was more stable than the spectral and spectral regression-based solutions.

While the proposed framework showed promising results over different datasets and can be applied on different domain applications, there are some limitations that can be taken into account

in the future. The methods exploit only a single Laplacian $\mathbf{L}_x$ to enforce local/global data relations relevant to the task. This can be enhanced by exploiting multiple graphs by combining the geometric data relationships using a weight parameter.

In the future, we plan to extend the proposed methods in the framework by investigating other kernel types in the non-linear case. The proposed framework can also be extended to multimodal one-class classification, where data is projected from multiple modalities to a joint subspace.

6. Acknowledgement

This work has been supported by NSF IUCRC CVDI, project AMALIA funded by Business Finland and DSB, as well as projects Mad@work and Stroke-Data funded by Haltian. The work of Jenni Raitoharju was supported by Academy of Finland project 324475.

References

[1] N. Vaswani, T. Bouwmans, S. Javed, P. Narayanamurthy, Robust subspace learning: Robust pca, robust subspace tracking, and robust subspace recovery, IEEE Signal Processing Magazine 35 (4) (2018) 32–55.

[2] X.-L. Xu, C.-X. Ren, R.-C. Wu, H. Yan, Sliced inverse regression with adaptive spectral sparsity for dimension reduction, IEEE Transactions on Cybernetics 47 (3) (2017) 759–771. doi:10.1109/TCYB.2016.2526630.

[3] J. Guo, X. Li, Based on statistics of the gradients the feature matching algorithm, in: International Workshop on Education Technology and Computer Science, Vol. 2, 2009, pp. 983–987.

[4] E. Rodriguez-Martinez, T. Mu, J. Y. Goulermas, Sequential projection pursuit with kernel matrix update and symbolic model selection, IEEE Transactions on Cybernetics 44 (12) (2014) 2458–2469.

[5] X. He, C. Zhang, L. Zhang, X. Li, A-optimal projection for image representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 38 (5) (2015) 1009–1015.

[6] L. Xu, J. Raitoharju, A. Iosifidis, M. Gabbouj, Saliency-based weighted multi-label linear discriminant analysis, IEEE Transactions on Cybernetics (early access).

[7] Y. Lim, J. Kwon, H.-S. Oh, Principal component analysis in the wavelet domain, Pattern Recognition (2021) 108096.

[8] R. Sheikh, M. Patel, A. Sinhal, Recognizing mnist handwritten data set using pca and lda, in: International Conference on Artificial Intelligence: Advances and Applications, 2020, pp. 169–177.

[9] D. Cai, X. He, J. Han, Srda: An efficient algorithm for large-scale discriminant analysis, IEEE Transactions on Knowledge and Data Engineering 20 (1) (2007) 1–12.

[10] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, C. A. Raffel, Mixmatch: A holistic approach to semi-supervised learning, in: Advances in Neural Information Processing Systems, 2019, pp. 5049–5059.

[11] T. Kefi-Fatteh, R. Ksantini, M.-B. Kaâniche, A. Bouhoula, A novel incremental one-class support vector machine based on low variance direction, Pattern Recognition 91 (2019) 308–321.

[12] O. U. Lenz, D. Peralta, C. Cornelis, Average localised proximity: A new data descriptor with good default one-class classification performance, Pattern Recognition 118 (2021) 107991.

[13] S. Alam, S. K. Sonbhadra, S. Agarwal, P. Nagabhushan, One-class support vector classifiers: A survey, Knowledge-Based Systems (2020) 105754.

[14] L. M. Manevitz, M. Yousef, One-class svms for document classification, Journal of Machine Learning Research 2 (Dec) (2001) 139–154.

[15] G. Cohen, H. Sax, A. Geissbuhler, et al., Novelty detection using one-class parzen density estimator. an application to surveillance of nosocomial infections., in: Mie, 2008, pp. 21–26.

[16] M. Hejazi, Y. P. Singh, One-class support vector machines approach to anomaly detection, Applied Artificial Intelligence 27 (5) (2013) 351–366.

[17] F. Sohrab, J. Raitoharju, Boosting rare benthic macroinvertebrates taxa identification with one-class classification, in: IEEE Symposium Series on Computational Intelligence, 2020, pp. 928–933.

[18] W. Khreich, B. Khosravifar, A. Hamou-Lhadj, C. Talhi, An anomaly detection system based on variable n-gram features and one-class svm, Information and Software Technology 91 (2017) 186–197.
[19] L. Yin, H. Wang, W. Fan, Active learning based support vector data description method for robust novelty detection, Knowledge-Based Systems 153 (2018) 40–52.
[20] H.-J. Xing, Y.-J. Liu, Z.-C. He, Robust sparse coding for one-class classification based on correntropy and logarithmic penalty function, Pattern Recognition 111 (2021) 107685.
[21] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, Sv estimation of a distribution's support, Advances in Neural Information Processing systems 12.
[22] D. M. Tax, R. P. Duin, Support vector data description, Machine learning 54 (1) (2004) 45–66.
[23] V. Mygdalis, A. Iosifidis, A. Tefas, I. Pitas, Graph embedded one-class classifiers for media data classification, Pattern Recognition 60 (2016) 585–595.
[24] M. Turkoz, S. Kim, Y. Son, M. K. Jeong, E. A. Elsayed, Generalized support vector data description for anomaly detection, Pattern Recognition 100 (2020) 107119.
[25] F. Sohrab, J. Raitoharju, A. Iosifidis, M. Gabbouj, Ellipsoidal subspace support vector data description, IEEE Access 8 (2020) 122013–122025.
[26] F. Sohrab, J. Raitoharju, M. Gabbouj, A. Iosifidis, Subspace support vector data description, in: International Conference on Pattern Recognition, 2018, pp. 722–727.
[27] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, S. Lin, Graph embedding and extensions: A general framework for dimensionality reduction, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (1) (2006) 40–51.
[28] D. Cai, X. He, J. Han, Spectral regression for dimensionality reduction, Tech. rep., Computer Science Department, UIUC (2007).
[29] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning: data mining, inference, and prediction, Springer Science & Business Media, 2009.
[30] F. Sohrab, J. Raitoharju, A. Iosifidis, M. Gabbouj, Multimodal subspace support vector data description, Pattern Recognition 110 (2021) 107648.
[31] K. B. Petersen, M. S. Pedersen, The matrix cookbook, version 20121115 (nov 2012).
URL `https://www.math.uwaterloo.ca/ hwolkowi/matrixcookbook.pdf`
[32] N. Kwak, Nonlinear projection trick in kernel methods: An alternative to the kernel trick, IEEE Transactions on Neural Networks and Learning Systems 24 (12) (2013) 2113–2119.
[33] H. Hoffmann, Kernel pca for novelty detection, Pattern Recognition 40 (3) (2007) 863–874.
[34] S. Zheng, Smoothly approximated support vector domain description, Pattern Recognition 49 (2016) 55–64.
[35] A. Sharma, K. K. Paliwal, S. Imoto, S. Miyano, Principal component analysis using QR decomposition, International Journal of Machine Learning and Cybernetics 4.
[36] L. Deng, The mnist database of handwritten digit images for machine learning research, IEEE Signal Processing Magazine 29 (6) (2012) 141–142.
[37] X. Zhu, X. Wu, Class noise vs. attribute noise: A quantitative study, Artificial intelligence review 22 (3) (2004) 177–210.
[38] J. Demšar, Statistical comparisons of classifiers over multiple data sets, Journal of Machine Learning Research 7 (2006) 1–30.

Graph-Embedded Subspace Support Vector Data Description
Supplementary Material

Fahad Sohrab, Alexandros Iosifidis, Moncef Gabbouj, Jenni Raitoharju

This document contains supplementary material for the proposed Graph-Embedded Subspace Support Vector Data Description (GESSVDD). In Section S1, we provide the complexity analysis of gradient-based, spectral-based, and spectral regression-based updates. In the other sections, we report more extensive results and analysis for different variants of the proposed GESSVDD: In Section S2, we report the Geometric Mean $Gmean$ results of linear data description for all datasets along with $Gmean$ results for **non-linear** data description for MNIST and Liver datasets. The non-linear $Gmean$ results for all other datasets besides MNIST, Liver, and Heart are given in the manuscript. Section S2 also provides $Gmean$ results of the linear and non-linear data description in the proposed framework for different variants over the Heart data set with added noise to train and test sets. In Section S3, we provide results in terms of True Positive Rate (TPR), True Negative Rate (TNR), False Positive Rate (FPR), and False Negative Rate (FNR) for all proposed variants along with the competing methods for all datasets for both linear and non-linear cases. We provide the standard deviation of $Gmean$ over test sets for the datasets where the 70-30$\%$ selection is made five times for forming the train and test sets in Section S4 for the linear case and Section S5 for non-linear cases. Finally, all the plots of sensitivity analysis over MNIST data with target class 0 are provided for all the variants in the proposed framework for the non-linear case in Section S6.

S1. Complexity analysis of GESSVDD

GESSVDD can be solved via spectral, spectral regression-based, and gradient-based techniques. We provide the detailed complexity analysis of the main algorithm in the manuscript (see Section 3.3. Complexity analysis) and focus on the different solution updates in the following subsections.

S1.1. Complexity analysis of gradient-based update

The gradient-based update has the following steps involved in the sub-algorithm.

1. Computing $\mathbf{S}_\alpha = \mathbf{X}\mathbf{L}_\alpha\mathbf{X}^\intercal$ for a given $\mathbf{L}_\alpha$ has a complexity of multiplying three matrices as well. Thus, computing $\mathbf{S}_\alpha$ has the complexity of $\mathcal{O}(DN^2 + ND^2)$.
2. Computing $\Delta L = 2\mathbf{S}_{inv}\mathbf{Q}\mathbf{S}_\alpha - 2\mathbf{S}_{inv}\mathbf{Q}\mathbf{S}_\alpha\mathbf{Q}^\intercal\mathbf{S}_{inv}\mathbf{Q}\mathbf{S}_x^\intercal$ requires several matrix multiplications with the highest complexity being $\mathcal{O}(dD^2)$ and, thus, the overall complexity of this step also becomes $\mathcal{O}(dD^2)$.
3. The complexity of updating the $\mathbf{Q}$ is $\mathcal{O}\big(dD\big)$.

After adding all complexities of the gradient-based updates along with the complexity of the main algorithm, the gradient-based solution has a complexity of $\mathcal{O}(N^3)$. In a non-linear case, the kernel matrix $\mathbf{K}$ is formed, centralized, and decomposed via eigendecomposition. These steps have the complexity of $\mathcal{O}(N^3)$. The dimensionality of data in non-linear case changes from D to N for all corresponding steps. The total complexity of the non-linear version stays at $\mathcal{O}(N^3)$.

S1.2. Complexity analysis of spectral-based update

The main relatively intensive computational steps in the spectral-based update are computing $\mathbf{S}_\alpha$ and the generalized eigenvalue problem.

1. Computing $\mathbf{S}_\alpha$ has the complexity of $\mathcal{O}(DN^2 + ND^2)$.
2. Solving the eigenvalue problem $\mathbf{S}_\alpha\mathbf{q} = \upsilon\mathbf{S}_x\mathbf{q}$ has the complexity of $\mathcal{O}(D^3)$.

By adding the above complexities, the complexity of spectral-based update becomes $\mathcal{O}(DN^2 + ND^2 + D^3)$. By adding this to the complexity of the main algorithm, the complexity becomes $\mathcal{O}\big(N^3 + DN^2 + ND^2 + D^3\big)$. Hence for **linear** case, the complexity is $\mathcal{O}\big(N^3\big)$. In a non-linear case, the dimensionality of the data changes from D to N. In this case, the complexity in terms $\mathcal{O}\big(4N^3\big)$. In terms of $\mathcal{O}$ notation, the complexity in non-linear case stays at $\mathcal{O}\big(N^3\big)$.

S1.3. Complexity analysis of spectral regression-based update

The main computational steps in the spectral regression-based update are computing the generalized eigenvalue problem and obtaining the projection matrix $\mathbf{Q}$ in a least-square sense.

1. Solving the generalized eigenvalue problem $\mathbf{L}_\alpha\mathbf{t} = \upsilon\mathbf{L}_x\mathbf{t}$ has the complexity of $\mathcal{O}(N^3)$.
2. Solving $\mathbf{Q} = \mathbf{T}^\intercal\mathbf{X}^\intercal(\mathbf{X}\mathbf{X}^\intercal + \eta\mathbf{I})^{-1}$ involves the following steps. Computing $\mathbf{X}\mathbf{X}^\intercal$ has the complexity of $\mathcal{O}(D^2N)$. The complexity of multiplying η with each element of $\mathbf{I}$ is $\mathcal{O}(D^2)$. Adding $\mathbf{X}\mathbf{X}^\intercal$ with $\eta\mathbf{I}$ has the complexity of $\mathcal{O}(D^2)$. Taking inverse of $(\mathbf{X}\mathbf{X}^\intercal + \eta\mathbf{I})$ has the complexity of D^3 and multiplying the rest of matrices has the complexity of $dND + D^3$. Adding the complexities of all these steps, the complexity of $\mathbf{Q} = \mathbf{T}^\intercal\mathbf{X}^\intercal(\mathbf{X}\mathbf{X}^\intercal + \eta\mathbf{I})^{-1}$ becomes $\mathcal{O}(D^2N + 2D^2 + D^3 + dND + D^3)$.

Adding the above two complexities and the complexity of the main algorithm and assuming that the dimensionality D of data is always lower than the number of samples N, the total complexity in terms of big $\mathcal{O}$ notation becomes $\mathcal{O}(N^3)$. In non-linear case, the steps involved in npt have the complexity of $\mathcal{O}(N^3)$, moreover the dimensionality changes from D to N. Thus, the complexity increases but in terms of the big $\mathcal{O}$ notation still stays as $\mathcal{O}(N^3)$.

S2. GESSVDD Gmean results

S2.1. Linear GESSVDD Gmean results

Table 1: *Gmean* results for **linear** data description in the proposed framework for MNIST and Liver datasets

Dataset	MNIST											Liver		
Target class	0	1	2	3	4	5	6	7	8	9	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.48	0.67	0.45	0.30	0.55	0.49	0.58	0.45	0.45	0.55	0.50	0.38	0.45	0.41
GESSVDD-Sb-GR-max	**0.62**	0.64	0.28	0.45	0.54	0.34	0.49	0.33	0.44	0.50	0.46	0.47	0.46	0.47
GESSVDD-Sb-SR-max	0.48	0.59	**0.48**	0.44	**0.63**	0.53	0.12	0.65	**0.56**	0.57	0.50	**0.51**	0.38	0.45
GESSVDD-Sb-$\mathcal{S}$-min	0.48	0.67	0.45	0.30	0.55	0.49	0.58	0.45	0.45	0.55	0.50	0.41	0.50	0.45
GESSVDD-Sb-GR-min	**0.62**	0.64	0.28	0.45	0.54	0.34	0.49	0.33	0.44	0.50	0.46	0.44	0.46	0.45
GESSVDD-Sb-SR-min	0.48	0.59	**0.48**	0.44	**0.63**	0.53	0.12	0.65	**0.56**	0.57	0.50	0.48	**0.51**	**0.50**
GESSVDD-Sw-$\mathcal{S}$-max	0.30	**0.86**	0.33	0.38	0.41	0.35	0.46	0.61	0.41	0.60	0.47	0.36	0.35	0.36
GESSVDD-Sw-GR-max	0.43	0.78	0.33	0.46	0.49	0.34	0.61	0.62	0.41	0.49	0.50	0.37	0.35	0.36
GESSVDD-Sw-SR-max	0.38	0.74	0.39	0.45	0.60	0.52	0.65	0.61	0.55	0.64	**0.55**	0.39	0.36	0.37
GESSVDD-Sw-$\mathcal{S}$-min	0.30	**0.86**	0.33	0.38	0.41	0.35	0.46	0.61	0.41	0.60	0.47	0.39	0.38	0.39
GESSVDD-Sw-GR-min	0.43	0.78	0.33	0.46	0.49	0.34	0.61	0.62	0.41	0.49	0.50	0.36	0.36	0.36
GESSVDD-Sw-SR-min	0.38	0.74	0.39	0.45	0.60	0.52	0.65	0.61	0.55	0.64	**0.55**	0.41	0.37	0.39
GESSVDD-kNN-$\mathcal{S}$-max	0.45	**0.86**	0.45	**0.50**	0.50	0.49	0.36	0.43	0.45	0.59	0.51	0.41	0.39	0.40
GESSVDD-kNN-GR-max	0.40	0.84	0.33	0.47	0.60	0.38	**0.69**	0.53	0.51	0.58	0.53	0.43	0.34	0.39
GESSVDD-kNN-SR-max	0.40	0.81	0.21	0.43	0.59	**0.56**	0.63	0.65	0.52	**0.68**	**0.55**	0.41	0.41	0.41
GESSVDD-kNN-$\mathcal{S}$-min	0.45	**0.86**	0.45	**0.50**	0.50	0.49	0.36	0.43	0.45	0.59	0.51	0.36	0.48	0.42
GESSVDD-kNN-GR-min	0.40	0.84	0.33	0.47	0.60	0.38	**0.69**	0.53	0.51	0.58	0.53	0.45	0.38	0.42
GESSVDD-kNN-SR-min	0.40	0.81	0.21	0.43	0.59	**0.56**	0.63	0.65	0.52	**0.68**	**0.55**	0.46	0.43	0.44
GESSVDD-PCA-$\mathcal{S}$-max	0.24	0.84	0.35	0.39	0.31	0.38	0.38	**0.68**	0.45	0.62	0.46	0.45	0.43	0.44
GESSVDD-PCA-GR-max	0.36	0.85	0.34	0.46	0.49	0.34	0.44	0.60	0.41	0.51	0.48	0.40	0.39	0.39
GESSVDD-PCA-SR-max	0.39	0.85	0.29	0.44	0.55	0.47	0.66	0.60	0.50	0.67	0.54	0.34	0.40	0.37
GESSVDD-PCA-$\mathcal{S}$-min	0.24	0.84	0.35	0.39	0.31	0.38	0.38	**0.68**	0.45	0.62	0.46	0.40	0.42	0.41
GESSVDD-PCA-GR-min	0.36	0.85	0.34	0.46	0.49	0.34	0.44	0.60	0.41	0.51	0.48	0.41	0.38	0.40
GESSVDD-PCA-SR-min	0.39	0.85	0.29	0.44	0.55	0.47	0.66	0.60	0.50	0.67	0.54	0.36	0.40	0.38
GESSVDD-I-$\mathcal{S}$-max	0.33	0.83	0.35	0.41	0.31	0.38	0.48	**0.68**	0.48	0.58	0.48	0.42	0.42	0.42
GESSVDD-I-GR-max	0.38	0.83	0.31	0.46	0.47	0.34	0.62	0.65	0.40	0.50	0.50	0.39	0.41	0.40
GESSVDD-I-SR-max	0.37	0.83	0.29	0.40	0.55	0.47	0.67	0.59	0.50	0.65	0.53	0.42	0.40	0.41
GESSVDD-I-$\mathcal{S}$-min	0.33	0.83	0.35	0.41	0.31	0.38	0.48	**0.68**	0.48	0.58	0.48	0.39	0.42	0.40
GESSVDD-I-GR-min (ESSVDD)	0.38	0.83	0.31	0.46	0.47	0.34	0.62	0.65	0.40	0.50	0.50	0.35	0.38	0.37
GESSVDD-I-SR-min	0.37	0.83	0.29	0.40	0.55	0.47	0.67	0.59	0.50	0.65	0.53	0.41	0.40	0.41
GESSVDD-0-$\mathcal{S}$-max	0.48	0.71	0.34	0.40	0.48	0.34	0.58	0.52	0.42	0.51	0.48	0.34	0.33	0.34
GESSVDD-0-GR-max	0.41	0.81	0.29	0.39	0.45	0.31	0.57	0.52	0.40	0.44	0.46	0.32	0.33	0.33
GESSVDD-0-$\mathcal{S}$-min	0.48	0.71	0.34	0.40	0.48	0.34	0.58	0.52	0.42	0.51	0.48	0.33	0.38	0.35
GESSVDD-0-GR-min (SSVDD)	0.41	0.81	0.29	0.39	0.45	0.31	0.57	0.52	0.40	0.44	0.46	0.32	0.34	0.33

Table 2: *Gmean* results for **linear** data description in the proposed framework for Seeds, Qualitative bankruptcy, Somerville happiness, Iris, Ionosphere and Sonar datasets

Dataset	Seeds				Qualitative bankruptcy			Somerville happiness		
Target class	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.71	0.60	0.85	0.72	0.64	0.31	0.47	0.35	0.33	0.34
GESSVDD-Sb-GR-max	0.74	0.69	0.67	0.70	0.65	0.44	0.55	0.45	**0.49**	**0.47**
GESSVDD-Sb-SR-max	0.76	0.90	0.91	0.86	0.59	0.30	0.44	0.49	0.31	0.40
GESSVDD-Sb-$\mathcal{S}$-min	0.76	0.85	0.85	0.82	0.60	0.44	0.52	0.45	0.37	0.41
GESSVDD-Sb-GR-min	0.79	0.80	0.90	0.83	0.68	0.47	0.57	0.38	0.47	0.42
GESSVDD-Sb-SR-min	0.77	0.88	0.91	0.86	0.61	0.24	0.42	0.44	0.30	0.37
GESSVDD-Sw-$\mathcal{S}$-max	0.80	0.93	0.93	0.89	0.84	0.00	0.42	0.46	0.43	0.45
GESSVDD-Sw-GR-max	**0.87**	0.90	0.94	**0.91**	0.72	0.03	0.38	0.39	0.44	0.41
GESSVDD-Sw-SR-max	0.77	0.91	0.85	0.84	0.81	0.15	0.48	0.49	0.36	0.42
GESSVDD-Sw-$\mathcal{S}$-min	0.75	0.93	0.93	0.87	0.86	0.07	0.47	0.50	0.34	0.42
GESSVDD-Sw-GR-min	0.82	0.81	0.93	0.85	0.78	0.15	0.46	0.51	0.39	0.45
GESSVDD-Sw-SR-min	0.74	0.93	**0.95**	0.87	0.89	0.06	0.47	0.41	0.37	0.39
GESSVDD-kNN-$\mathcal{S}$-max	0.80	0.90	0.93	0.87	0.77	0.22	0.50	**0.52**	0.36	0.44
GESSVDD-kNN-GR-max	0.83	0.93	0.93	0.90	0.77	0.18	0.48	0.48	0.46	**0.47**
GESSVDD-kNN-SR-max	0.79	0.92	0.93	0.88	0.80	0.19	0.50	0.49	0.43	0.46
GESSVDD-kNN-$\mathcal{S}$-min	0.73	0.88	0.92	0.85	0.77	0.28	0.53	0.46	0.37	0.42
GESSVDD-kNN-GR-min	0.83	**0.94**	**0.95**	**0.91**	0.80	**0.46**	**0.63**	0.44	0.44	0.44
GESSVDD-kNN-SR-min	0.79	0.93	0.93	0.88	0.80	0.32	0.56	0.43	0.39	0.41
GESSVDD-PCA-$\mathcal{S}$-max	0.81	0.92	0.93	0.89	0.85	0.19	0.52	0.47	0.38	0.43
GESSVDD-PCA-GR-max	0.86	0.90	0.89	0.88	**0.94**	0.10	0.52	0.48	0.40	0.44
GESSVDD-PCA-SR-max	0.82	0.92	0.94	0.89	0.91	0.17	0.54	0.44	0.41	0.43
GESSVDD-PCA-$\mathcal{S}$-min	0.76	0.92	0.93	0.87	0.85	0.14	0.50	0.50	0.33	0.42
GESSVDD-PCA-GR-min	0.86	0.86	0.85	0.86	0.90	0.09	0.49	0.46	0.42	0.44
GESSVDD-PCA-SR-min	0.78	0.93	0.94	0.88	0.89	0.21	0.55	0.43	0.39	0.41
GESSVDD-I-$\mathcal{S}$-max	0.78	0.93	0.93	0.88	0.85	0.19	0.52	0.46	0.41	0.43
GESSVDD-I-GR-max	0.83	0.92	0.93	0.89	0.93	0.09	0.51	0.50	0.44	**0.47**
GESSVDD-I-SR-max	0.81	0.92	0.93	0.89	0.92	0.07	0.49	0.49	0.40	0.45
GESSVDD-I-$\mathcal{S}$-min	0.77	0.92	0.93	0.87	0.87	0.13	0.50	0.49	0.32	0.40
GESSVDD-I-GR-min (ESSVDD)	**0.87**	0.92	0.90	0.90	0.90	0.12	0.51	0.51	0.39	0.45
GESSVDD-I-SR-min	0.78	0.92	0.93	0.88	0.91	0.19	0.55	0.41	0.39	0.40
GESSVDD-0-$\mathcal{S}$-max	**0.87**	0.92	0.93	**0.91**	0.81	0.00	0.40	0.36	0.44	0.40
GESSVDD-0-GR-max	0.85	0.93	0.94	**0.91**	0.87	0.00	0.44	0.42	0.43	0.42
GESSVDD-0-$\mathcal{S}$-min	0.73	0.93	0.91	0.85	0.82	0.23	0.53	0.46	0.32	0.39
GESSVDD-0-GR-min (SSVDD)	0.85	0.93	**0.95**	**0.91**	0.90	0.17	0.53	0.49	0.43	0.46

Dataset	Iris				Ionosphere			Sonar		
Target class	I-S	S-VC	S-V	Av.	I-B	I-G	Av.	S-R	S-M	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.92	0.86	0.81	0.87	0.22	0.39	0.30	0.46	0.44	0.45
GESSVDD-Sb-GR-max	0.90	0.87	0.89	0.89	0.21	0.88	0.55	0.52	0.45	0.48
GESSVDD-Sb-SR-max	0.84	0.88	0.82	0.85	**0.46**	0.23	0.35	0.31	0.53	0.42
GESSVDD-Sb-$\mathcal{S}$-min	0.91	0.83	0.81	0.85	0.39	0.41	0.40	0.44	0.41	0.43
GESSVDD-Sb-GR-min	0.94	0.86	0.89	0.90	0.21	0.88	0.55	0.49	0.47	0.48
GESSVDD-Sb-SR-min	0.84	0.68	0.85	0.79	0.40	0.33	0.37	0.35	0.28	0.32
GESSVDD-Sw-$\mathcal{S}$-max	0.95	0.90	0.88	0.91	0.35	0.87	0.61	0.49	0.44	0.46
GESSVDD-Sw-GR-max	0.91	0.90	0.85	0.89	0.22	0.90	0.56	0.55	0.47	0.51
GESSVDD-Sw-SR-max	0.95	0.86	0.85	0.89	0.28	0.86	0.57	0.51	0.45	0.48
GESSVDD-Sw-$\mathcal{S}$-min	0.95	0.89	0.87	0.90	0.27	0.86	0.57	0.52	0.46	0.49
GESSVDD-Sw-GR-min	0.96	0.92	0.79	0.89	0.13	0.91	0.52	0.55	0.46	0.50
GESSVDD-Sw-SR-min	0.95	0.86	0.85	0.89	0.23	0.85	0.54	0.50	0.45	0.47
GESSVDD-kNN-$\mathcal{S}$-max	0.93	0.88	0.86	0.89	0.23	0.87	0.55	0.49	0.37	0.43
GESSVDD-kNN-GR-max	0.95	0.80	0.75	0.83	**0.46**	0.89	**0.67**	0.53	0.48	0.50
GESSVDD-kNN-SR-max	0.97	0.86	0.88	0.90	0.23	0.87	0.55	0.47	0.43	0.45
GESSVDD-kNN-$\mathcal{S}$-min	0.93	0.89	0.82	0.88	0.28	0.87	0.58	0.49	0.37	0.43
GESSVDD-kNN-GR-min	0.97	0.89	**0.91**	0.92	0.42	**0.92**	**0.67**	0.54	0.57	**0.56**
GESSVDD-kNN-SR-min	0.97	0.86	0.87	0.90	0.31	0.87	0.59	0.47	0.44	0.45
GESSVDD-PCA-$\mathcal{S}$-max	**0.99**	0.93	0.85	0.92	0.32	0.87	0.60	0.50	0.33	0.41
GESSVDD-PCA-GR-max	0.93	0.87	0.87	0.89	0.13	0.91	0.52	**0.57**	0.47	0.52
GESSVDD-PCA-SR-max	**0.99**	0.93	0.84	0.92	0.35	0.86	0.61	0.41	0.35	0.38
GESSVDD-PCA-$\mathcal{S}$-min	0.97	0.93	0.85	0.92	0.27	0.88	0.58	0.50	0.44	0.47
GESSVDD-PCA-GR-min	0.94	0.83	0.85	0.87	0.34	0.89	0.61	0.50	**0.59**	0.55
GESSVDD-PCA-SR-min	**0.99**	0.91	0.85	0.92	0.26	0.87	0.56	0.49	0.36	0.42
GESSVDD-I-$\mathcal{S}$-max	**0.99**	0.93	0.85	0.92	0.33	0.88	0.60	0.49	0.35	0.42
GESSVDD-I-GR-max	0.95	0.89	0.85	0.89	0.20	0.90	0.55	0.54	0.51	0.52
GESSVDD-I-SR-max	**0.99**	0.93	0.77	0.90	0.27	0.88	0.57	0.49	0.35	0.42
GESSVDD-I-$\mathcal{S}$-min	0.95	**0.94**	0.81	0.90	0.33	0.88	0.61	0.49	0.35	0.42
GESSVDD-I-GR-min (ESSVDD)	0.93	0.82	0.89	0.88	0.36	0.90	0.63	0.52	0.58	0.55
GESSVDD-I-SR-min	0.98	**0.94**	0.87	**0.93**	0.32	0.87	0.59	0.51	0.35	0.43
GESSVDD-0-$\mathcal{S}$-max	0.94	0.88	**0.91**	0.91	0.12	0.62	0.37	0.43	0.37	0.40
GESSVDD-0-GR-max	0.97	0.91	0.89	0.92	0.15	0.80	0.47	0.49	0.44	0.47
GESSVDD-0-$\mathcal{S}$-min	0.92	0.89	0.87	0.90	0.18	0.81	0.49	0.47	0.32	0.39
GESSVDD-0-GR-min (SSVDD)	0.96	0.91	0.90	0.92	0.12	0.78	0.45	0.51	0.55	0.53

Table 3: *Gmean* results for **linear** data description over Heart dataset and its manually created corrupted versions

Dataset	Heart Clean train set Clean test set			Heart Clean train set Corrupted test set			Heart Corrupted train set Clean test set			Heart Corrupted train set Corrupted test set		
Target class	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.52	0.44	0.48	0.20	0.22	0.21	0.12	0.07	0.09	0.42	0.36	0.39
GESSVDD-Sb-GR-max	0.49	0.48	0.49	0.17	0.22	0.20	0.24	**0.36**	**0.30**	0.36	0.30	0.33
GESSVDD-Sb-SR-max	0.52	0.43	0.47	0.15	0.11	0.13	**0.35**	0.17	0.26	0.44	0.41	0.43
GESSVDD-Sb-$\mathcal{S}$-min	0.51	0.48	0.49	0.26	**0.26**	**0.26**	0.05	0.06	0.05	0.35	0.42	0.38
GESSVDD-Sb-GR-min	0.59	0.68	**0.64**	0.14	0.23	0.18	0.21	0.23	0.22	0.32	**0.46**	0.39
GESSVDD-Sb-SR-min	0.55	0.53	0.54	**0.27**	0.11	0.19	0.24	0.09	0.16	0.44	0.43	0.44
GESSVDD-Sw-$\mathcal{S}$-max	0.51	0.49	0.50	0.13	0.14	0.13	0.00	0.00	0.00	0.26	0.40	0.33
GESSVDD-Sw-GR-max	0.54	0.66	0.60	0.21	0.19	0.20	0.00	0.00	0.00	0.30	0.40	0.35
GESSVDD-Sw-SR-max	0.52	0.57	0.55	0.03	0.00	0.01	0.00	0.00	0.00	0.34	0.42	0.38
GESSVDD-Sw-$\mathcal{S}$-min	0.49	0.61	0.55	0.09	0.12	0.10	0.00	0.00	0.00	0.30	0.43	0.37
GESSVDD-Sw-GR-min	**0.62**	0.60	0.61	0.21	0.20	0.20	0.00	0.00	0.00	0.33	0.37	0.35
GESSVDD-Sw-SR-min	0.50	0.58	0.54	0.00	0.03	0.01	0.03	0.00	0.01	0.29	0.33	0.31
GESSVDD-kNN-$\mathcal{S}$-max	0.56	0.55	0.55	0.04	0.13	0.08	0.04	0.00	0.02	0.41	0.44	0.42
GESSVDD-kNN-GR-max	0.53	0.60	0.57	0.08	0.19	0.14	0.14	0.08	0.11	0.40	0.38	0.39
GESSVDD-kNN-SR-max	0.47	0.57	0.52	0.00	0.06	0.03	0.00	0.00	0.00	0.39	0.43	0.41
GESSVDD-kNN-$\mathcal{S}$-min	0.53	0.54	0.54	0.08	0.03	0.05	0.08	0.00	0.04	**0.48**	0.43	**0.45**
GESSVDD-kNN-GR-min	0.54	0.61	0.58	0.09	0.00	0.04	0.32	0.08	0.20	0.35	0.44	0.39
GESSVDD-kNN-SR-min	0.53	0.58	0.55	0.00	0.13	0.06	0.00	0.00	0.00	0.35	0.42	0.38
GESSVDD-PCA-$\mathcal{S}$-max	0.50	0.58	0.54	0.05	0.00	0.02	0.00	0.00	0.00	0.33	0.38	0.35
GESSVDD-PCA-GR-max	0.53	0.53	0.53	0.00	0.00	0.00	0.00	0.00	0.00	0.38	0.43	0.40
GESSVDD-PCA-SR-max	0.48	0.54	0.51	0.00	0.12	0.06	0.00	0.00	0.00	0.31	0.39	0.35
GESSVDD-PCA-$\mathcal{S}$-min	0.54	0.66	0.60	0.00	0.13	0.07	0.00	0.00	0.00	0.30	0.45	0.37
GESSVDD-PCA-GR-min	0.54	0.62	0.58	0.12	0.00	0.06	0.00	0.00	0.00	0.38	0.45	0.41
GESSVDD-PCA-SR-min	0.54	0.55	0.54	0.07	0.12	0.10	0.00	0.00	0.00	0.31	0.36	0.34
GESSVDD-I-$\mathcal{S}$-max	0.53	0.54	0.53	0.06	0.06	0.06	0.00	0.00	0.00	0.25	0.34	0.30
GESSVDD-I-GR-max	0.56	0.60	0.58	0.20	0.10	0.15	0.00	0.00	0.00	0.31	0.42	0.36
GESSVDD-I-SR-max	0.48	0.57	0.53	0.03	0.03	0.03	0.00	0.00	0.00	0.41	0.36	0.39
GESSVDD-I-$\mathcal{S}$-min	0.56	0.57	0.57	0.07	0.03	0.05	0.00	0.00	0.00	0.38	0.43	0.41
GESSVDD-I-GR-min (ESSVDD)	0.53	**0.69**	0.61	0.00	0.00	0.00	0.00	0.00	0.00	0.35	0.39	0.37
GESSVDD-I-SR-min	0.50	0.49	0.49	0.03	0.00	0.01	0.00	0.00	0.00	0.35	0.32	0.34
GESSVDD-0-$\mathcal{S}$-max	0.48	0.47	0.47	0.10	0.19	0.14	0.00	0.00	0.00	0.14	0.28	0.21
GESSVDD-0-GR-max	0.55	0.64	0.59	0.14	0.18	0.16	0.00	0.00	0.00	0.25	0.31	0.28
GESSVDD-0-$\mathcal{S}$-min	0.53	0.62	0.57	0.18	0.00	0.09	0.00	0.00	0.00	0.31	0.40	0.36
GESSVDD-0-GR-min (SSVDD)	0.59	0.62	0.61	0.00	0.17	0.09	0.00	0.00	0.00	0.36	0.40	0.38

S2.2. Non-linear GESSVDD Gmean results

Table 4: *Gmean* results for **non-linear** data description in the proposed framework for MNIST and Liver datasets

Dataset	MNIST											Liver		
Target class	0	1	2	3	4	5	6	7	8	9	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.16	0.16	0.06	0.08	0.39	0.29	0.31	0.00	0.12	0.09	0.17	0.37	0.35	0.36
GESSVDD-Sb-GR-max	**0.62**	0.32	0.48	0.41	0.20	0.33	0.21	0.38	0.11	**0.58**	0.36	0.29	0.35	0.32
GESSVDD-Sb-SR-max	0.12	0.50	0.26	**0.51**	0.19	0.08	0.29	0.28	0.29	0.47	0.30	0.29	0.27	0.28
GESSVDD-Sb-$\mathcal{S}$-min	0.16	0.16	0.06	0.08	0.39	0.29	0.31	0.00	0.12	0.09	0.17	0.34	0.33	0.33
GESSVDD-Sb-GR-min	**0.62**	0.32	0.48	0.41	0.20	0.33	0.21	0.38	0.11	**0.58**	0.36	0.35	0.24	0.30
GESSVDD-Sb-SR-min	0.12	0.50	0.26	**0.51**	0.19	0.08	0.29	0.28	0.29	0.47	0.30	0.34	0.29	0.31
GESSVDD-Sw-$\mathcal{S}$-max	0.07	0.05	0.25	0.14	0.28	0.11	0.18	0.33	0.05	0.10	0.16	**0.45**	0.32	0.38
GESSVDD-Sw-GR-max	0.25	0.22	0.09	0.44	0.12	**0.52**	0.23	0.43	0.20	0.17	0.27	0.36	0.40	0.38
GESSVDD-Sw-SR-max	0.24	**0.54**	0.15	0.27	0.46	0.32	0.45	0.43	0.19	0.31	0.34	0.39	0.47	0.43
GESSVDD-Sw-$\mathcal{S}$-min	0.07	0.05	0.25	0.14	0.28	0.11	0.18	0.33	0.05	0.10	0.16	0.24	0.33	0.28
GESSVDD-Sw-GR-min	0.25	0.22	0.09	0.44	0.12	**0.52**	0.23	0.43	0.20	0.17	0.27	0.33	0.28	0.30
GESSVDD-Sw-SR-min	0.24	**0.54**	0.15	0.27	0.46	0.32	0.45	0.43	0.19	0.31	0.34	0.31	0.40	0.36
GESSVDD-kNN-$\mathcal{S}$-max	0.08	0.23	**0.53**	0.06	**0.55**	0.31	0.31	0.26	0.27	0.25	0.28	0.39	0.32	0.36
GESSVDD-kNN-GR-max	**0.62**	0.32	0.49	0.49	0.24	0.47	0.18	0.40	0.37	0.55	**0.41**	0.25	0.45	0.35
GESSVDD-kNN-SR-max	0.38	0.53	0.16	0.34	0.49	0.46	**0.48**	0.43	0.31	0.50	**0.41**	0.41	0.42	0.41
GESSVDD-kNN-$\mathcal{S}$-min	0.08	0.23	**0.53**	0.06	**0.55**	0.31	0.31	0.26	0.27	0.25	0.28	0.33	0.31	0.32
GESSVDD-kNN-GR-min	**0.62**	0.32	0.49	0.49	0.24	0.47	0.18	0.40	0.37	0.55	**0.41**	0.39	0.39	0.39
GESSVDD-kNN-SR-min	0.38	0.53	0.16	0.34	0.49	0.46	**0.48**	0.43	0.31	0.50	**0.41**	0.40	0.44	0.42
GESSVDD-PCA-$\mathcal{S}$-max	0.19	0.33	0.15	0.17	0.18	0.31	0.05	**0.54**	0.24	0.17	0.23	0.28	0.37	0.32
GESSVDD-PCA-GR-max	0.45	0.25	0.19	0.29	0.27	0.27	0.41	0.08	0.28	0.28	0.28	0.39	0.34	0.37
GESSVDD-PCA-SR-max	0.60	0.51	0.20	0.26	0.47	0.16	0.43	0.16	0.30	0.39	0.35	0.36	0.27	0.32
GESSVDD-PCA-$\mathcal{S}$-min	0.19	0.33	0.15	0.17	0.18	0.31	0.05	**0.54**	0.24	0.17	0.23	0.34	0.28	0.31
GESSVDD-PCA-GR-min	0.45	0.25	0.19	0.29	0.27	0.27	0.41	0.08	0.28	0.28	0.28	0.31	0.34	0.32
GESSVDD-PCA-SR-min	0.60	0.51	0.20	0.26	0.47	0.16	0.43	0.16	0.30	0.39	0.35	0.32	0.26	0.29
GESSVDD-I-$\mathcal{S}$-max	0.30	0.33	0.17	0.17	0.30	0.32	0.42	**0.54**	0.21	0.17	0.29	0.42	0.27	0.34
GESSVDD-I-GR-max	0.36	0.34	0.18	0.09	0.19	**0.52**	0.46	0.43	0.36	0.21	0.31	0.35	0.42	0.39
GESSVDD-I-SR-max	0.60	0.51	0.21	0.28	0.47	0.30	0.43	0.25	0.30	0.39	0.37	0.35	0.42	0.39
GESSVDD-I-$\mathcal{S}$-min	0.30	0.33	0.17	0.17	0.30	0.32	0.42	**0.54**	0.21	0.17	0.29	0.39	0.26	0.32
GESSVDD-I-GR-min (ESSVDD)	0.36	0.34	0.18	0.09	0.19	**0.52**	0.46	0.43	0.36	0.21	0.31	0.40	**0.49**	**0.45**
GESSVDD-I-SR-min	0.60	0.51	0.21	0.28	0.47	0.30	0.43	0.25	0.30	0.39	0.37	0.30	0.37	0.33
GESSVDD-0-$\mathcal{S}$-max	0.22	0.09	0.06	0.28	0.23	0.40	0.08	0.28	0.11	0.05	0.18	0.36	0.37	0.36
GESSVDD-0-GR-max	0.60	0.34	0.48	0.39	0.43	0.49	0.43	0.35	**0.44**	0.17	**0.41**	0.36	0.44	0.40
GESSVDD-0-$\mathcal{S}$-min	0.22	0.09	0.06	0.28	0.23	0.40	0.08	0.28	0.11	0.05	0.18	0.37	0.37	0.37
GESSVDD-0-GR-min (SSVDD)	0.60	0.34	0.48	0.39	0.43	0.49	0.43	0.35	**0.44**	0.17	**0.41**	0.37	0.39	0.38

Table 5: *Gmean* results for **non-linear** data description over Heart dataset and its manually created corrupted versions

Dataset	Heart Clean train set Clean test set			Heart Clean train set Corrupted test set			Heart Corrupted train set Clean test set			Heart Corrupted train set Corrupted test set		
Target class	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.29	0.33	0.31	0.42	0.29	0.36	0.13	0.05	0.09	0.35	0.40	0.37
GESSVDD-Sb-GR-max	0.34	0.31	0.32	0.35	0.23	0.29	0.33	0.31	**0.32**	0.27	0.45	0.36
GESSVDD-Sb-SR-max	0.32	0.24	0.28	0.22	0.24	0.23	0.30	0.22	0.26	0.43	0.44	0.44
GESSVDD-Sb-$\mathcal{S}$-min	0.41	0.43	0.42	**0.44**	0.19	0.32	0.19	0.19	0.19	0.35	0.20	0.28
GESSVDD-Sb-GR-min	0.45	0.33	0.39	0.32	0.18	0.25	0.21	0.32	0.26	0.33	0.33	0.33
GESSVDD-Sb-SR-min	0.46	0.32	0.39	0.25	0.13	0.19	**0.36**	0.16	0.26	0.43	0.44	0.43
GESSVDD-Sw-$\mathcal{S}$-max	0.43	0.38	0.41	0.00	0.12	0.06	0.13	0.00	0.06	0.36	0.34	0.35
GESSVDD-Sw-GR-max	0.16	0.32	0.24	0.36	**0.39**	**0.38**	0.07	0.11	0.09	0.42	0.39	0.41
GESSVDD-Sw-SR-max	0.43	0.40	0.41	0.11	0.06	0.08	0.00	0.06	0.03	0.43	0.41	0.42
GESSVDD-Sw-$\mathcal{S}$-min	0.34	0.46	0.40	0.00	0.16	0.08	0.00	0.15	0.08	0.44	0.43	0.43
GESSVDD-Sw-GR-min	0.40	0.47	0.44	0.18	0.29	0.24	0.04	0.05	0.05	0.45	0.44	0.45
GESSVDD-Sw-SR-min	0.44	0.30	0.37	0.03	0.19	0.11	0.00	0.00	0.00	0.41	0.33	0.37
GESSVDD-kNN-$\mathcal{S}$-max	0.44	0.47	0.45	0.00	0.00	0.00	0.14	0.04	0.09	0.45	0.45	0.45
GESSVDD-kNN-GR-max	0.43	**0.57**	0.50	0.28	0.19	0.24	0.12	0.00	0.06	0.44	0.38	0.41
GESSVDD-kNN-SR-max	0.42	0.43	0.42	0.04	0.03	0.03	0.00	0.00	0.00	0.47	0.40	0.44
GESSVDD-kNN-$\mathcal{S}$-min	0.49	0.51	0.50	0.03	0.00	0.02	0.03	0.00	0.02	0.46	0.43	0.44
GESSVDD-kNN-GR-min	0.47	0.44	0.46	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.43	0.46
GESSVDD-kNN-SR-min	0.47	0.50	0.49	0.04	0.07	0.06	0.06	0.00	0.03	0.45	0.38	0.41
GESSVDD-PCA-$\mathcal{S}$-max	0.37	0.31	0.34	0.00	0.03	0.01	0.00	0.00	0.00	**0.49**	0.36	0.42
GESSVDD-PCA-GR-max	0.47	0.30	0.39	0.30	0.34	0.32	0.13	0.16	0.14	0.45	0.36	0.40
GESSVDD-PCA-SR-max	0.46	0.42	0.44	0.19	0.05	0.12	0.06	0.09	0.08	0.34	0.41	0.37
GESSVDD-PCA-$\mathcal{S}$-min	0.24	0.27	0.25	0.06	0.11	0.08	0.00	0.00	0.00	0.43	0.39	0.41
GESSVDD-PCA-GR-min	0.30	0.28	0.29	0.28	0.22	0.25	0.00	0.04	0.02	0.46	**0.47**	**0.47**
GESSVDD-PCA-SR-min	0.44	0.44	0.44	0.10	0.07	0.09	0.03	0.00	0.02	0.37	0.41	0.39
GESSVDD-I-$\mathcal{S}$-max	0.38	0.39	0.38	0.06	0.03	0.04	0.04	0.00	0.02	0.45	0.38	0.42
GESSVDD-I-GR-max	0.42	0.28	0.35	0.37	0.35	0.36	0.03	0.05	0.04	0.48	0.31	0.39
GESSVDD-I-SR-max	0.49	0.46	0.47	0.13	0.03	0.08	0.00	0.00	0.00	0.43	0.44	0.43
GESSVDD-I-$\mathcal{S}$-min	0.44	0.35	0.39	0.00	0.25	0.12	0.08	0.00	0.04	0.36	0.39	0.38
GESSVDD-I-GR-min (ESSVDD)	0.38	0.37	0.37	0.37	0.15	0.26	0.00	0.11	0.06	0.38	0.46	0.42
GESSVDD-I-SR-min	0.46	0.44	0.45	0.15	0.14	0.14	0.00	0.08	0.04	0.42	0.45	0.44
GESSVDD-0-$\mathcal{S}$-max	0.35	0.34	0.34	0.24	0.21	0.22	0.26	0.32	0.29	0.39	0.36	0.38
GESSVDD-0-GR-max	0.45	0.46	0.46	0.06	0.22	0.14	0.06	**0.44**	0.25	0.46	0.36	0.41
GESSVDD-0-$\mathcal{S}$-min	0.32	0.41	0.37	0.14	0.16	0.15	**0.36**	0.09	0.23	0.34	0.40	0.37
GESSVDD-0-GR-min (SSVDD)	**0.53**	0.49	**0.51**	0.12	0.24	0.18	0.00	0.09	0.04	0.37	**0.47**	0.42

S3. GESSVDD TPR TNR FPR FNR results

*S3.1. TPR TNR FPR FNR results for **linear** data description*

*S3.1.1. TPR results **linear** data description*

Table 6: *TPR* results for **linear** data description

Dataset	Seeds				Qualitative bankruptcy			Somerville happiness		
Target class	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.77	0.76	0.88	0.80	0.76	0.46	0.61	**0.83**	0.76	0.80
GESSVDD-Sb-GR-max	0.90	0.62	0.65	0.72	0.83	0.44	0.63	0.73	0.73	0.73
GESSVDD-Sb-SR-max	0.82	0.86	0.96	0.88	0.58	0.46	0.52	0.61	0.68	0.65
GESSVDD-Sb-$\mathcal{S}$-min	0.81	0.75	0.92	0.83	0.79	0.59	0.69	0.68	0.68	0.68
GESSVDD-Sb-GR-min	0.88	0.77	0.93	0.86	0.91	0.35	0.63	0.70	0.79	0.74
GESSVDD-Sb-SR-min	0.85	0.86	0.96	0.89	0.67	0.68	0.68	0.80	0.69	0.75
GESSVDD-Sw-$\mathcal{S}$-max	0.85	0.92	**0.97**	0.91	0.82	**0.95**	0.89	0.80	0.80	0.80
GESSVDD-Sw-GR-max	**0.90**	0.86	0.93	0.90	0.92	0.91	0.92	0.76	0.80	0.78
GESSVDD-Sw-SR-max	**0.90**	0.90	0.96	0.92	0.89	0.92	0.90	0.75	0.74	0.74
GESSVDD-Sw-$\mathcal{S}$-min	0.88	0.91	0.96	0.92	0.87	0.89	0.88	0.80	0.73	0.76
GESSVDD-Sw-GR-min	0.87	0.88	0.90	0.88	0.88	0.91	0.89	0.73	0.80	0.77
GESSVDD-Sw-SR-min	0.84	0.94	0.96	0.91	0.93	0.88	0.90	0.75	0.77	0.76
GESSVDD-kNN-$\mathcal{S}$-max	0.79	0.91	0.96	0.89	0.84	0.86	0.85	0.71	0.66	0.69
GESSVDD-kNN-GR-max	0.89	0.91	0.96	0.92	0.88	0.71	0.79	0.77	0.61	0.69
GESSVDD-kNN-SR-max	0.76	0.90	0.93	0.86	0.89	0.76	0.82	0.71	0.56	0.64
GESSVDD-kNN-$\mathcal{S}$-min	0.77	0.85	0.96	0.86	0.84	0.72	0.78	0.74	0.61	0.67
GESSVDD-kNN-GR-min	0.82	0.93	**0.97**	0.91	0.81	0.38	0.60	0.77	0.54	0.65
GESSVDD-kNN-SR-min	0.74	0.90	0.95	0.86	0.86	0.66	0.76	0.74	0.67	0.71
GESSVDD-PCA-$\mathcal{S}$-max	0.87	**0.95**	0.93	0.92	0.86	0.80	0.83	0.72	0.74	0.73
GESSVDD-PCA-GR-max	0.85	0.89	0.92	0.89	0.95	0.88	0.91	0.77	0.80	0.79
GESSVDD-PCA-SR-max	0.79	**0.95**	0.95	0.90	0.91	0.81	0.86	0.77	0.73	0.75
GESSVDD-PCA-$\mathcal{S}$-min	0.85	0.93	0.92	0.90	0.83	0.81	0.82	0.77	0.71	0.74
GESSVDD-PCA-GR-min	0.86	0.82	0.94	0.87	0.95	0.84	0.90	0.81	0.78	0.79
GESSVDD-PCA-SR-min	0.87	0.94	0.96	**0.92**	0.93	0.87	0.90	0.77	0.67	0.72
GESSVDD-I-$\mathcal{S}$-max	0.81	**0.95**	0.92	0.90	0.84	0.83	0.84	0.72	0.79	0.76
GESSVDD-I-GR-max	0.77	0.91	0.92	0.87	0.93	0.81	0.87	0.72	0.85	0.79
GESSVDD-I-SR-max	0.84	**0.95**	0.92	0.90	0.92	0.90	0.91	0.82	0.68	0.75
GESSVDD-I-$\mathcal{S}$-min	0.84	0.92	0.91	0.89	0.83	0.86	0.85	0.73	0.78	0.75
GESSVDD-I-GR-min (ESSVDD)	0.86	0.88	0.93	0.89	0.90	0.82	0.86	0.77	0.82	0.80
GESSVDD-I-SR-min	0.83	**0.95**	0.92	0.90	0.93	0.83	0.88	0.73	0.78	0.75
GESSVDD-0-$\mathcal{S}$-max	0.87	0.90	0.94	0.90	0.80	0.90	0.85	0.80	**0.87**	**0.84**
GESSVDD-0-GR-max	0.87	0.90	**0.97**	0.91	0.86	0.89	0.87	0.78	0.86	0.82
GESSVDD-0-$\mathcal{S}$-min	0.84	0.94	0.90	0.89	0.80	0.85	0.83	0.75	0.62	0.68
GESSVDD-0-GR-min (SSVDD)	0.87	0.90	**0.97**	0.91	0.90	0.85	0.88	0.81	0.86	0.84
ESVDD	0.70	0.76	0.79	0.75	0.93	0.79	0.86	0.68	0.66	0.67
SVDD	0.87	0.90	0.94	0.90	**0.96**	0.88	**0.92**	0.75	0.84	0.79
OCSVM	0.59	0.60	0.50	0.56	0.37	0.54	0.46	0.58	0.62	0.60

Table 7: *TPR* results for **linear** data description

Dataset	Iris				Ionosphere			Sonar		
Target class	I-S	S-VC	S-V	Av.	I-B	I-G	Av.	S-R	S-M	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.85	0.88	0.79	0.84	0.68	0.68	0.68	0.62	0.31	0.46
GESSVDD-Sb-GR-max	0.87	0.89	0.85	0.87	0.80	0.96	0.88	0.86	0.75	0.80
GESSVDD-Sb-SR-max	0.87	0.83	0.85	0.85	0.58	0.21	0.39	0.59	0.69	0.64
GESSVDD-Sb-$\mathcal{S}$-min	0.83	0.84	0.79	0.82	0.71	0.49	0.60	0.52	0.27	0.39
GESSVDD-Sb-GR-min	0.89	0.89	0.85	0.88	0.81	0.96	0.89	0.86	0.58	0.72
GESSVDD-Sb-SR-min	0.87	0.61	0.84	0.77	0.56	0.39	0.48	0.50	0.38	0.44
GESSVDD-Sw-$\mathcal{S}$-max	0.92	0.92	0.84	0.89	0.77	**0.97**	0.87	0.56	0.76	0.66
GESSVDD-Sw-GR-max	0.89	0.89	0.89	0.89	**0.91**	0.94	0.93	0.82	0.84	0.83
GESSVDD-Sw-SR-max	0.92	0.79	0.79	0.83	0.65	0.96	0.80	0.61	0.88	0.75
GESSVDD-Sw-$\mathcal{S}$-min	0.92	0.85	0.83	0.87	0.76	0.97	0.87	0.54	0.65	0.60
GESSVDD-Sw-GR-min	0.92	0.89	0.88	0.90	0.83	0.95	0.89	0.81	0.82	0.82
GESSVDD-Sw-SR-min	0.92	0.80	0.84	0.85	0.80	0.96	0.88	0.69	0.85	0.77
GESSVDD-kNN-$\mathcal{S}$-max	0.91	0.93	0.85	0.90	0.67	0.94	0.81	0.53	0.82	0.68
GESSVDD-kNN-GR-max	0.93	0.83	0.83	0.86	0.39	0.93	0.66	0.68	0.77	0.72
GESSVDD-kNN-SR-max	0.88	0.88	0.84	0.87	0.64	0.95	0.79	0.56	0.82	0.69
GESSVDD-kNN-$\mathcal{S}$-min	0.93	0.88	0.89	0.90	0.70	0.94	0.82	0.53	0.82	0.68
GESSVDD-kNN-GR-min	0.95	0.84	0.88	0.89	0.28	0.93	0.60	0.72	0.77	0.75
GESSVDD-kNN-SR-min	**0.97**	0.91	0.79	0.89	0.58	0.95	0.76	0.56	0.79	0.67
GESSVDD-PCA-$\mathcal{S}$-max	**0.97**	0.91	0.79	0.89	0.62	0.97	0.80	0.53	0.92	0.72
GESSVDD-PCA-GR-max	0.88	0.81	0.79	0.83	0.84	0.94	0.89	0.78	0.79	0.79
GESSVDD-PCA-SR-max	**0.97**	0.91	0.76	0.88	0.78	0.95	0.86	0.43	0.90	0.67
GESSVDD-PCA-$\mathcal{S}$-min	0.93	**0.95**	0.76	0.88	0.67	0.95	0.81	0.53	0.72	0.63
GESSVDD-PCA-GR-min	0.91	0.84	0.83	0.86	0.86	0.93	0.89	0.66	0.80	0.73
GESSVDD-PCA-SR-min	**0.97**	0.88	0.75	0.87	0.67	0.94	0.81	0.48	0.90	0.69
GESSVDD-I-$\mathcal{S}$-max	**0.97**	0.91	0.79	0.89	0.70	0.95	0.82	0.51	0.90	0.71
GESSVDD-I-GR-max	0.91	0.89	0.88	0.89	0.83	0.94	0.88	0.68	0.79	0.73
GESSVDD-I-SR-max	**0.97**	0.92	0.77	0.89	0.69	0.97	0.83	0.49	0.90	0.70
GESSVDD-I-$\mathcal{S}$-min	0.91	0.92	0.84	0.89	0.65	0.95	0.80	0.51	0.87	0.69
GESSVDD-I-GR-min (ESSVDD)	0.88	0.89	**0.95**	**0.91**	0.85	0.93	0.89	0.63	0.81	0.72
GESSVDD-I-SR-min	0.96	0.92	0.79	0.89	0.77	0.97	0.87	0.53	0.87	0.70
GESSVDD-0-$\mathcal{S}$-max	0.89	0.87	0.88	0.88	0.90	0.95	**0.93**	0.89	0.92	0.91
GESSVDD-0-GR-max	0.93	0.89	0.88	0.90	0.89	0.94	0.92	**0.90**	0.93	0.91
GESSVDD-0-$\mathcal{S}$-min	0.91	0.88	0.91	0.90	0.89	0.92	0.91	0.83	0.78	0.80
GESSVDD-0-GR-min (SSVDD)	0.92	0.89	0.88	0.90	0.90	0.94	0.92	0.83	0.79	0.81
ESVDD	0.79	0.73	0.77	0.76	0.50	0.83	0.66	0.00	0.01	0.00
SVDD	0.85	0.88	0.88	0.87	0.86	0.93	0.90	0.75	0.84	0.80
OCSVM	0.55	0.57	0.51	0.54	0.49	0.54	0.52	0.52	0.54	0.53

Table 8: *TPR* results for **linear** data description

Dataset	MNIST											Liver		
Target class	0	1	2	3	4	5	6	7	8	9	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.89	0.69	0.33	0.30	0.46	0.65	0.78	0.64	0.79	0.65	0.62	0.92	0.63	0.78
GESSVDD-Sb-GR-max	0.84	0.82	**0.98**	0.43	0.69	0.82	0.41	0.12	0.64	0.90	0.67	0.51	0.60	0.56
GESSVDD-Sb-SR-max	0.64	0.47	0.62	0.32	0.69	0.37	0.15	0.69	0.45	0.51	0.49	0.53	0.80	0.67
GESSVDD-Sb-$\mathcal{S}$-min	0.89	0.69	0.33	0.30	0.46	0.65	0.78	0.64	0.79	0.65	0.62	0.46	0.56	0.51
GESSVDD-Sb-GR-min	0.84	0.82	**0.98**	0.43	0.69	0.82	0.41	0.12	0.64	0.90	0.67	0.62	0.59	0.60
GESSVDD-Sb-SR-min	0.64	0.47	0.62	0.32	0.69	0.37	0.15	0.69	0.45	0.51	0.49	0.67	0.57	0.62
GESSVDD-Sw-$\mathcal{S}$-max	0.97	0.99	**0.98**	0.97	0.98	**0.98**	0.98	0.98	0.98	**0.98**	0.98	0.92	**0.93**	**0.92**
GESSVDD-Sw-GR-max	**0.99**	0.99	**0.98**	0.97	0.98	**0.98**	0.98	**0.99**	**0.99**	**0.98**	0.98	0.89	0.86	0.87
GESSVDD-Sw-SR-max	0.96	0.99	0.96	0.96	0.93	0.89	0.91	0.93	0.92	0.90	0.94	0.90	0.87	0.88
GESSVDD-Sw-$\mathcal{S}$-min	0.97	0.99	**0.98**	0.97	0.98	**0.98**	0.98	0.98	0.98	**0.98**	0.98	0.88	0.87	0.88
GESSVDD-Sw-GR-min	**0.99**	0.99	**0.98**	0.97	0.98	**0.98**	0.98	**0.99**	**0.99**	**0.98**	0.98	0.92	0.87	0.90
GESSVDD-Sw-SR-min	0.96	0.99	0.96	0.96	0.93	0.89	0.91	0.93	0.92	0.90	0.94	0.91	0.87	0.89
GESSVDD-kNN-$\mathcal{S}$-max	0.72	0.99	0.55	0.58	0.51	0.68	0.80	**0.99**	0.81	**0.98**	0.76	0.86	0.89	0.87
GESSVDD-kNN-GR-max	0.90	0.99	0.87	0.37	0.84	0.55	0.98	**0.99**	0.57	**0.98**	0.80	0.63	0.52	0.57
GESSVDD-kNN-SR-max	0.92	0.91	0.39	0.95	0.86	0.79	0.86	0.90	0.82	0.81	0.82	0.78	0.76	0.77
GESSVDD-kNN-$\mathcal{S}$-min	0.72	0.99	0.55	0.58	0.51	0.68	0.80	**0.99**	0.81	**0.98**	0.76	0.53	0.72	0.63
GESSVDD-kNN-GR-min	0.90	0.99	0.87	0.37	0.84	0.55	0.98	**0.99**	0.57	**0.98**	0.80	0.47	0.75	0.61
GESSVDD-kNN-SR-min	0.92	0.91	0.39	0.95	0.86	0.79	0.86	0.90	0.82	0.81	0.82	0.70	0.71	0.70
GESSVDD-PCA-$\mathcal{S}$-max	0.93	0.99	**0.98**	0.84	0.96	0.97	**0.99**	0.98	0.98	**0.98**	0.96	0.88	0.90	0.89
GESSVDD-PCA-GR-max	0.98	0.99	**0.98**	0.97	0.98	0.97	0.98	**0.99**	0.98	**0.98**	0.98	0.90	0.91	0.90
GESSVDD-PCA-SR-max	0.96	0.94	0.97	0.96	0.94	0.91	0.91	0.95	0.93	0.92	0.94	0.89	0.88	0.89
GESSVDD-PCA-$\mathcal{S}$-min	0.93	0.99	**0.98**	0.84	0.96	0.97	**0.99**	0.98	0.98	**0.98**	0.96	0.86	0.85	0.86
GESSVDD-PCA-GR-min	0.98	0.99	**0.98**	0.97	0.98	0.97	0.98	**0.99**	0.98	**0.98**	0.98	0.92	0.87	0.90
GESSVDD-PCA-SR-min	0.96	0.94	0.97	0.96	0.94	0.91	0.91	0.95	0.93	0.92	0.94	**0.93**	0.89	0.91
GESSVDD-I-$\mathcal{S}$-max	0.91	0.99	0.96	0.82	0.59	**0.98**	0.98	0.97	0.73	**0.98**	0.89	0.89	0.90	0.90
GESSVDD-I-GR-max	**0.99**	0.99	**0.98**	0.97	0.98	0.97	0.98	0.98	0.98	**0.98**	0.98	0.91	0.87	0.89
GESSVDD-I-SR-max	0.96	0.94	0.97	0.96	0.94	0.91	0.91	0.96	0.93	0.89	0.94	0.89	0.89	0.89
GESSVDD-I-$\mathcal{S}$-min	0.91	0.99	0.96	0.82	0.59	**0.98**	0.98	0.97	0.73	**0.98**	0.89	0.84	0.85	0.85
GESSVDD-I-GR-min (ESSVDD)	**0.99**	0.99	**0.98**	0.97	0.98	0.97	0.98	0.98	0.98	**0.98**	0.98	0.91	0.88	0.90
GESSVDD-I-SR-min	0.96	0.94	0.97	0.96	0.94	0.91	0.91	0.96	0.93	0.89	0.94	0.84	0.91	0.87
GESSVDD-0-$\mathcal{S}$-max	0.98	**1.00**	**0.98**	**0.98**	0.98	0.97	0.98	**0.99**	**0.99**	**0.98**	0.98	0.90	0.91	0.90
GESSVDD-0-GR-max	0.98	0.99	**0.98**	**0.98**	**0.99**	**0.98**	0.98	**0.99**	**0.99**	**0.98**	**0.99**	0.90	0.88	0.89
GESSVDD-0-$\mathcal{S}$-min	0.98	**1.00**	**0.98**	**0.98**	0.98	0.97	0.98	**0.99**	**0.99**	**0.98**	0.98	0.91	0.90	0.90
GESSVDD-0-GR-min (SSVDD)	0.98	0.99	**0.98**	**0.98**	**0.99**	**0.98**	0.98	**0.99**	**0.99**	**0.98**	**0.99**	0.90	0.89	0.89
ESVDD	0.00	0.66	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.03	0.07	0.89	0.90	0.90
SVDD	0.96	0.98	0.96	0.97	0.97	0.96	0.97	0.96	0.97	0.97	0.97	0.60	0.49	0.55
OCSVM	0.52	0.47	0.50	0.49	0.46	0.49	0.49	0.52	0.48	0.45	0.49	0.90	0.90	0.90

Table 9: *TPR* results for **linear** data description

Dataset	Heart Clean train set Clean test set			Heart Clean train set Corrupted test set			Heart Corrupted train set Clean test set			Heart Corrupted train set Corrupted test set		
Target class	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.74	0.57	0.65	0.06	0.08	0.07	0.68	0.58	0.63	0.76	0.69	0.73
GESSVDD-Sb-GR-max	0.64	0.44	0.54	0.07	0.08	0.07	0.80	0.66	0.73	0.56	0.57	0.56
GESSVDD-Sb-SR-max	0.65	0.57	0.61	0.04	0.04	0.04	0.60	0.40	0.50	0.63	0.55	0.59
GESSVDD-Sb-$\mathcal{S}$-min	0.72	0.48	0.60	0.08	0.12	0.10	0.70	0.64	0.67	0.70	0.75	0.73
GESSVDD-Sb-GR-min	0.58	0.70	0.64	0.05	0.09	0.07	0.68	0.33	0.50	0.54	0.39	0.46
GESSVDD-Sb-SR-min	0.48	0.63	0.55	0.12	0.04	0.08	0.75	0.53	0.64	0.70	0.48	0.59
GESSVDD-Sw-$\mathcal{S}$-max	**0.92**	0.90	0.91	0.06	0.04	0.05	**1.00**	**1.00**	**1.00**	0.94	0.92	0.93
GESSVDD-Sw-GR-max	0.89	0.86	0.88	0.14	0.12	0.13	**1.00**	**1.00**	**1.00**	0.90	0.90	0.90
GESSVDD-Sw-SR-max	0.84	0.83	0.83	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.84	0.84	0.84
GESSVDD-Sw-$\mathcal{S}$-min	0.86	0.90	0.88	0.03	0.05	0.04	0.99	**1.00**	**1.00**	0.88	0.82	0.85
GESSVDD-Sw-GR-min	0.83	0.82	0.83	0.15	0.13	0.14	**1.00**	**1.00**	**1.00**	0.84	0.82	0.83
GESSVDD-Sw-SR-min	0.86	0.85	0.85	0.00	0.00	0.00	0.95	**1.00**	0.98	0.90	0.84	0.87
GESSVDD-kNN-$\mathcal{S}$-max	0.75	0.80	0.77	0.01	0.03	0.02	0.94	**1.00**	0.97	0.86	0.82	0.84
GESSVDD-kNN-GR-max	0.78	0.82	0.80	0.04	0.07	0.05	0.94	0.96	0.95	0.62	0.63	0.62
GESSVDD-kNN-SR-max	0.74	0.83	0.78	0.00	0.01	0.00	**1.00**	**1.00**	**1.00**	0.88	0.73	0.80
GESSVDD-kNN-$\mathcal{S}$-min	0.80	0.75	0.78	0.02	0.00	0.01	**1.00**	**1.00**	**1.00**	0.76	0.76	0.76
GESSVDD-kNN-GR-min	0.75	0.78	0.76	0.02	0.00	0.01	0.88	0.96	0.92	0.63	0.77	0.70
GESSVDD-kNN-SR-min	0.75	0.83	0.79	0.00	0.05	0.03	**1.00**	**1.00**	**1.00**	0.70	0.64	0.67
GESSVDD-PCA-$\mathcal{S}$-max	0.88	0.88	0.88	0.01	0.00	0.01	0.99	**1.00**	**1.00**	0.87	**0.93**	0.90
GESSVDD-PCA-GR-max	0.83	0.85	0.84	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.85	0.83	0.84
GESSVDD-PCA-SR-max	0.85	0.84	0.85	0.00	0.02	0.01	**1.00**	**1.00**	**1.00**	0.88	0.80	0.84
GESSVDD-PCA-$\mathcal{S}$-min	0.85	0.84	0.84	0.00	0.05	0.03	**1.00**	**1.00**	**1.00**	0.90	0.79	0.84
GESSVDD-PCA-GR-min	0.82	0.83	0.83	0.10	0.00	0.05	**1.00**	**1.00**	**1.00**	0.82	0.80	0.81
GESSVDD-PCA-SR-min	0.86	0.86	0.86	0.03	0.03	0.03	**1.00**	**1.00**	**1.00**	0.86	0.88	0.87
GESSVDD-I-$\mathcal{S}$-max	0.86	0.86	0.86	0.02	0.02	0.02	0.98	**1.00**	0.99	0.89	0.85	0.87
GESSVDD-I-GR-max	0.83	0.85	0.84	0.14	0.06	0.10	**1.00**	**1.00**	**1.00**	0.89	0.83	0.86
GESSVDD-I-SR-max	0.85	0.85	0.85	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.90	0.86	0.88
GESSVDD-I-$\mathcal{S}$-min	0.8	0.84	0.82	0.03	0.00	0.01	**1.00**	**1.00**	**1.00**	0.81	0.80	0.81
GESSVDD-I-GR-min (ESSVDD)	0.86	0.83	0.85	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.85	0.80	0.83
GESSVDD-I-SR-min	0.83	0.89	0.86	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.88	0.92	0.90
GESSVDD-0-$\mathcal{S}$-max	**0.92**	**0.92**	**0.92**	0.03	0.08	0.06	**1.00**	**1.00**	**1.00**	**0.96**	0.92	**0.94**
GESSVDD-0-GR-max	**0.92**	0.89	0.91	0.06	0.11	0.09	**1.00**	**1.00**	**1.00**	0.95	**0.93**	**0.94**
GESSVDD-0-$\mathcal{S}$-min	0.84	0.88	0.86	0.07	0.00	0.04	**1.00**	**1.00**	**1.00**	0.91	0.85	0.88
GESSVDD-0-GR-min (SSVDD)	0.78	0.84	0.81	0.00	0.10	0.05	**1.00**	**1.00**	**1.00**	0.87	0.82	0.85
ESVDD	0.76	0.80	0.78	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.82	0.75	0.79
SVDD	0.59	0.48	0.54	**0.53**	**0.73**	**0.63**	0.67	0.26	0.47	0.54	0.49	0.51
OCSVM	0.90	0.85	0.87	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.87	0.85	0.86

*S3.1.2. TNR results **linear** data description*

Table 10: *TNR* results for **linear** data description

Dataset	Seeds				Qualitative bankruptcy			Somerville happiness		
Target class	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.67	0.50	0.84	0.67	0.62	0.59	0.60	0.17	0.15	0.16
GESSVDD-Sb-GR-max	0.63	0.95	0.90	0.83	0.57	0.53	0.55	0.33	0.37	0.35
GESSVDD-Sb-SR-max	0.73	0.94	0.87	0.85	0.62	0.28	0.45	**0.42**	0.20	0.31
GESSVDD-Sb-$\mathcal{S}$-min	0.73	0.97	0.79	0.83	0.49	0.50	0.50	0.37	0.21	0.29
GESSVDD-Sb-GR-min	0.72	0.94	0.88	0.85	0.54	**0.69**	0.62	0.31	0.35	0.33
GESSVDD-Sb-SR-min	0.72	0.91	0.87	0.84	0.58	0.24	0.41	0.26	0.17	0.22
GESSVDD-Sw-$\mathcal{S}$-max	0.75	0.94	0.90	0.86	0.86	0.00	0.43	0.31	0.25	0.28
GESSVDD-Sw-GR-max	0.84	0.96	0.95	0.92	0.57	0.01	0.29	0.26	0.25	0.26
GESSVDD-Sw-SR-max	0.67	0.92	0.78	0.79	0.78	0.04	0.41	0.34	0.18	0.26
GESSVDD-Sw-$\mathcal{S}$-min	0.65	0.94	0.90	0.83	0.85	0.04	0.45	0.32	0.18	0.25
GESSVDD-Sw-GR-min	0.79	0.80	0.95	0.85	0.71	0.04	0.38	0.37	0.20	0.28
GESSVDD-Sw-SR-min	0.66	0.91	0.93	0.83	0.85	0.02	0.44	0.23	0.19	0.21
GESSVDD-kNN-$\mathcal{S}$-max	0.82	0.89	0.90	0.87	0.73	0.08	0.40	0.40	0.21	0.30
GESSVDD-kNN-GR-max	0.79	0.94	0.90	0.88	0.69	0.21	0.45	0.31	0.44	0.37
GESSVDD-kNN-SR-max	0.82	0.95	0.93	0.90	0.72	0.09	0.40	0.36	0.36	0.36
GESSVDD-kNN-$\mathcal{S}$-min	0.73	0.92	0.89	0.85	0.73	0.20	0.46	0.31	0.33	0.32
GESSVDD-kNN-GR-min	0.85	0.94	0.92	0.91	0.81	0.60	**0.70**	0.26	**0.59**	0.43
GESSVDD-kNN-SR-min	0.85	0.98	0.91	0.91	0.75	0.27	0.51	0.26	0.25	0.26
GESSVDD-PCA-$\mathcal{S}$-max	0.77	0.90	0.94	0.87	0.85	0.09	0.47	0.31	0.21	0.26
GESSVDD-PCA-GR-max	0.87	0.91	0.87	0.88	0.93	0.03	0.48	0.32	0.20	0.26
GESSVDD-PCA-SR-max	0.85	0.90	0.93	0.89	0.91	0.07	0.49	0.26	0.24	0.25
GESSVDD-PCA-$\mathcal{S}$-min	0.70	0.91	0.93	0.85	0.88	0.03	0.46	0.34	0.18	0.26
GESSVDD-PCA-GR-min	0.86	0.91	0.80	0.86	0.85	0.03	0.44	0.28	0.23	0.26
GESSVDD-PCA-SR-min	0.71	0.92	0.91	0.85	0.87	0.09	0.48	0.25	0.23	0.24
GESSVDD-I-$\mathcal{S}$-max	0.76	0.90	0.94	0.87	0.87	0.09	0.48	0.29	0.22	0.26
GESSVDD-I-GR-max	0.90	0.93	0.94	0.92	0.94	0.02	0.48	0.35	0.24	0.30
GESSVDD-I-SR-max	0.79	0.89	0.94	0.87	0.91	0.01	0.46	0.32	0.24	0.28
GESSVDD-I-$\mathcal{S}$-min	0.71	0.92	0.94	0.86	0.91	0.06	0.48	0.34	0.15	0.24
GESSVDD-I-GR-min (ESSVDD)	0.89	0.96	0.88	0.91	0.90	0.03	0.47	0.34	0.19	0.26
GESSVDD-I-SR-min	0.75	0.90	0.94	0.86	0.89	0.08	0.49	0.26	0.20	0.23
GESSVDD-0-$\mathcal{S}$-max	0.87	0.96	0.91	0.91	0.81	0.00	0.41	0.22	0.23	0.22
GESSVDD-0-GR-max	0.83	0.96	0.91	0.90	0.89	0.00	0.44	0.28	0.22	0.25
GESSVDD-0-$\mathcal{S}$-min	0.65	0.91	0.92	0.83	0.85	0.09	0.47	0.29	0.17	0.23
GESSVDD-0-GR-min (SSVDD)	0.84	0.96	0.92	0.91	0.91	0.06	0.48	0.33	0.23	0.28
ESVDD	**0.91**	**0.99**	**0.98**	**0.96**	**0.99**	0.07	0.53	0.27	0.26	0.27
SVDD	0.83	0.95	0.93	0.90	0.93	0.00	0.47	0.25	0.16	0.20
OCSVM	0.44	0.80	0.57	0.60	0.39	0.39	0.39	0.37	0.50	**0.43**

Table 11: *TNR* results for **linear** data description

Dataset	Iris				Ionosphere			Sonar		
Target class	I-S	S-VC	S-V	Av.	I-B	I-G	Av.	S-R	S-M	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.99	0.85	0.85	0.90	0.24	0.44	0.34	0.44	0.72	0.58
GESSVDD-Sb-GR-max	0.94	0.86	0.92	0.91	0.23	0.82	0.52	0.34	0.34	0.34
GESSVDD-Sb-SR-max	0.85	0.94	0.83	0.87	0.44	0.54	0.49	0.20	0.46	0.33
GESSVDD-Sb-$\mathcal{S}$-min	0.99	0.83	0.84	0.89	0.38	0.59	0.48	0.53	0.76	0.65
GESSVDD-Sb-GR-min	**1.00**	0.85	0.94	0.93	0.18	0.82	0.50	0.30	0.52	0.41
GESSVDD-Sb-SR-min	0.85	0.76	0.87	0.83	0.44	0.45	0.44	0.29	0.30	0.29
GESSVDD-Sw-$\mathcal{S}$-max	0.98	0.89	0.92	0.93	0.18	0.79	0.48	0.50	0.27	0.38
GESSVDD-Sw-GR-max	0.93	0.92	0.83	0.89	0.07	0.86	0.46	0.38	0.30	0.34
GESSVDD-Sw-SR-max	0.99	0.95	0.92	0.95	0.15	0.77	0.46	0.44	0.25	0.35
GESSVDD-Sw-$\mathcal{S}$-min	0.98	0.93	0.92	0.94	0.11	0.77	0.44	0.55	0.37	0.46
GESSVDD-Sw-GR-min	**1.00**	0.95	0.73	0.89	0.03	0.87	0.45	0.38	0.28	0.33
GESSVDD-Sw-SR-min	0.99	0.92	0.87	0.93	0.10	0.75	0.42	0.37	0.26	0.31
GESSVDD-kNN-$\mathcal{S}$-max	0.99	0.85	0.89	0.91	0.12	0.81	0.46	0.48	0.17	0.33
GESSVDD-kNN-GR-max	**1.00**	0.76	0.73	0.83	0.62	0.85	0.74	0.42	0.31	0.36
GESSVDD-kNN-SR-max	**1.00**	0.90	0.94	0.95	0.15	0.80	0.47	0.42	0.23	0.32
GESSVDD-kNN-$\mathcal{S}$-min	0.99	0.91	0.81	0.90	0.15	0.81	0.48	0.48	0.17	0.33
GESSVDD-kNN-GR-min	**1.00**	0.91	0.93	0.95	**0.84**	0.92	**0.88**	0.42	0.43	0.43
GESSVDD-kNN-SR-min	**1.00**	0.89	0.88	0.92	0.26	0.80	0.53	0.42	0.26	0.34
GESSVDD-PCA-$\mathcal{S}$-max	**1.00**	0.95	0.93	0.96	0.20	0.79	0.49	0.51	0.12	0.32
GESSVDD-PCA-GR-max	0.99	0.94	**0.97**	0.97	0.04	0.88	0.46	0.44	0.28	0.36
GESSVDD-PCA-SR-max	**1.00**	0.95	0.94	0.96	0.16	0.79	0.48	0.61	0.16	0.38
GESSVDD-PCA-$\mathcal{S}$-min	**1.00**	0.92	0.96	0.96	0.12	0.82	0.47	0.51	0.35	0.43
GESSVDD-PCA-GR-min	0.97	0.84	0.87	0.90	0.14	0.86	0.50	0.41	0.43	0.42
GESSVDD-PCA-SR-min	**1.00**	0.95	0.96	0.97	0.13	0.80	0.47	0.54	0.17	0.35
GESSVDD-I-$\mathcal{S}$-max	**1.00**	0.95	0.93	0.96	0.17	0.82	0.50	0.51	0.16	0.33
GESSVDD-I-GR-max	**1.00**	0.89	0.83	0.91	0.05	0.87	0.46	0.44	0.33	0.39
GESSVDD-I-SR-max	**1.00**	0.95	0.82	0.92	0.14	0.79	0.47	0.53	0.16	0.35
GESSVDD-I-$\mathcal{S}$-min	**1.00**	0.95	0.80	0.92	0.17	0.82	0.50	0.51	0.17	0.34
GESSVDD-I-GR-min (ESSVDD)	0.98	0.79	0.85	0.87	0.15	0.87	0.51	0.44	0.42	0.43
GESSVDD-I-SR-min	**1.00**	0.95	**0.97**	0.98	0.14	0.79	0.46	0.52	0.17	0.34
GESSVDD-0-$\mathcal{S}$-max	0.98	0.91	0.93	0.94	0.03	0.41	0.22	0.22	0.16	0.19
GESSVDD-0-GR-max	**1.00**	0.93	0.90	0.94	0.04	0.68	0.36	0.28	0.22	0.25
GESSVDD-0-$\mathcal{S}$-min	0.94	0.91	0.85	0.90	0.07	0.72	0.40	0.28	0.14	0.21
GESSVDD-0-GR-min (SSVDD)	**1.00**	0.93	0.91	0.95	0.03	0.65	0.34	0.32	0.40	0.36
ESVDD	**1.00**	**1.00**	0.97	**0.99**	0.23	**0.94**	0.58	**1.00**	**1.00**	**1.00**
SVDD	**1.00**	0.93	0.91	0.94	0.00	0.79	0.39	0.38	0.39	0.38
OCSVM	0.79	0.47	0.63	0.63	0.51	0.48	0.49	0.44	0.38	0.41

Table 12: *TNR* results for **linear** data description

Dataset	MNIST											Liver		
Target class	0	1	2	3	4	5	6	7	8	9	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.31	0.70	0.74	0.69	0.69	0.40	0.47	0.43	0.38	0.56	0.54	0.16	0.46	0.31
GESSVDD-Sb-GR-max	0.45	0.53	0.08	0.70	0.47	0.22	0.64	0.89	0.46	0.28	0.47	0.54	0.44	0.49
GESSVDD-Sb-SR-max	0.49	0.89	0.51	0.70	0.62	0.77	0.22	0.62	0.69	0.77	0.63	0.52	0.21	0.37
GESSVDD-Sb-$\mathcal{S}$-min	0.31	0.70	0.74	0.69	0.69	0.40	0.47	0.43	0.38	0.56	0.54	**0.59**	0.53	**0.56**
GESSVDD-Sb-GR-min	0.45	0.53	0.08	0.70	0.47	0.22	0.64	0.89	0.46	0.28	0.47	0.42	0.43	0.43
GESSVDD-Sb-SR-min	0.49	0.89	0.51	0.70	0.62	0.77	0.22	0.62	0.69	0.77	0.63	0.34	**0.54**	0.44
GESSVDD-Sw-$\mathcal{S}$-max	0.10	0.76	0.11	0.16	0.19	0.12	0.22	0.37	0.18	0.38	0.26	0.15	0.14	0.14
GESSVDD-Sw-GR-max	0.22	0.62	0.12	0.22	0.24	0.12	0.38	0.40	0.18	0.25	0.27	0.16	0.15	0.15
GESSVDD-Sw-SR-max	0.17	0.60	0.16	0.22	0.40	0.35	0.50	0.49	0.36	0.54	0.38	0.17	0.15	0.16
GESSVDD-Sw-$\mathcal{S}$-min	0.10	0.76	0.11	0.16	0.19	0.12	0.22	0.37	0.18	0.38	0.26	0.18	0.17	0.18
GESSVDD-Sw-GR-min	0.22	0.62	0.12	0.22	0.24	0.12	0.38	0.40	0.18	0.25	0.27	0.15	0.15	0.15
GESSVDD-Sw-SR-min	0.17	0.60	0.16	0.22	0.40	0.35	0.50	0.49	0.36	0.54	0.38	0.19	0.16	0.18
GESSVDD-kNN-$\mathcal{S}$-max	0.36	0.76	0.47	0.49	0.52	0.42	0.19	0.21	0.31	0.36	0.41	0.21	0.17	0.19
GESSVDD-kNN-GR-max	0.27	0.71	0.14	0.70	0.44	0.37	0.49	0.31	0.55	0.34	0.43	0.36	0.43	0.40
GESSVDD-kNN-SR-max	0.19	0.80	0.66	0.21	0.45	0.44	0.49	0.51	0.36	0.63	0.47	0.23	0.25	0.24
GESSVDD-kNN-$\mathcal{S}$-min	0.36	0.76	0.47	0.49	0.52	0.42	0.19	0.21	0.31	0.36	0.41	0.51	0.36	0.44
GESSVDD-kNN-GR-min	0.27	0.71	0.14	0.70	0.44	0.37	0.49	0.31	0.55	0.34	0.43	0.58	0.24	0.41
GESSVDD-kNN-SR-min	0.19	0.80	0.66	0.21	0.45	0.44	0.49	0.51	0.36	0.63	0.47	0.39	0.32	0.36
GESSVDD-PCA-$\mathcal{S}$-max	0.09	0.71	0.13	0.26	0.11	0.15	0.15	0.47	0.20	0.39	0.27	0.24	0.20	0.22
GESSVDD-PCA-GR-max	0.14	0.74	0.13	0.22	0.24	0.12	0.20	0.37	0.17	0.26	0.26	0.18	0.17	0.18
GESSVDD-PCA-SR-max	0.17	0.82	0.10	0.22	0.37	0.28	0.51	0.44	0.29	0.54	0.37	0.13	0.19	0.16
GESSVDD-PCA-$\mathcal{S}$-min	0.09	0.71	0.13	0.26	0.11	0.15	0.15	0.47	0.20	0.39	0.27	0.20	0.21	0.20
GESSVDD-PCA-GR-min	0.14	0.74	0.13	0.22	0.24	0.12	0.20	0.37	0.17	0.26	0.26	0.19	0.17	0.18
GESSVDD-PCA-SR-min	0.17	0.82	0.10	0.22	0.37	0.28	0.51	0.44	0.29	0.54	0.37	0.14	0.18	0.16
GESSVDD-I-$\mathcal{S}$-max	0.17	0.70	0.14	0.27	0.21	0.15	0.24	0.48	0.36	0.35	0.31	0.20	0.20	0.20
GESSVDD-I-GR-max	0.18	0.70	0.10	0.22	0.23	0.12	0.39	0.43	0.17	0.25	0.28	0.18	0.20	0.19
GESSVDD-I-SR-max	0.16	0.79	0.10	0.18	0.37	0.28	0.52	0.43	0.29	0.56	0.37	0.21	0.18	0.20
GESSVDD-I-$\mathcal{S}$-min	0.17	0.70	0.14	0.27	0.21	0.15	0.24	0.48	0.36	0.35	0.31	0.19	0.20	0.20
GESSVDD-I-GR-min (ESSVDD)	0.18	0.70	0.10	0.22	0.23	0.12	0.39	0.43	0.17	0.25	0.28	0.14	0.17	0.16
GESSVDD-I-SR-min	0.16	0.79	0.10	0.18	0.37	0.28	0.52	0.43	0.29	0.56	0.37	0.21	0.18	0.19
GESSVDD-0-$\mathcal{S}$-max	0.24	0.51	0.12	0.16	0.24	0.12	0.34	0.28	0.18	0.27	0.25	0.13	0.12	0.13
GESSVDD-0-GR-max	0.17	0.66	0.09	0.16	0.20	0.10	0.33	0.28	0.17	0.20	0.24	0.11	0.13	0.12
GESSVDD-0-$\mathcal{S}$-min	0.24	0.51	0.12	0.16	0.24	0.12	0.34	0.28	0.18	0.27	0.25	0.12	0.16	0.14
GESSVDD-0-GR-min (SSVDD)	0.17	0.66	0.09	0.16	0.20	0.10	0.33	0.28	0.17	0.20	0.24	0.12	0.13	0.13
ESVDD	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.14	0.18	0.16
SVDD	0.34	0.86	0.23	0.28	0.42	0.18	0.55	0.57	0.29	0.40	0.41	0.45	0.45	0.45
OCSVM	0.43	0.65	0.53	0.50	0.57	0.54	0.41	0.62	0.51	0.58	0.53	0.19	0.14	0.17

Table 13: *TNR* results for **linear** data description

Dataset	Heart Clean train set Clean test set			Heart Clean train set Corrupted test set			Heart Corrupted train set Clean test set			Heart Corrupted train set Corrupted test set		
Target class	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.46	0.60	0.53	0.96	0.87	0.91	0.22	0.21	0.22	0.28	0.31	0.30
GESSVDD-Sb-GR-max	0.58	**0.83**	**0.70**	0.94	0.97	0.95	0.17	0.38	0.27	0.41	0.40	0.41
GESSVDD-Sb-SR-max	0.56	0.39	0.47	0.98	0.98	0.98	0.48	0.51	0.50	0.42	0.42	0.42
GESSVDD-Sb-$\mathcal{S}$-min	0.46	0.75	0.60	0.99	0.86	0.92	0.18	0.23	0.21	0.29	0.31	0.30
GESSVDD-Sb-GR-min	0.70	0.69	**0.70**	0.95	0.93	0.94	0.29	0.59	0.44	0.40	**0.63**	**0.52**
GESSVDD-Sb-SR-min	**0.75**	0.57	0.66	0.90	0.97	0.94	0.27	0.39	0.33	0.32	0.49	0.40
GESSVDD-Sw-$\mathcal{S}$-max	0.29	0.28	0.28	0.96	0.95	0.96	0.00	0.00	0.00	0.09	0.18	0.13
GESSVDD-Sw-GR-max	0.35	0.51	0.43	0.90	0.89	0.89	0.00	0.00	0.00	0.12	0.18	0.15
GESSVDD-Sw-SR-max	0.34	0.42	0.38	0.99	**1.00**	**1.00**	0.00	0.00	0.00	0.14	0.22	0.18
GESSVDD-Sw-$\mathcal{S}$-min	0.31	0.42	0.37	0.98	0.92	0.95	0.00	0.00	0.00	0.11	0.23	0.17
GESSVDD-Sw-GR-min	0.46	0.46	0.46	0.89	0.89	0.89	0.00	0.00	0.00	0.14	0.17	0.15
GESSVDD-Sw-SR-min	0.30	0.41	0.35	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	0.10	0.14	0.12
GESSVDD-kNN-$\mathcal{S}$-max	0.42	0.40	0.41	**1.00**	0.95	0.98	0.01	0.00	0.00	0.22	0.24	0.23
GESSVDD-kNN-GR-max	0.38	0.44	0.41	0.99	0.97	0.98	0.08	0.04	0.06	0.40	0.36	0.38
GESSVDD-kNN-SR-max	0.35	0.41	0.38	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	0.18	0.27	0.22
GESSVDD-kNN-$\mathcal{S}$-min	0.37	0.45	0.41	0.99	**1.00**	0.99	0.03	0.00	0.01	0.30	0.25	0.28
GESSVDD-kNN-GR-min	0.40	0.49	0.45	**1.00**	**1.00**	**1.00**	0.23	0.04	0.14	0.30	0.26	0.28
GESSVDD-kNN-SR-min	0.39	0.41	0.40	**1.00**	0.96	0.98	0.00	0.00	0.00	0.33	0.38	0.35
GESSVDD-PCA-$\mathcal{S}$-max	0.29	0.39	0.34	0.98	**1.00**	0.99	0.00	0.00	0.00	0.13	0.16	0.14
GESSVDD-PCA-GR-max	0.34	0.34	0.34	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	0.18	0.24	0.21
GESSVDD-PCA-SR-max	0.28	0.35	0.32	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	0.13	0.20	0.17
GESSVDD-PCA-$\mathcal{S}$-min	0.36	0.52	0.44	0.99	0.98	0.99	0.00	0.00	0.00	0.10	0.26	0.18
GESSVDD-PCA-GR-min	0.37	0.49	0.43	0.95	**1.00**	0.98	0.00	0.00	0.00	0.19	0.25	0.22
GESSVDD-PCA-SR-min	0.34	0.36	0.35	0.97	**1.00**	0.98	0.00	0.00	0.00	0.12	0.16	0.14
GESSVDD-I-$\mathcal{S}$-max	0.34	0.35	0.34	**1.00**	0.99	0.99	0.00	0.00	0.00	0.08	0.15	0.11
GESSVDD-I-GR-max	0.39	0.45	0.42	0.90	0.94	0.92	0.00	0.00	0.00	0.12	0.21	0.17
GESSVDD-I-SR-max	0.28	0.39	0.33	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	0.20	0.16	0.18
GESSVDD-I-$\mathcal{S}$-min	0.4	0.40	0.40	0.99	**1.00**	**1.00**	0.00	0.00	0.00	0.18	0.24	0.21
GESSVDD-I-GR-min (ESSVDD)	0.34	0.58	0.46	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	0.15	0.20	0.17
GESSVDD-I-SR-min	0.31	0.28	0.30	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	0.14	0.13	0.13
GESSVDD-0-$\mathcal{S}$-max	0.26	0.26	0.26	0.97	0.94	0.95	0.00	0.00	0.00	0.05	0.10	0.07
GESSVDD-0-GR-max	0.34	0.47	0.40	0.95	0.88	0.91	0.00	0.00	0.00	0.07	0.11	0.09
GESSVDD-0-$\mathcal{S}$-min	0.34	0.45	0.40	0.97	**1.00**	0.98	0.00	0.00	0.00	0.12	0.19	0.15
GESSVDD-0-GR-min (SSVDD)	0.46	0.47	0.46	**1.00**	0.90	0.95	0.00	0.00	0.00	0.15	0.20	0.18
ESVDD	0.42	0.48	0.45	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	0.23	0.32	0.27
SVDD	0.40	0.28	0.34	0.37	0.31	0.34	**0.54**	**0.73**	**0.63**	**0.44**	0.55	0.50
OCSVM	0.37	0.47	0.42	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	0.14	0.20	0.17

*S3.1.3. FPR results **linear** data description*

Table 14: *FPR* results for **linear** data description

Dataset	Seeds				Qualitative bankruptcy			Somerville happiness		
Target class	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.33	**0.50**	0.16	0.33	0.38	0.41	0.40	**0.83**	**0.85**	**0.84**
GESSVDD-Sb-GR-max	0.37	0.05	0.10	0.17	0.43	0.47	0.45	0.67	0.63	0.65
GESSVDD-Sb-SR-max	0.27	0.06	0.13	0.15	0.38	0.72	0.55	0.58	0.80	0.69
GESSVDD-Sb-$\mathcal{S}$-min	0.27	0.03	0.21	0.17	0.51	0.50	0.50	0.63	0.79	0.71
GESSVDD-Sb-GR-min	0.28	0.06	0.12	0.15	0.46	0.31	0.38	0.69	0.65	0.67
GESSVDD-Sb-SR-min	0.28	0.09	0.13	0.16	0.42	0.76	0.59	0.74	0.83	0.78
GESSVDD-Sw-$\mathcal{S}$-max	0.25	0.06	0.10	0.14	0.14	**1.00**	0.57	0.69	0.75	0.72
GESSVDD-Sw-GR-max	0.16	0.04	0.05	0.08	0.43	0.99	**0.71**	0.74	0.75	0.74
GESSVDD-Sw-SR-max	0.33	0.08	0.22	0.21	0.22	0.96	0.59	0.66	0.82	0.74
GESSVDD-Sw-$\mathcal{S}$-min	0.35	0.06	0.10	0.17	0.15	0.96	0.55	0.68	0.82	0.75
GESSVDD-Sw-GR-min	0.21	0.20	0.05	0.15	0.29	0.96	0.62	0.63	0.80	0.72
GESSVDD-Sw-SR-min	0.34	0.09	0.07	0.17	0.15	0.98	0.56	0.77	0.81	0.79
GESSVDD-kNN-$\mathcal{S}$-max	0.18	0.11	0.10	0.13	0.27	0.92	0.60	0.60	0.79	0.70
GESSVDD-kNN-GR-max	0.21	0.06	0.10	0.12	0.31	0.79	0.55	0.69	0.56	0.63
GESSVDD-kNN-SR-max	0.18	0.05	0.07	0.10	0.28	0.91	0.60	0.64	0.64	0.64
GESSVDD-kNN-$\mathcal{S}$-min	0.27	0.08	0.11	0.15	0.27	0.80	0.54	0.69	0.67	0.68
GESSVDD-kNN-GR-min	0.15	0.06	0.08	0.09	0.19	0.40	0.30	0.74	0.41	0.57
GESSVDD-kNN-SR-min	0.15	0.02	0.09	0.09	0.25	0.73	0.49	0.74	0.75	0.74
GESSVDD-PCA-$\mathcal{S}$-max	0.23	0.10	0.06	0.13	0.15	0.91	0.53	0.69	0.79	0.74
GESSVDD-PCA-GR-max	0.13	0.09	0.13	0.12	0.07	0.97	0.52	0.68	0.80	0.74
GESSVDD-PCA-SR-max	0.15	0.10	0.07	0.11	0.09	0.93	0.51	0.74	0.76	0.75
GESSVDD-PCA-$\mathcal{S}$-min	0.30	0.09	0.07	0.15	0.12	0.97	0.54	0.66	0.82	0.74
GESSVDD-PCA-GR-min	0.14	0.09	0.20	0.14	0.15	0.98	0.56	0.72	0.77	0.74
GESSVDD-PCA-SR-min	0.29	0.08	0.09	0.15	0.13	0.91	0.52	0.75	0.77	0.76
GESSVDD-I-$\mathcal{S}$-max	0.24	0.10	0.06	0.13	0.13	0.91	0.52	0.71	0.78	0.74
GESSVDD-I-GR-max	0.10	0.07	0.06	0.08	0.06	0.98	0.52	0.65	0.76	0.70
GESSVDD-I-SR-max	0.21	0.11	0.06	0.13	0.09	0.99	0.54	0.68	0.76	0.72
GESSVDD-I-$\mathcal{S}$-min	0.29	0.08	0.06	0.14	0.09	0.94	0.52	0.66	**0.85**	0.76
GESSVDD-I-GR-min (ESSVDD)	0.11	0.04	0.12	0.09	0.10	0.97	0.53	0.66	0.81	0.74
GESSVDD-I-SR-min	0.25	0.10	0.06	0.14	0.11	0.92	0.51	0.74	0.80	0.77
GESSVDD-0-$\mathcal{S}$-max	0.13	0.04	0.09	0.09	0.19	**1.00**	0.59	0.78	0.77	0.78
GESSVDD-0-GR-max	0.17	0.04	0.09	0.10	0.11	**1.00**	0.56	0.72	0.78	0.75
GESSVDD-0-$\mathcal{S}$-min	0.35	0.09	0.08	0.17	0.15	0.91	0.53	0.71	0.83	0.77
GESSVDD-0-GR-min (SSVDD)	0.16	0.04	0.08	0.09	0.09	0.94	0.52	0.67	0.77	0.72
ESVDD	0.09	0.01	0.02	0.04	0.01	0.93	0.47	0.73	0.74	0.73
SVDD	0.17	0.05	0.07	0.10	0.07	**1.00**	0.53	0.75	0.84	0.80
OCSVM	**0.56**	0.20	**0.43**	**0.40**	**0.61**	0.61	0.61	0.63	0.50	0.57

Table 15: *FPR* results for **linear** data description

Dataset	Iris				Ionosphere			Sonar		
Target class	I-S	S-VC	S-V	Av.	I-B	I-G	Av.	S-R	S-M	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.01	0.15	0.15	0.10	0.76	0.56	0.66	0.56	0.28	0.42
GESSVDD-Sb-GR-max	0.06	0.14	0.08	0.09	0.77	0.18	0.48	0.66	0.66	0.66
GESSVDD-Sb-SR-max	0.15	0.06	0.17	0.13	0.56	0.46	0.51	**0.80**	0.54	0.67
GESSVDD-Sb-$\mathcal{S}$-min	0.01	0.17	0.16	0.11	0.62	0.41	0.52	0.47	0.24	0.35
GESSVDD-Sb-GR-min	0.00	0.15	0.06	0.07	0.82	0.18	0.50	0.70	0.48	0.59
GESSVDD-Sb-SR-min	0.15	0.24	0.13	0.17	0.56	0.55	0.56	0.71	0.70	0.71
GESSVDD-Sw-$\mathcal{S}$-max	0.02	0.11	0.08	0.07	0.82	0.21	0.52	0.50	0.73	0.62
GESSVDD-Sw-GR-max	0.07	0.08	0.17	0.11	0.93	0.14	0.54	0.62	0.70	0.66
GESSVDD-Sw-SR-max	0.01	0.05	0.08	0.05	0.85	0.23	0.54	0.56	0.75	0.65
GESSVDD-Sw-$\mathcal{S}$-min	0.02	0.07	0.08	0.06	0.89	0.23	0.56	0.45	0.63	0.54
GESSVDD-Sw-GR-min	0.00	0.05	0.27	0.11	0.97	0.13	0.55	0.62	0.72	0.67
GESSVDD-Sw-SR-min	0.01	0.08	0.13	0.07	0.90	0.25	0.58	0.63	0.74	0.69
GESSVDD-kNN-$\mathcal{S}$-max	0.01	0.15	0.11	0.09	0.88	0.19	0.54	0.52	0.83	0.67
GESSVDD-kNN-GR-max	0.00	0.24	0.27	0.17	0.38	0.15	0.26	0.58	0.69	0.64
GESSVDD-kNN-SR-max	0.00	0.10	0.06	0.05	0.85	0.20	0.53	0.58	0.77	0.68
GESSVDD-kNN-$\mathcal{S}$-min	0.01	0.09	0.19	0.10	0.85	0.19	0.52	0.52	0.83	0.67
GESSVDD-kNN-GR-min	0.00	0.09	0.07	0.05	0.16	0.08	0.12	0.58	0.57	0.57
GESSVDD-kNN-SR-min	0.00	0.11	0.12	0.08	0.74	0.20	0.47	0.58	0.74	0.66
GESSVDD-PCA-$\mathcal{S}$-max	0.00	0.05	0.07	0.04	0.80	0.21	0.51	0.49	**0.88**	0.68
GESSVDD-PCA-GR-max	0.01	0.06	0.03	0.03	0.96	0.12	0.54	0.56	0.72	0.64
GESSVDD-PCA-SR-max	0.00	0.05	0.06	0.04	0.84	0.21	0.52	0.39	0.84	0.62
GESSVDD-PCA-$\mathcal{S}$-min	0.00	0.08	0.04	0.04	0.88	0.18	0.53	0.49	0.65	0.57
GESSVDD-PCA-GR-min	0.03	0.16	0.13	0.10	0.86	0.14	0.50	0.59	0.57	0.58
GESSVDD-PCA-SR-min	0.00	0.05	0.04	0.03	0.87	0.20	0.53	0.46	0.83	0.65
GESSVDD-I-$\mathcal{S}$-max	0.00	0.05	0.07	0.04	0.83	0.18	0.50	0.49	0.84	0.67
GESSVDD-I-GR-max	0.00	0.11	0.17	0.09	0.95	0.13	0.54	0.56	0.67	0.61
GESSVDD-I-SR-max	0.00	0.05	0.18	0.08	0.86	0.21	0.53	0.47	0.84	0.65
GESSVDD-I-$\mathcal{S}$-min	0.00	0.05	0.20	0.08	0.83	0.18	0.50	0.49	0.83	0.66
GESSVDD-I-GR-min (ESSVDD)	0.02	0.21	0.15	0.13	0.85	0.13	0.49	0.56	0.58	0.57
GESSVDD-I-SR-min	0.00	0.05	0.03	0.02	0.86	0.21	0.54	0.48	0.83	0.66
GESSVDD-0-$\mathcal{S}$-max	0.02	0.09	0.07	0.06	0.97	**0.59**	**0.78**	0.78	0.84	**0.81**
GESSVDD-0-GR-max	0.00	0.07	0.10	0.06	0.96	0.32	0.64	0.72	0.78	0.75
GESSVDD-0-$\mathcal{S}$-min	0.06	0.09	0.15	0.10	0.93	0.28	0.60	0.72	0.86	0.79
GESSVDD-0-GR-min (SSVDD)	0.00	0.07	0.09	0.05	0.97	0.35	0.66	0.68	0.60	0.64
ESVDD	0.00	0.00	0.03	0.01	0.77	0.06	0.42	0.00	0.00	0.00
SVDD	0.00	0.07	0.09	0.06	**1.00**	0.21	0.61	0.62	0.61	0.62
OCSVM	**0.21**	**0.53**	**0.37**	**0.37**	0.49	0.52	0.51	0.56	0.62	0.59

Table 16: *FPR* results for **linear** data description

Dataset	MNIST											Liver		
Target class	0	1	2	3	4	5	6	7	8	9	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.69	0.30	0.26	0.31	0.31	0.60	0.53	0.57	0.62	0.44	0.46	0.84	0.54	0.69
GESSVDD-Sb-GR-max	0.55	0.47	**0.92**	0.30	0.53	0.78	0.36	0.11	0.54	0.72	0.53	0.46	0.56	0.51
GESSVDD-Sb-SR-max	0.51	0.11	0.49	0.30	0.38	0.23	0.78	0.38	0.31	0.23	0.37	0.48	0.79	0.63
GESSVDD-Sb-$\mathcal{S}$-min	0.69	0.30	0.26	0.31	0.31	0.60	0.53	0.57	0.62	0.44	0.46	0.41	0.47	0.44
GESSVDD-Sb-GR-min	0.55	0.47	**0.92**	0.30	0.53	0.78	0.36	0.11	0.54	0.72	0.53	0.58	0.57	0.57
GESSVDD-Sb-SR-min	0.51	0.11	0.49	0.30	0.38	0.23	0.78	0.38	0.31	0.23	0.37	0.66	0.46	0.56
GESSVDD-Sw-$\mathcal{S}$-max	0.90	0.24	0.89	**0.84**	0.81	0.88	0.78	0.63	0.82	0.62	0.74	0.85	0.86	0.86
GESSVDD-Sw-GR-max	0.78	0.38	0.88	0.78	0.76	0.88	0.62	0.60	0.82	0.75	0.73	0.84	0.85	0.85
GESSVDD-Sw-SR-max	0.83	0.40	0.84	0.78	0.60	0.65	0.50	0.51	0.64	0.46	0.62	0.83	0.85	0.84
GESSVDD-Sw-$\mathcal{S}$-min	0.90	0.24	0.89	**0.84**	0.81	0.88	0.78	0.63	0.82	0.62	0.74	0.82	0.83	0.82
GESSVDD-Sw-GR-min	0.78	0.38	0.88	0.78	0.76	0.88	0.62	0.60	0.82	0.75	0.73	0.85	0.85	0.85
GESSVDD-Sw-SR-min	0.83	0.40	0.84	0.78	0.60	0.65	0.50	0.51	0.64	0.46	0.62	0.81	0.84	0.82
GESSVDD-kNN-$\mathcal{S}$-max	0.64	0.24	0.53	0.51	0.48	0.58	0.81	**0.79**	0.69	0.64	0.59	0.79	0.83	0.81
GESSVDD-kNN-GR-max	0.73	0.29	0.86	0.30	0.56	0.63	0.51	0.69	0.45	0.66	0.57	0.64	0.57	0.60
GESSVDD-kNN-SR-max	0.81	0.20	0.34	0.79	0.55	0.56	0.51	0.49	0.64	0.37	0.53	0.77	0.75	0.76
GESSVDD-kNN-$\mathcal{S}$-min	0.64	0.24	0.53	0.51	0.48	0.58	0.81	**0.79**	0.69	0.64	0.59	0.49	0.64	0.56
GESSVDD-kNN-GR-min	0.73	0.29	0.86	0.30	0.56	0.63	0.51	0.69	0.45	0.66	0.57	0.42	0.76	0.59
GESSVDD-kNN-SR-min	0.81	0.20	0.34	0.79	0.55	0.56	0.51	0.49	0.64	0.37	0.53	0.61	0.68	0.64
GESSVDD-PCA-$\mathcal{S}$-max	**0.91**	0.29	0.87	0.74	**0.89**	0.85	**0.85**	0.53	0.80	0.61	0.73	0.76	0.80	0.78
GESSVDD-PCA-GR-max	0.86	0.26	0.87	0.78	0.76	0.88	0.80	0.63	**0.83**	0.74	0.74	0.82	0.83	0.82
GESSVDD-PCA-SR-max	0.83	0.18	0.90	0.78	0.63	0.72	0.49	0.56	0.71	0.46	0.63	0.87	0.81	0.84
GESSVDD-PCA-$\mathcal{S}$-min	**0.91**	0.29	0.87	0.74	**0.89**	0.85	**0.85**	0.53	0.80	0.61	0.73	0.80	0.79	0.80
GESSVDD-PCA-GR-min	0.86	0.26	0.87	0.78	0.76	0.88	0.80	0.63	**0.83**	0.74	0.74	0.81	0.83	0.82
GESSVDD-PCA-SR-min	0.83	0.18	0.90	0.78	0.63	0.72	0.49	0.56	0.71	0.46	0.63	0.86	0.82	0.84
GESSVDD-I-$\mathcal{S}$-max	0.83	0.30	0.86	0.73	0.79	0.85	0.76	0.52	0.64	0.65	0.69	0.80	0.80	0.80
GESSVDD-I-GR-max	0.82	0.30	0.90	0.78	0.77	0.88	0.61	0.57	**0.83**	0.75	0.72	0.82	0.80	0.81
GESSVDD-I-SR-max	0.84	0.21	0.90	0.82	0.63	0.72	0.48	0.57	0.71	0.44	0.63	0.79	0.82	0.80
GESSVDD-I-$\mathcal{S}$-min	0.83	0.30	0.86	0.73	0.79	0.85	0.76	0.52	0.64	0.65	0.69	0.81	0.80	0.80
GESSVDD-I-GR-min (ESSVDD)	0.82	0.30	0.90	0.78	0.77	0.88	0.61	0.57	**0.83**	0.75	0.72	0.86	0.83	0.84
GESSVDD-I-SR-min	0.84	0.21	0.90	0.82	0.63	0.72	0.48	0.57	0.71	0.44	0.63	0.79	0.82	0.81
GESSVDD-0-$\mathcal{S}$-max	0.76	**0.49**	0.88	**0.84**	0.76	0.88	0.66	0.72	0.82	0.73	0.75	0.87	**0.88**	0.87
GESSVDD-0-GR-max	0.83	0.34	0.91	**0.84**	0.80	**0.90**	0.67	0.72	**0.83**	**0.80**	**0.76**	**0.89**	0.87	**0.88**
GESSVDD-0-$\mathcal{S}$-min	0.76	**0.49**	0.88	**0.84**	0.76	0.88	0.66	0.72	0.82	0.73	0.75	0.88	0.84	0.86
GESSVDD-0-GR-min (SSVDD)	0.83	0.34	0.91	**0.84**	0.80	**0.90**	0.67	0.72	**0.83**	**0.80**	**0.76**	0.88	0.87	0.87
ESVDD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.86	0.82	0.84
SVDD	0.66	0.14	0.77	0.72	0.58	0.82	0.45	0.43	0.71	0.60	0.59	0.55	0.55	0.55
OCSVM	0.57	0.35	0.47	0.50	0.43	0.46	0.59	0.38	0.49	0.42	0.47	0.81	0.86	0.83

Table 17: *FPR* results for **linear** data description

Dataset	Heart Clean train set Clean test set			Heart Clean train set Corrupted test set			Heart Corrupted train set Clean test set			Heart Corrupted train set Corrupted test set		
Target class	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.54	0.40	0.47	0.04	0.13	0.09	0.78	0.79	0.78	0.72	0.69	0.70
GESSVDD-Sb-GR-max	0.42	0.17	0.30	0.06	0.03	0.05	0.83	0.62	0.73	0.59	0.60	0.59
GESSVDD-Sb-SR-max	0.44	0.61	0.53	0.02	0.02	0.02	0.52	0.49	0.50	0.58	0.58	0.58
GESSVDD-Sb-$\mathcal{S}$-min	0.54	0.25	0.40	0.01	0.14	0.08	0.82	0.77	0.79	0.71	0.69	0.70
GESSVDD-Sb-GR-min	0.30	0.31	0.30	0.05	0.07	0.06	0.71	0.41	0.56	0.60	0.37	0.48
GESSVDD-Sb-SR-min	0.25	0.43	0.34	0.10	0.03	0.06	0.73	0.61	0.67	0.68	0.51	0.60
GESSVDD-Sw-$\mathcal{S}$-max	0.71	0.72	0.72	0.04	0.05	0.04	**1.00**	**1.00**	**1.00**	0.91	0.82	0.87
GESSVDD-Sw-GR-max	0.65	0.49	0.57	0.10	0.11	0.11	**1.00**	**1.00**	**1.00**	0.88	0.82	0.85
GESSVDD-Sw-SR-max	0.66	0.58	0.62	0.01	0.00	0.00	**1.00**	**1.00**	**1.00**	0.86	0.78	0.82
GESSVDD-Sw-$\mathcal{S}$-min	0.69	0.58	0.63	0.02	0.08	0.05	**1.00**	**1.00**	**1.00**	0.89	0.77	0.83
GESSVDD-Sw-GR-min	0.54	0.54	0.54	0.11	0.11	0.11	**1.00**	**1.00**	**1.00**	0.86	0.83	0.85
GESSVDD-Sw-SR-min	0.70	0.59	0.65	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.90	0.86	0.88
GESSVDD-kNN-$\mathcal{S}$-max	0.58	0.60	0.59	0.00	0.05	0.02	0.99	**1.00**	**1.00**	0.78	0.76	0.77
GESSVDD-kNN-GR-max	0.62	0.56	0.59	0.01	0.03	0.02	0.92	0.96	0.94	0.60	0.64	0.62
GESSVDD-kNN-SR-max	0.65	0.59	0.62	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.82	0.73	0.78
GESSVDD-kNN-$\mathcal{S}$-min	0.63	0.55	0.59	0.01	0.00	0.01	0.97	**1.00**	0.99	0.70	0.75	0.72
GESSVDD-kNN-GR-min	0.60	0.51	0.55	0.00	0.00	0.00	0.77	0.96	0.86	0.70	0.74	0.72
GESSVDD-kNN-SR-min	0.61	0.59	0.60	0.00	0.04	0.02	**1.00**	**1.00**	**1.00**	0.67	0.62	0.65
GESSVDD-PCA-$\mathcal{S}$-max	0.71	0.61	0.66	0.02	0.00	0.01	**1.00**	**1.00**	**1.00**	0.87	0.84	0.86
GESSVDD-PCA-GR-max	0.66	0.66	0.66	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.82	0.76	0.79
GESSVDD-PCA-SR-max	0.72	0.65	0.68	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.87	0.80	0.83
GESSVDD-PCA-$\mathcal{S}$-min	0.64	0.48	0.56	0.01	0.02	0.01	**1.00**	**1.00**	**1.00**	0.90	0.74	0.82
GESSVDD-PCA-GR-min	0.63	0.51	0.57	0.05	0.00	0.02	**1.00**	**1.00**	**1.00**	0.81	0.75	0.78
GESSVDD-PCA-SR-min	0.66	0.64	0.65	0.03	0.00	0.02	**1.00**	**1.00**	**1.00**	0.88	0.84	0.86
GESSVDD-I-$\mathcal{S}$-max	0.66	0.65	0.66	0.00	0.01	0.01	**1.00**	**1.00**	**1.00**	0.92	0.85	0.89
GESSVDD-I-GR-max	0.61	0.55	0.58	0.10	0.06	0.08	**1.00**	**1.00**	**1.00**	0.88	0.79	0.83
GESSVDD-I-SR-max	0.72	0.61	0.67	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.80	0.84	0.82
GESSVDD-I-$\mathcal{S}$-min	0.6	0.60	0.60	0.01	0.00	0.00	**1.00**	**1.00**	**1.00**	0.82	0.76	0.79
GESSVDD-I-GR-min (ESSVDD)	0.66	0.42	0.54	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.85	0.80	0.83
GESSVDD-I-SR-min	0.69	0.72	0.70	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.86	0.87	0.87
GESSVDD-0-$\mathcal{S}$-max	**0.74**	**0.74**	**0.74**	0.03	0.06	0.05	**1.00**	**1.00**	**1.00**	**0.95**	**0.90**	**0.93**
GESSVDD-0-GR-max	0.66	0.53	0.60	0.05	0.12	0.09	**1.00**	**1.00**	**1.00**	0.93	0.89	0.91
GESSVDD-0-$\mathcal{S}$-min	0.66	0.55	0.60	0.03	0.00	0.02	**1.00**	**1.00**	**1.00**	0.88	0.81	0.85
GESSVDD-0-GR-min (SSVDD)	0.54	0.53	0.54	0.00	0.10	0.05	**1.00**	**1.00**	**1.00**	0.85	0.80	0.82
ESVDD	0.58	0.52	0.55	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.77	0.68	0.73
SVDD	0.60	0.72	0.66	**0.63**	**0.69**	**0.66**	0.46	0.27	0.37	0.56	0.45	0.50
OCSVM	0.63	0.53	0.58	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.86	0.80	0.83

*S3.1.4. FNR result **linear** data description*

Table 18: *FNR* results for **linear** data description

Dataset	Seeds				Qualitative bankruptcy			Somerville happiness		
Target class	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.23	0.24	0.12	0.20	0.24	0.54	0.39	0.17	0.24	0.20
GESSVDD-Sb-GR-max	0.10	0.38	0.35	0.28	0.17	0.56	0.37	0.27	0.27	0.27
GESSVDD-Sb-SR-max	0.18	0.14	0.04	0.12	0.42	0.54	0.48	0.39	0.32	0.35
GESSVDD-Sb-$\mathcal{S}$-min	0.19	0.25	0.08	0.17	0.21	0.41	0.31	0.32	0.32	0.32
GESSVDD-Sb-GR-min	0.12	0.23	0.07	0.14	0.09	**0.65**	0.37	0.30	0.21	0.26
GESSVDD-Sb-SR-min	0.15	0.14	0.04	0.11	0.33	0.32	0.32	0.20	0.31	0.25
GESSVDD-Sw-$\mathcal{S}$-max	0.15	0.08	0.03	0.09	0.18	0.05	0.11	0.20	0.20	0.20
GESSVDD-Sw-GR-max	0.10	0.14	0.07	0.10	0.08	0.09	0.08	0.24	0.20	0.22
GESSVDD-Sw-SR-max	0.10	0.10	0.04	0.08	0.11	0.08	0.10	0.25	0.26	0.26
GESSVDD-Sw-$\mathcal{S}$-min	0.12	0.09	0.04	0.08	0.13	0.11	0.12	0.20	0.27	0.24
GESSVDD-Sw-GR-min	0.13	0.12	0.10	0.12	0.12	0.09	0.11	0.27	0.20	0.23
GESSVDD-Sw-SR-min	0.16	0.06	0.04	0.09	0.07	0.12	0.10	0.25	0.23	0.24
GESSVDD-kNN-$\mathcal{S}$-max	0.21	0.09	0.04	0.11	0.16	0.14	0.15	0.29	0.34	0.31
GESSVDD-kNN-GR-max	0.11	0.09	0.04	0.08	0.12	0.29	0.21	0.23	0.39	0.31
GESSVDD-kNN-SR-max	0.24	0.10	0.07	0.14	0.11	0.24	0.18	0.29	0.44	0.36
GESSVDD-kNN-$\mathcal{S}$-min	0.23	0.15	0.04	0.14	0.16	0.28	0.22	0.26	0.39	0.33
GESSVDD-kNN-GR-min	0.18	0.07	0.03	0.09	0.19	0.62	0.40	0.23	**0.46**	0.35
GESSVDD-kNN-SR-min	0.26	0.10	0.05	0.14	0.14	0.34	0.24	0.26	0.33	0.29
GESSVDD-PCA-$\mathcal{S}$-max	0.13	0.05	0.07	0.08	0.14	0.20	0.17	0.28	0.26	0.27
GESSVDD-PCA-GR-max	0.15	0.11	0.08	0.11	0.05	0.12	0.09	0.23	0.20	0.21
GESSVDD-PCA-SR-max	0.21	0.05	0.05	0.10	0.09	0.19	0.14	0.23	0.27	0.25
GESSVDD-PCA-$\mathcal{S}$-min	0.15	0.07	0.08	0.10	0.17	0.19	0.18	0.23	0.29	0.26
GESSVDD-PCA-GR-min	0.14	0.18	0.06	0.13	0.05	0.16	0.10	0.19	0.22	0.21
GESSVDD-PCA-SR-min	0.13	0.06	0.04	0.08	0.07	0.13	0.10	0.23	0.33	0.28
GESSVDD-I-$\mathcal{S}$-max	0.19	0.05	0.08	0.10	0.16	0.17	0.16	0.28	0.21	0.24
GESSVDD-I-GR-max	0.23	0.09	0.08	0.13	0.07	0.19	0.13	0.28	0.15	0.21
GESSVDD-I-SR-max	0.16	0.05	0.08	0.10	0.08	0.10	0.09	0.18	0.32	0.25
GESSVDD-I-$\mathcal{S}$-min	0.16	0.08	0.09	0.11	0.17	0.14	0.15	0.27	0.22	0.25
GESSVDD-I-GR-min (ESSVDD)	0.14	0.12	0.07	0.11	0.10	0.18	0.14	0.23	0.18	0.20
GESSVDD-I-SR-min	0.17	0.05	0.08	0.10	0.07	0.17	0.12	0.27	0.22	0.25
GESSVDD-0-$\mathcal{S}$-max	0.13	0.10	0.06	0.10	0.20	0.10	0.15	0.20	0.13	0.16
GESSVDD-0-GR-max	0.13	0.10	0.03	0.09	0.14	0.11	0.13	0.22	0.14	0.18
GESSVDD-0-$\mathcal{S}$-min	0.16	0.06	0.10	0.11	0.20	0.15	0.17	0.25	0.38	0.32
GESSVDD-0-GR-min (SSVDD)	0.13	0.10	0.03	0.09	0.10	0.15	0.12	0.19	0.14	0.16
ESVDD	0.30	0.24	0.21	0.25	0.07	0.21	0.14	0.32	0.34	0.33
SVDD	0.13	0.10	0.06	0.10	0.04	0.12	0.08	0.25	0.16	0.21
OCSVM	**0.41**	**0.40**	**0.50**	**0.44**	**0.63**	0.46	**0.54**	**0.42**	0.38	**0.40**

Table 19: *FNR* results for **linear** data description

Dataset	Iris				Ionosphere			Sonar		
Target class	I-S	S-VC	S-V	Av.	I-B	I-G	Av.	S-R	S-M	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.15	0.12	0.21	0.16	0.32	0.32	0.32	0.38	0.69	0.54
GESSVDD-Sb-GR-max	0.13	0.11	0.15	0.13	0.20	0.04	0.12	0.14	0.25	0.20
GESSVDD-Sb-SR-max	0.13	0.17	0.15	0.15	0.42	**0.79**	**0.61**	0.41	0.31	0.36
GESSVDD-Sb-$\mathcal{S}$-min	0.17	0.16	0.21	0.18	0.29	0.51	0.40	0.48	0.73	0.61
GESSVDD-Sb-GR-min	0.11	0.11	0.15	0.12	0.19	0.04	0.11	0.14	0.42	0.28
GESSVDD-Sb-SR-min	0.13	0.39	0.16	0.23	0.44	0.61	0.52	0.50	0.62	0.56
GESSVDD-Sw-$\mathcal{S}$-max	0.08	0.08	0.16	0.11	0.23	0.03	0.13	0.44	0.24	0.34
GESSVDD-Sw-GR-max	0.11	0.11	0.11	0.11	0.09	0.06	0.07	0.18	0.16	0.17
GESSVDD-Sw-SR-max	0.08	0.21	0.21	0.17	0.35	0.04	0.20	0.39	0.12	0.25
GESSVDD-Sw-$\mathcal{S}$-min	0.08	0.15	0.17	0.13	0.24	0.03	0.13	0.46	0.35	0.40
GESSVDD-Sw-GR-min	0.08	0.11	0.12	0.10	0.17	0.05	0.11	0.19	0.18	0.18
GESSVDD-Sw-SR-min	0.08	0.20	0.16	0.15	0.20	0.04	0.12	0.31	0.15	0.23
GESSVDD-kNN-$\mathcal{S}$-max	0.12	0.09	0.16	0.12	0.33	0.06	0.19	0.47	0.18	0.32
GESSVDD-kNN-GR-max	0.09	0.07	0.15	0.10	0.61	0.07	0.34	0.32	0.23	0.28
GESSVDD-kNN-SR-max	0.07	0.17	0.17	0.14	0.36	0.05	0.21	0.44	0.18	0.31
GESSVDD-kNN-$\mathcal{S}$-min	0.12	0.12	0.16	0.13	0.30	0.06	0.18	0.47	0.18	0.32
GESSVDD-kNN-GR-min	0.07	0.12	0.11	0.10	**0.72**	0.07	0.40	0.28	0.23	0.25
GESSVDD-kNN-SR-min	0.05	0.16	0.12	0.11	0.42	0.05	0.24	0.44	0.21	0.33
GESSVDD-PCA-$\mathcal{S}$-max	0.03	0.09	0.21	0.11	0.38	0.03	0.20	0.47	0.08	0.28
GESSVDD-PCA-GR-max	0.12	0.19	0.21	0.17	0.16	0.06	0.11	0.22	0.21	0.21
GESSVDD-PCA-SR-max	0.03	0.09	0.24	0.12	0.22	0.05	0.14	0.57	0.10	0.33
GESSVDD-PCA-$\mathcal{S}$-min	0.07	0.05	0.24	0.12	0.33	0.05	0.19	0.47	0.28	0.37
GESSVDD-PCA-GR-min	0.09	0.16	0.17	0.14	0.14	0.07	0.11	0.34	0.20	0.27
GESSVDD-PCA-SR-min	0.03	0.12	0.25	0.13	0.33	0.06	0.19	0.52	0.10	0.31
GESSVDD-I-$\mathcal{S}$-max	0.03	0.09	0.21	0.11	0.30	0.05	0.18	0.49	0.10	0.29
GESSVDD-I-GR-max	0.09	0.11	0.12	0.11	0.17	0.06	0.12	0.32	0.21	0.27
GESSVDD-I-SR-max	0.03	0.08	0.23	0.11	0.31	0.03	0.17	0.51	0.10	0.30
GESSVDD-I-$\mathcal{S}$-min	0.09	0.08	0.16	0.11	0.35	0.05	0.20	0.49	0.13	0.31
GESSVDD-I-GR-min (ESSVDD)	0.12	0.11	0.05	0.09	0.15	0.07	0.11	0.37	0.19	0.28
GESSVDD-I-SR-min	0.04	0.08	0.21	0.11	0.23	0.03	0.13	0.47	0.13	0.30
GESSVDD-0-$\mathcal{S}$-max	0.11	0.13	0.12	0.12	0.10	0.05	0.07	0.11	0.08	0.09
GESSVDD-0-GR-max	0.07	0.11	0.12	0.10	0.11	0.06	0.08	0.10	0.07	0.09
GESSVDD-0-$\mathcal{S}$-min	0.09	0.12	0.09	0.10	0.11	0.08	0.09	0.17	0.22	0.20
GESSVDD-0-GR-min (SSVDD)	0.08	0.11	0.12	0.10	0.10	0.06	0.08	0.17	0.21	0.19
ESVDD	0.21	0.27	0.23	0.24	0.50	0.17	0.34	**1.00**	**0.99**	**1.00**
SVDD	0.15	0.12	0.12	0.13	0.14	0.07	0.10	0.25	0.16	0.20
OCSVM	**0.45**	**0.43**	**0.49**	**0.46**	0.51	0.46	0.48	0.48	0.46	0.47

Table 20: *FNR* results for **linear** data description

Dataset	MNIST											Liver		
Target class	0	1	2	3	4	5	6	7	8	9	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.11	0.31	0.67	0.70	0.54	0.35	0.22	0.36	0.21	0.35	0.38	0.08	0.37	0.22
GESSVDD-Sb-GR-max	0.16	0.18	0.02	0.57	0.31	0.18	0.59	0.88	0.36	0.10	0.33	0.49	0.40	0.44
GESSVDD-Sb-SR-max	0.36	**0.53**	0.38	0.68	0.31	0.63	0.85	0.31	0.55	0.49	0.51	0.47	0.20	0.33
GESSVDD-Sb-$\mathcal{S}$-min	0.11	0.31	0.67	0.70	0.54	0.35	0.22	0.36	0.21	0.35	0.38	**0.54**	0.44	**0.49**
GESSVDD-Sb-GR-min	0.16	0.18	0.02	0.57	0.31	0.18	0.59	0.88	0.36	0.10	0.33	0.38	0.41	0.40
GESSVDD-Sb-SR-min	0.36	**0.53**	0.38	0.68	0.31	0.63	0.85	0.31	0.55	0.49	0.51	0.33	0.43	0.38
GESSVDD-Sw-$\mathcal{S}$-max	0.03	0.01	0.02	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.08	0.07	0.08
GESSVDD-Sw-GR-max	0.01	0.01	0.02	0.03	0.02	0.02	0.02	0.01	0.01	0.02	0.02	0.11	0.14	0.13
GESSVDD-Sw-SR-max	0.04	0.01	0.04	0.04	0.07	0.11	0.09	0.07	0.08	0.10	0.06	0.10	0.13	0.12
GESSVDD-Sw-$\mathcal{S}$-min	0.03	0.01	0.02	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.12	0.13	0.12
GESSVDD-Sw-GR-min	0.01	0.01	0.02	0.03	0.02	0.02	0.02	0.01	0.01	0.02	0.02	0.08	0.13	0.10
GESSVDD-Sw-SR-min	0.04	0.01	0.04	0.04	0.07	0.11	0.09	0.07	0.08	0.10	0.06	0.09	0.13	0.11
GESSVDD-kNN-$\mathcal{S}$-max	0.28	0.01	0.45	0.42	0.49	0.32	0.20	0.01	0.19	0.02	0.24	0.14	0.11	0.13
GESSVDD-kNN-GR-max	0.10	0.01	0.13	0.63	0.16	0.45	0.02	0.01	0.43	0.02	0.20	0.37	0.48	0.43
GESSVDD-kNN-SR-max	0.08	0.09	0.61	0.05	0.14	0.21	0.14	0.10	0.18	0.19	0.18	0.22	0.24	0.23
GESSVDD-kNN-$\mathcal{S}$-min	0.28	0.01	0.45	0.42	0.49	0.32	0.20	0.01	0.19	0.02	0.24	0.47	0.28	0.37
GESSVDD-kNN-GR-min	0.10	0.01	0.13	0.63	0.16	0.45	0.02	0.01	0.43	0.02	0.20	0.53	0.25	0.39
GESSVDD-kNN-SR-min	0.08	0.09	0.61	0.05	0.14	0.21	0.14	0.10	0.18	0.19	0.18	0.30	0.29	0.30
GESSVDD-PCA-$\mathcal{S}$-max	0.07	0.01	0.02	0.16	0.04	0.03	0.01	0.02	0.02	0.02	0.04	0.12	0.10	0.11
GESSVDD-PCA-GR-max	0.02	0.01	0.02	0.03	0.02	0.03	0.02	0.01	0.02	0.02	0.02	0.10	0.09	0.10
GESSVDD-PCA-SR-max	0.04	0.06	0.03	0.04	0.06	0.09	0.09	0.05	0.07	0.08	0.06	0.11	0.12	0.11
GESSVDD-PCA-$\mathcal{S}$-min	0.07	0.01	0.02	0.16	0.04	0.03	0.01	0.02	0.02	0.02	0.04	0.14	0.15	0.14
GESSVDD-PCA-GR-min	0.02	0.01	0.02	0.03	0.02	0.03	0.02	0.01	0.02	0.02	0.02	0.08	0.13	0.10
GESSVDD-PCA-SR-min	0.04	0.06	0.03	0.04	0.06	0.09	0.09	0.05	0.07	0.08	0.06	0.07	0.11	0.09
GESSVDD-I-$\mathcal{S}$-max	0.09	0.01	0.04	0.18	0.41	0.02	0.02	0.03	0.27	0.02	0.11	0.11	0.10	0.10
GESSVDD-I-GR-max	0.01	0.01	0.02	0.03	0.02	0.03	0.02	0.02	0.02	0.02	0.02	0.09	0.13	0.11
GESSVDD-I-SR-max	0.04	0.06	0.03	0.04	0.06	0.09	0.09	0.04	0.07	0.11	0.06	0.11	0.11	0.11
GESSVDD-I-$\mathcal{S}$-min	0.09	0.01	0.04	0.18	0.41	0.02	0.02	0.03	0.27	0.02	0.11	0.16	0.15	0.15
GESSVDD-I-GR-min (ESSVDD)	0.01	0.01	0.02	0.03	0.02	0.03	0.02	0.02	0.02	0.02	0.02	0.09	0.12	0.10
GESSVDD-I-SR-min	0.04	0.06	0.03	0.04	0.06	0.09	0.09	0.04	0.07	0.11	0.06	0.16	0.09	0.13
GESSVDD-0-$\mathcal{S}$-max	0.02	0.00	0.02	0.02	0.02	0.03	0.02	0.01	0.01	0.02	0.02	0.10	0.09	0.10
GESSVDD-0-GR-max	0.02	0.01	0.02	0.02	0.01	0.02	0.02	0.01	0.01	0.02	0.01	0.10	0.12	0.11
GESSVDD-0-$\mathcal{S}$-min	0.02	0.00	0.02	0.02	0.02	0.03	0.02	0.01	0.01	0.02	0.02	0.09	0.10	0.10
GESSVDD-0-GR-min (SSVDD)	0.02	0.01	0.02	0.02	0.01	0.02	0.02	0.01	0.01	0.02	0.01	0.10	0.11	0.11
ESVDD	**1.00**	0.34	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**0.95**	**1.00**	**0.97**	**0.93**	0.11	0.10	0.10
SVDD	0.04	0.02	0.04	0.03	0.03	0.04	0.03	0.04	0.03	0.03	0.03	0.40	**0.51**	0.45
OCSVM	0.48	**0.53**	0.50	0.51	0.54	0.51	0.51	0.48	0.52	0.55	0.51	0.10	0.10	0.10

Table 21: *FNR* results for **linear** data description

Dataset	Heart Clean train set Clean test set			Heart Clean train set Corrupted test set			Heart Corrupted train set Clean test set			Heart Corrupted train set Corrupted test set		
Target class	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.26	0.43	0.35	0.94	0.92	0.93	0.32	0.42	0.37	0.24	0.31	0.27
GESSVDD-Sb-GR-max	0.36	**0.56**	**0.46**	0.93	0.92	0.93	0.20	0.34	0.27	0.44	0.43	0.44
GESSVDD-Sb-SR-max	0.35	0.43	0.39	0.96	0.96	0.96	**0.40**	0.60	0.50	0.37	0.45	0.41
GESSVDD-Sb-$\mathcal{S}$-min	0.28	0.52	0.40	0.92	0.88	0.90	0.30	0.36	0.33	0.30	0.25	0.27
GESSVDD-Sb-GR-min	0.42	0.30	0.36	0.95	0.91	0.93	0.32	0.67	0.50	**0.46**	**0.61**	**0.54**
GESSVDD-Sb-SR-min	**0.52**	0.37	0.45	0.88	0.96	0.92	0.25	0.47	0.36	0.30	0.52	0.41
GESSVDD-Sw-$\mathcal{S}$-max	0.08	0.10	0.09	0.94	0.96	0.95	0.00	0.00	0.00	0.06	0.08	0.07
GESSVDD-Sw-GR-max	0.11	0.14	0.12	0.86	0.88	0.87	0.00	0.00	0.00	0.10	0.10	0.10
GESSVDD-Sw-SR-max	0.16	0.17	0.17	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	0.16	0.16	0.16
GESSVDD-Sw-$\mathcal{S}$-min	0.14	0.10	0.12	0.97	0.95	0.96	0.01	0.00	0.00	0.12	0.18	0.15
GESSVDD-Sw-GR-min	0.17	0.18	0.17	0.85	0.87	0.86	0.00	0.00	0.00	0.16	0.18	0.17
GESSVDD-Sw-SR-min	0.14	0.15	0.15	**1.00**	**1.00**	**1.00**	0.05	0.00	0.02	0.10	0.16	0.13
GESSVDD-kNN-$\mathcal{S}$-max	0.25	0.20	0.23	0.99	0.97	0.98	0.06	0.00	0.03	0.14	0.18	0.16
GESSVDD-kNN-GR-max	0.22	0.18	0.20	0.96	0.93	0.95	0.06	0.04	0.05	0.38	0.37	0.38
GESSVDD-kNN-SR-max	0.26	0.17	0.22	**1.00**	0.99	**1.00**	0.00	0.00	0.00	0.12	0.27	0.20
GESSVDD-kNN-$\mathcal{S}$-min	0.20	0.25	0.22	0.98	**1.00**	0.99	0.00	0.00	0.00	0.24	0.24	0.24
GESSVDD-kNN-GR-min	0.25	0.22	0.24	0.98	**1.00**	0.99	0.12	0.04	0.08	0.37	0.23	0.30
GESSVDD-kNN-SR-min	0.25	0.17	0.21	**1.00**	0.95	0.97	0.00	0.00	0.00	0.30	0.36	0.33
GESSVDD-PCA-$\mathcal{S}$-max	0.12	0.12	0.12	0.99	**1.00**	0.99	0.01	0.00	0.00	0.13	0.07	0.10
GESSVDD-PCA-GR-max	0.17	0.15	0.16	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	0.15	0.17	0.16
GESSVDD-PCA-SR-max	0.15	0.16	0.15	**1.00**	0.98	0.99	0.00	0.00	0.00	0.12	0.20	0.16
GESSVDD-PCA-$\mathcal{S}$-min	0.15	0.16	0.16	**1.00**	0.95	0.97	0.00	0.00	0.00	0.10	0.21	0.16
GESSVDD-PCA-GR-min	0.18	0.17	0.17	0.90	**1.00**	0.95	0.00	0.00	0.00	0.18	0.20	0.19
GESSVDD-PCA-SR-min	0.14	0.14	0.14	0.97	0.97	0.97	0.00	0.00	0.00	0.14	0.12	0.13
GESSVDD-I-$\mathcal{S}$-max	0.14	0.14	0.14	0.98	0.98	0.98	0.02	0.00	0.01	0.11	0.15	0.13
GESSVDD-I-GR-max	0.17	0.15	0.16	0.86	0.94	0.90	0.00	0.00	0.00	0.11	0.17	0.14
GESSVDD-I-SR-max	0.15	0.15	0.15	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	0.10	0.14	0.12
GESSVDD-I-$\mathcal{S}$-min	0.2	0.16	0.18	0.97	**1.00**	0.99	0.00	0.00	0.00	0.19	0.20	0.19
GESSVDD-I-GR-min (ESSVDD)	0.14	0.17	0.15	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	0.15	0.20	0.17
GESSVDD-I-SR-min	0.17	0.11	0.14	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	0.12	0.08	0.10
GESSVDD-0-$\mathcal{S}$-max	0.08	0.08	0.08	0.97	0.92	0.94	0.00	0.00	0.00	0.04	0.08	0.06
GESSVDD-0-GR-max	0.08	0.11	0.09	0.94	0.89	0.91	0.00	0.00	0.00	0.05	0.07	0.06
GESSVDD-0-$\mathcal{S}$-min	0.16	0.12	0.14	0.93	**1.00**	0.96	0.00	0.00	0.00	0.09	0.15	0.12
GESSVDD-0-GR-min (SSVDD)	0.22	0.16	0.19	**1.00**	0.90	0.95	0.00	0.00	0.00	0.13	0.18	0.15
ESVDD	0.24	0.20	0.22	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	0.18	0.25	0.21
SVDD	0.41	0.52	**0.46**	0.47	0.27	0.37	0.33	**0.74**	**0.53**	**0.46**	0.51	0.49
OCSVM	0.10	0.15	0.13	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	0.13	0.15	0.14

*S3.2. TPR TNR FPR FNR results for **non-linear** data description*

*S3.2.1. TPR results **non-linear** data description*

Table 22: *TPR* results for **non-linear** data description

Dataset	Seeds				Qualitative bankruptcy			Somerville happiness		
Target class	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.86	0.94	0.89	0.90	0.76	0.46	0.61	**0.78**	0.77	0.78
GESSVDD-Sb-GR-max	0.78	0.64	0.92	0.78	0.93	0.59	0.76	0.74	0.66	0.70
GESSVDD-Sb-SR-max	0.78	0.74	0.97	0.83	0.70	0.45	0.58	0.47	0.49	0.48
GESSVDD-Sb-$\mathcal{S}$-min	0.86	0.90	0.81	0.85	0.76	0.51	0.63	0.62	0.72	0.67
GESSVDD-Sb-GR-min	0.78	0.58	0.94	0.77	0.94	0.54	0.74	0.74	0.64	0.69
GESSVDD-Sb-SR-min	0.81	0.78	0.95	0.85	0.70	0.40	0.55	0.60	0.55	0.57
GESSVDD-Sw-$\mathcal{S}$-max	0.89	0.95	**0.99**	0.94	0.88	**0.89**	0.89	0.71	0.71	0.71
GESSVDD-Sw-GR-max	0.88	0.90	0.94	0.90	0.80	0.89	0.85	0.71	0.71	0.71
GESSVDD-Sw-SR-max	0.86	0.93	0.88	0.89	0.91	0.58	0.75	0.68	0.61	0.64
GESSVDD-Sw-$\mathcal{S}$-min	0.78	0.95	**0.99**	0.91	0.92	0.86	0.89	0.63	0.57	0.60
GESSVDD-Sw-GR-min	0.89	0.93	0.91	0.91	0.83	0.89	0.86	0.74	0.72	0.73
GESSVDD-Sw-SR-min	0.88	0.89	0.90	0.89	0.91	0.62	0.76	0.57	0.63	0.60
GESSVDD-kNN-$\mathcal{S}$-max	0.93	0.90	0.94	0.93	0.90	0.78	0.84	0.64	0.58	0.61
GESSVDD-kNN-GR-max	0.84	0.90	0.91	0.88	0.91	0.77	0.84	0.71	0.58	0.65
GESSVDD-kNN-SR-max	0.91	0.89	**0.99**	0.93	0.92	0.70	0.81	0.60	0.57	0.58
GESSVDD-kNN-$\mathcal{S}$-min	0.93	0.90	0.95	0.93	0.83	0.72	0.78	0.67	0.57	0.62
GESSVDD-kNN-GR-min	0.81	0.93	0.96	0.90	0.87	0.62	0.75	0.75	0.52	0.63
GESSVDD-kNN-SR-min	0.92	0.91	0.93	0.92	0.87	0.63	0.75	0.54	0.63	0.59
GESSVDD-PCA-$\mathcal{S}$-max	0.89	0.92	0.95	0.92	0.56	0.67	0.61	0.71	0.73	0.72
GESSVDD-PCA-GR-max	0.85	0.93	0.94	0.91	0.90	0.86	0.88	0.66	0.68	0.67
GESSVDD-PCA-SR-max	0.90	0.73	0.95	0.86	0.87	0.60	0.73	0.69	0.65	0.67
GESSVDD-PCA-$\mathcal{S}$-min	0.93	0.95	0.95	0.95	0.83	0.64	0.73	0.55	0.60	0.57
GESSVDD-PCA-GR-min	0.87	0.81	0.95	0.88	0.94	0.84	0.89	0.68	0.66	0.67
GESSVDD-PCA-SR-min	0.93	0.95	0.94	0.94	0.93	0.72	0.83	0.70	0.67	0.68
GESSVDD-I-$\mathcal{S}$-max	**0.94**	0.94	0.78	0.89	0.87	0.63	0.75	0.72	0.62	0.67
GESSVDD-I-GR-max	0.83	0.90	0.92	0.88	0.94	0.85	**0.90**	0.71	0.71	0.71
GESSVDD-I-SR-max	**0.94**	0.94	0.98	0.96	0.85	0.83	0.84	0.69	0.56	0.62
GESSVDD-I-$\mathcal{S}$-min	**0.94**	0.93	0.96	0.95	0.79	0.62	0.70	0.59	0.63	0.61
GESSVDD-I-GR-min (ESSVDD)	0.84	0.90	0.93	0.89	**0.97**	0.81	0.89	0.71	0.58	0.65
GESSVDD-I-SR-min	**0.94**	**0.96**	0.97	**0.96**	0.93	0.65	0.79	0.66	0.68	0.67
GESSVDD-0-$\mathcal{S}$-max	0.88	0.90	0.93	0.90	0.68	0.52	0.60	0.73	**0.84**	**0.79**
GESSVDD-0-GR-max	0.86	0.90	0.94	0.90	0.95	0.43	0.69	0.64	0.73	0.68
GESSVDD-0-$\mathcal{S}$-min	0.87	0.90	0.90	0.89	0.66	0.54	0.60	0.57	0.54	0.56
GESSVDD-0-GR-min (SSVDD)	0.87	0.92	0.95	0.91	0.96	0.57	0.76	0.61	0.45	0.53
ESVDD	0.71	0.79	0.78	0.76	0.00	0.00	0.00	0.00	0.51	0.25
SVDD	0.86	0.89	0.96	0.90	0.68	0.27	0.47	0.76	0.34	0.55
OCSVM	0.41	0.52	0.48	0.47	0.72	0.50	0.61	0.49	0.55	0.52
GESVDD-PCA	0.82	0.89	0.91	0.87	0.91	0.64	0.78	0.63	0.59	0.61
GESVDD-Sw	0.75	0.89	0.90	0.85	0.91	0.62	0.77	0.44	0.45	0.45
GESVDD-kNN	0.78	0.87	0.92	0.86	0.87	0.65	0.76	0.66	0.33	0.49
GESVM-PCA	0.84	0.84	0.90	0.86	0.91	0.37	0.64	0.61	0.48	0.55
GESVM-Sw	0.85	0.85	0.87	0.85	0.87	0.42	0.65	0.57	0.54	0.55
GESVM-kNN	0.82	0.83	0.85	0.83	0.86	0.31	0.58	0.63	0.53	0.58

Table 23: *TPR* results for **non-linear** data description

Dataset	Iris				Ionosphere			Sonar		
Target class	I-S	S-VC	S-V	Av.	I-B	I-G	Av.	S-R	S-M	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.92	0.83	0.87	0.87	0.58	0.73	0.66	0.63	0.64	0.64
GESSVDD-Sb-GR-max	0.91	0.83	0.85	0.86	0.80	0.96	0.88	0.86	0.69	0.78
GESSVDD-Sb-SR-max	0.91	0.85	0.81	0.86	0.81	0.27	0.54	0.76	0.42	0.59
GESSVDD-Sb-$\mathcal{S}$-min	0.92	0.87	0.89	0.89	0.57	0.53	0.55	0.59	0.48	0.54
GESSVDD-Sb-GR-min	0.91	0.84	**0.91**	0.88	0.76	0.96	0.86	0.86	0.50	0.68
GESSVDD-Sb-SR-min	0.95	0.79	0.69	0.81	0.64	0.27	0.46	0.78	0.24	0.51
GESSVDD-Sw-$\mathcal{S}$-max	0.89	0.87	0.85	0.87	0.54	0.95	0.75	0.68	0.77	0.72
GESSVDD-Sw-GR-max	0.88	0.83	0.88	0.86	**0.98**	0.94	**0.96**	0.82	0.81	**0.81**
GESSVDD-Sw-SR-max	0.95	0.87	0.85	0.89	0.44	0.95	0.69	0.81	0.45	0.63
GESSVDD-Sw-$\mathcal{S}$-min	0.84	0.88	0.84	0.85	0.55	0.95	0.75	0.59	0.67	0.63
GESSVDD-Sw-GR-min	0.91	0.87	0.83	0.87	0.97	0.93	0.95	0.76	**0.85**	0.80
GESSVDD-Sw-SR-min	0.93	0.87	0.88	0.89	0.19	0.92	0.55	0.81	0.73	0.77
GESSVDD-kNN-$\mathcal{S}$-max	0.88	0.87	0.79	0.84	0.79	0.97	0.88	0.71	0.72	0.71
GESSVDD-kNN-GR-max	0.89	0.92	0.84	0.88	0.81	0.92	0.86	0.69	0.73	0.71
GESSVDD-kNN-SR-max	0.89	0.92	0.77	0.86	0.79	0.96	0.88	0.73	0.59	0.66
GESSVDD-kNN-$\mathcal{S}$-min	0.89	0.88	0.77	0.85	0.68	0.97	0.82	0.61	0.68	0.64
GESSVDD-kNN-GR-min	0.91	0.80	0.88	0.86	0.97	0.94	0.96	0.75	0.75	0.75
GESSVDD-kNN-SR-min	0.89	0.92	0.73	0.85	0.66	0.96	0.81	0.66	0.60	0.63
GESSVDD-PCA-$\mathcal{S}$-max	0.88	0.93	0.83	0.88	0.62	0.93	0.77	0.82	0.76	0.79
GESSVDD-PCA-GR-max	0.88	0.84	0.80	0.84	0.91	0.95	0.93	0.79	0.82	0.81
GESSVDD-PCA-SR-max	0.88	**0.95**	0.81	0.88	0.36	0.89	0.63	0.74	0.48	0.61
GESSVDD-PCA-$\mathcal{S}$-min	0.88	**0.95**	0.88	**0.90**	0.44	0.92	0.68	0.57	0.48	0.53
GESSVDD-PCA-GR-min	0.88	0.87	0.83	0.86	0.91	0.93	0.92	0.72	0.79	0.76
GESSVDD-PCA-SR-min	0.88	**0.95**	0.81	0.88	0.52	0.87	0.70	0.55	0.64	0.60
GESSVDD-I-$\mathcal{S}$-max	**0.96**	0.89	0.83	0.89	0.70	0.98	0.84	0.69	0.70	0.70
GESSVDD-I-GR-max	0.92	0.83	0.88	0.88	0.93	0.94	0.93	0.70	0.79	0.75
GESSVDD-I-SR-max	**0.96**	0.91	0.83	0.90	0.68	0.98	0.83	0.70	0.47	0.58
GESSVDD-I-$\mathcal{S}$-min	0.89	0.91	0.83	0.88	0.66	**0.98**	0.82	0.59	0.49	0.54
GESSVDD-I-GR-min (ESSVDD)	0.92	0.85	0.87	0.88	0.97	0.93	0.95	0.66	0.82	0.74
GESSVDD-I-SR-min	0.91	0.91	0.85	0.89	0.63	**0.98**	0.81	0.61	0.59	0.60
GESSVDD-0-$\mathcal{S}$-max	0.88	0.91	0.87	0.88	0.86	0.79	0.83	**0.87**	0.51	0.69
GESSVDD-0-GR-max	0.92	0.89	0.88	0.90	0.68	0.93	0.81	0.71	0.76	0.73
GESSVDD-0-$\mathcal{S}$-min	0.92	0.88	0.87	0.89	0.78	0.91	0.85	0.67	0.63	0.65
GESSVDD-0-GR-min (SSVDD)	0.89	0.91	0.85	0.88	0.96	0.92	0.94	0.82	0.72	0.77
ESVDD	0.57	0.71	0.73	0.67	0.69	0.83	0.76	0.42	0.29	0.36
SVDD	0.84	0.89	0.85	0.86	0.24	0.93	0.58	0.76	0.78	0.77
OCSVM	0.53	0.43	0.35	0.44	0.44	0.41	0.42	0.49	0.60	0.54
GESVDD-PCA	0.71	0.88	0.83	0.80	0.31	0.89	0.60	0.68	0.73	0.70
GESVDD-Sw	0.80	0.80	0.84	0.81	0.25	0.90	0.57	0.68	0.65	0.67
GESVDD-kNN	0.71	0.87	0.84	0.80	0.18	0.89	0.54	0.61	0.64	0.63
GESVM-PCA	0.81	0.84	0.87	0.84	0.50	0.88	0.69	0.64	0.64	0.64
GESVM-Sw	0.79	0.93	0.83	0.85	0.47	0.88	0.67	0.65	0.52	0.58
GESVM-kNN	0.80	0.84	0.84	0.83	0.53	0.86	0.69	0.66	0.56	0.61

Table 24: *TPR* results for **non-linear** data description

Dataset	MNIST											Liver		
Target class	0	1	2	3	4	5	6	7	8	9	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.16	0.16	0.06	0.08	0.39	0.29	0.31	0.00	0.12	0.09	0.17	0.37	0.35	0.36
GESSVDD-Sb-GR-max	0.62	0.32	0.48	0.41	0.20	0.33	0.21	0.38	0.11	0.58	0.36	0.29	0.35	0.32
GESSVDD-Sb-SR-max	0.01	0.59	0.26	0.67	0.05	0.04	0.10	0.26	0.11	0.47	0.26	0.29	0.27	0.28
GESSVDD-Sb-$\mathcal{S}$-min	0.79	0.08	0.00	0.10	0.69	0.82	0.80	0.00	0.21	0.20	0.37	0.34	0.33	0.33
GESSVDD-Sb-GR-min	0.62	0.18	0.67	0.24	0.04	0.61	0.71	0.18	**1.00**	0.38	0.46	0.35	0.24	0.30
GESSVDD-Sb-SR-min	0.01	0.59	0.26	0.67	0.05	0.04	0.10	0.26	0.11	0.47	0.26	0.34	0.29	0.31
GESSVDD-Sw-$\mathcal{S}$-max	0.07	0.05	0.25	0.14	0.28	0.11	0.18	0.33	0.05	0.10	0.16	0.45	0.32	0.38
GESSVDD-Sw-GR-max	0.25	0.22	0.09	0.44	0.12	0.52	0.23	0.43	0.20	0.17	0.27	0.36	0.40	0.38
GESSVDD-Sw-SR-max	0.44	0.80	0.20	0.55	0.79	0.32	0.84	0.79	0.15	0.48	0.53	0.39	0.47	0.43
GESSVDD-Sw-$\mathcal{S}$-min	0.00	0.01	0.86	0.22	0.48	0.04	0.21	0.44	0.00	0.01	0.23	0.24	0.33	0.28
GESSVDD-Sw-GR-min	0.52	0.48	0.55	0.27	0.02	0.39	0.45	0.34	0.85	0.36	0.42	0.33	0.28	0.30
GESSVDD-Sw-SR-min	0.44	0.80	0.20	0.55	0.79	0.32	0.84	0.79	0.15	0.48	0.53	0.31	0.40	0.36
GESSVDD-kNN-$\mathcal{S}$-max	0.08	0.23	0.53	0.06	0.55	0.31	0.31	0.26	0.27	0.25	0.28	0.39	0.32	0.36
GESSVDD-kNN-GR-max	0.62	0.32	0.49	0.49	0.24	0.47	0.18	0.40	0.37	0.55	0.41	0.25	0.45	0.35
GESSVDD-kNN-SR-max	0.89	0.80	0.04	0.81	0.78	0.52	0.79	0.79	0.66	0.76	0.68	0.41	0.42	0.41
GESSVDD-kNN-$\mathcal{S}$-min	0.01	0.40	0.70	0.00	0.77	0.75	0.39	0.43	0.24	0.12	0.38	0.33	0.31	0.32
GESSVDD-kNN-GR-min	0.40	0.18	0.67	0.60	0.26	0.34	0.43	0.22	0.31	0.35	0.38	0.39	0.39	0.39
GESSVDD-kNN-SR-min	0.89	0.80	0.04	0.81	0.78	0.52	0.79	0.79	0.66	0.76	0.68	0.40	0.44	0.42
GESSVDD-PCA-$\mathcal{S}$-max	0.19	0.33	0.15	0.17	0.18	0.31	0.05	0.54	0.24	0.17	0.23	0.28	0.37	0.32
GESSVDD-PCA-GR-max	0.45	0.25	0.19	0.29	0.27	0.27	0.41	0.08	0.28	0.28	0.28	0.39	0.34	0.37
GESSVDD-PCA-SR-max	0.96	0.80	0.95	0.67	0.78	0.80	0.84	0.09	0.42	0.79	0.71	0.36	0.27	0.32
GESSVDD-PCA-$\mathcal{S}$-min	0.41	0.99	0.21	0.18	0.60	0.32	0.01	0.88	0.14	0.05	0.38	0.34	0.28	0.31
GESSVDD-PCA-GR-min	0.60	0.68	0.29	0.50	0.37	0.30	0.51	0.01	0.77	0.21	0.42	0.31	0.34	0.32
GESSVDD-PCA-SR-min	0.96	0.80	0.95	0.67	0.78	0.80	0.84	0.09	0.42	0.79	0.71	0.32	0.26	0.29
GESSVDD-I-$\mathcal{S}$-max	0.30	0.33	0.17	0.17	0.30	0.32	0.42	0.54	0.21	0.17	0.29	0.42	0.27	0.34
GESSVDD-I-GR-max	0.36	0.34	0.18	0.09	0.19	0.52	0.46	0.43	0.36	0.21	0.31	0.35	0.42	0.39
GESSVDD-I-SR-max	0.96	0.80	0.39	0.78	0.78	0.27	0.84	0.90	0.79	0.79	0.73	0.35	0.42	0.39
GESSVDD-I-$\mathcal{S}$-min	0.32	0.99	0.24	0.13	0.30	0.33	0.37	0.88	0.16	0.05	0.38	0.39	0.26	0.32
GESSVDD-I-GR-min (ESSVDD)	**0.99**	0.95	0.21	0.97	0.22	0.42	0.24	0.33	0.63	0.26	0.52	0.40	0.49	0.45
GESSVDD-I-SR-min	0.96	0.80	0.39	0.78	0.78	0.27	0.84	0.90	0.79	0.79	0.73	0.30	0.37	0.33
GESSVDD-0-$\mathcal{S}$-max	0.22	0.09	0.06	0.28	0.23	0.40	0.08	0.28	0.11	0.05	0.18	0.36	0.37	0.36
GESSVDD-0-GR-max	0.60	0.34	0.48	0.39	0.43	0.49	0.43	0.35	0.44	0.17	0.41	0.36	0.44	0.40
GESSVDD-0-$\mathcal{S}$-min	0.39	**1.00**	**0.97**	0.13	0.07	0.23	0.01	0.58	0.02	0.00	0.34	0.37	0.37	0.37
GESSVDD-0-GR-min (SSVDD)	0.89	0.29	0.67	0.24	0.27	0.32	0.24	0.41	0.30	0.06	0.37	0.37	0.39	0.38
ESVDD	0.95	0.96	**0.97**	**0.99**	**0.98**	**0.99**	**0.97**	**0.98**	0.99	**0.98**	**0.98**	0.43	0.54	0.49
SVDD	0.03	0.02	0.46	0.03	0.02	0.03	0.03	0.04	0.02	0.03	0.07	0.49	0.40	0.45
OCSVM	0.38	0.48	0.46	0.25	0.39	0.54	0.47	0.34	0.41	0.52	0.42	0.27	0.08	0.17
GESVDD-PCA	0.90	0.95	0.77	0.74	0.79	0.67	0.88	0.84	0.68	0.87	0.81	**0.51**	0.49	0.50
GESVDD-Sw	0.89	0.95	0.78	0.74	0.79	0.66	0.00	0.86	0.63	0.87	0.72	**0.51**	0.52	0.51
GESVDD-kNN	0.89	0.95	0.77	0.73	0.79	0.64	0.85	0.86	0.71	0.87	0.81	**0.51**	0.52	**0.52**
GESVM-PCA	0.90	0.90	0.81	0.71	0.89	0.58	0.83	0.85	0.75	0.90	0.81	0.50	**0.55**	**0.52**
GESVM-Sw	0.92	0.90	0.72	0.71	0.81	0.72	0.80	0.89	0.68	0.82	0.80	0.50	0.51	0.51
GESVM-kNN	0.91	0.90	0.81	0.70	0.90	0.58	0.83	0.84	0.66	0.87	0.80	0.44	0.47	0.46

Table 25: *TPR* results for **non-linear** data description in the proposed framework with added noise

Dataset	Heart Clean train set Clean test set			Heart Clean train set Corrupted test set			Heart Corrupted train set Clean test set			Heart Corrupted train set Corrupted test set		
Target class	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.29	0.29	0.29	0.42	0.29	0.36	0.13	0.05	0.09	0.35	0.40	0.37
GESSVDD-Sb-GR-max	0.34	0.34	0.34	0.35	0.23	0.29	0.33	0.31	0.32	0.27	0.45	0.36
GESSVDD-Sb-SR-max	0.32	0.32	0.32	0.22	0.24	0.23	0.30	0.22	0.26	0.43	0.44	0.44
GESSVDD-Sb-$\mathcal{S}$-min	0.41	0.41	0.41	**0.44**	0.19	0.32	0.19	0.19	0.19	0.35	0.20	0.28
GESSVDD-Sb-GR-min	0.45	0.45	0.45	0.32	0.18	0.25	0.21	0.32	0.26	0.33	0.33	0.33
GESSVDD-Sb-SR-min	0.46	0.46	0.46	0.25	0.13	0.19	0.36	0.16	0.26	0.43	0.44	0.43
GESSVDD-Sw-$\mathcal{S}$-max	0.43	0.43	0.43	0.00	0.12	0.06	0.13	0.00	0.06	0.36	0.34	0.35
GESSVDD-Sw-GR-max	0.16	0.16	0.16	0.36	**0.39**	**0.38**	0.07	0.11	0.09	0.42	0.39	0.41
GESSVDD-Sw-SR-max	0.43	0.43	0.43	0.11	0.06	0.08	0.00	0.06	0.03	0.43	0.41	0.42
GESSVDD-Sw-$\mathcal{S}$-min	0.34	0.34	0.34	0.00	0.16	0.08	0.00	0.15	0.08	0.44	0.43	0.43
GESSVDD-Sw-GR-min	0.40	0.40	0.40	0.18	0.29	0.24	0.04	0.05	0.05	0.45	0.44	0.45
GESSVDD-Sw-SR-min	0.44	0.44	0.44	0.03	0.19	0.11	0.00	0.00	0.00	0.41	0.33	0.37
GESSVDD-kNN-$\mathcal{S}$-max	0.44	0.44	0.44	0.00	0.00	0.00	0.14	0.04	0.09	0.45	0.45	0.45
GESSVDD-kNN-GR-max	0.43	0.43	0.43	0.28	0.19	0.24	0.12	0.00	0.06	0.44	0.38	0.41
GESSVDD-kNN-SR-max	0.42	0.42	0.42	0.04	0.03	0.03	0.00	0.00	0.00	0.47	0.40	0.44
GESSVDD-kNN-$\mathcal{S}$-min	0.49	0.49	0.49	0.03	0.00	0.02	0.03	0.00	0.02	0.46	0.43	0.44
GESSVDD-kNN-GR-min	0.47	0.47	0.47	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.43	0.46
GESSVDD-kNN-SR-min	0.47	0.47	0.47	0.04	0.07	0.06	0.06	0.00	0.03	0.45	0.38	0.41
GESSVDD-PCA-$\mathcal{S}$-max	0.37	0.37	0.37	0.00	0.03	0.01	0.00	0.00	0.00	0.49	0.36	0.42
GESSVDD-PCA-GR-max	0.47	0.47	0.47	0.30	0.34	0.32	0.13	0.16	0.14	0.45	0.36	0.40
GESSVDD-PCA-SR-max	0.46	0.46	0.46	0.19	0.05	0.12	0.06	0.09	0.08	0.34	0.41	0.37
GESSVDD-PCA-$\mathcal{S}$-min	0.24	0.24	0.24	0.06	0.11	0.08	0.00	0.00	0.00	0.43	0.39	0.41
GESSVDD-PCA-GR-min	0.30	0.30	0.30	0.28	0.22	0.25	0.00	0.04	0.02	0.46	0.47	0.47
GESSVDD-PCA-SR-min	0.44	0.44	0.44	0.10	0.07	0.09	0.03	0.00	0.02	0.37	0.41	0.39
GESSVDD-I-$\mathcal{S}$-max	0.38	0.38	0.38	0.06	0.03	0.04	0.04	0.00	0.02	0.45	0.38	0.42
GESSVDD-I-GR-max	0.42	0.42	0.42	0.37	0.35	0.36	0.03	0.05	0.04	0.48	0.31	0.39
GESSVDD-I-SR-max	0.49	0.49	0.49	0.13	0.03	0.08	0.00	0.00	0.00	0.43	0.44	0.43
GESSVDD-I-$\mathcal{S}$-min	0.44	0.44	0.44	0.00	0.25	0.12	0.08	0.00	0.04	0.36	0.39	0.38
GESSVDD-I-GR-min (ESSVDD)	0.38	0.38	0.38	0.37	0.15	0.26	0.00	0.11	0.06	0.38	0.46	0.42
GESSVDD-I-SR-min	0.46	0.46	0.46	0.15	0.14	0.14	0.00	0.08	0.04	0.42	0.45	0.44
GESSVDD-0-$\mathcal{S}$-max	0.35	0.35	0.35	0.24	0.21	0.22	0.26	0.32	0.29	0.39	0.36	0.38
GESSVDD-0-GR-max	0.45	0.45	0.45	0.06	0.22	0.14	0.06	0.44	0.25	0.46	0.36	0.41
GESSVDD-0-$\mathcal{S}$-min	0.32	0.32	0.32	0.14	0.16	0.15	0.36	0.09	0.23	0.34	0.40	0.37
GESSVDD-0-GR-min (SSVDD)	0.53	0.53	0.53	0.12	0.24	0.18	0.00	0.09	0.04	0.37	0.47	0.42
ESVDD	0.34	0.34	0.34	0.00	0.00	0.00	0.00	0.03	0.02	0.49	0.52	0.51
SVDD	0.53	0.53	0.53	0.07	0.15	0.11	0.22	0.07	0.15	0.47	0.45	0.46
OCSVM	0.20	0.20	0.20	0.00	0.00	0.00	0.00	0.07	0.03	0.50	0.49	0.50
GESVDD-PCA	0.68	0.68	0.68	0.00	0.00	0.00	0.00	0.00	0.00	0.51	**0.53**	**0.52**
GESVDD-Sw	0.68	0.68	0.68	0.00	0.00	0.00	0.00	0.00	0.00	**0.53**	0.52	**0.52**
GESVDD-kNN	**0.70**	**0.70**	**0.70**	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.52	0.50
GESVM-PCA	0.66	0.66	0.66	0.00	0.00	0.00	0.00	0.00	0.00	0.49	0.49	0.49
GESVM-Sw	0.67	0.67	0.67	0.00	0.00	0.00	0.00	0.05	0.03	0.50	0.48	0.49
GESVM-kNN	0.60	0.60	0.60	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.45	0.45	0.45

*S3.2.2. TNR results **non-linear** data description*

Table 26: *TNR* results for **non-linear** data description

Dataset	Seeds				Qualitative bankruptcy			Somerville happiness		
Target class	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.69	0.89	0.88	0.82	0.50	0.70	0.60	0.19	0.16	0.17
GESSVDD-Sb-GR-max	0.89	0.95	0.86	0.90	0.74	0.53	0.63	0.25	0.34	0.30
GESSVDD-Sb-SR-max	0.78	0.90	0.89	0.85	0.48	0.41	0.45	0.61	0.24	0.43
GESSVDD-Sb-$\mathcal{S}$-min	0.74	0.92	0.88	0.85	0.53	0.61	0.57	0.45	0.26	0.36
GESSVDD-Sb-GR-min	0.89	0.94	0.88	0.90	0.69	0.62	0.66	0.37	0.36	0.36
GESSVDD-Sb-SR-min	0.65	0.96	0.89	0.83	0.48	0.61	0.55	0.41	0.19	0.30
GESSVDD-Sw-$\mathcal{S}$-max	0.83	0.84	0.86	0.84	0.81	0.17	0.49	0.41	0.23	0.32
GESSVDD-Sw-GR-max	0.90	0.83	0.89	0.87	0.81	0.05	0.43	0.33	0.27	0.30
GESSVDD-Sw-SR-max	0.79	0.86	0.90	0.85	0.80	0.65	0.73	0.44	0.27	0.36
GESSVDD-Sw-$\mathcal{S}$-min	0.81	0.84	0.86	0.84	0.81	0.17	0.49	0.52	0.44	0.48
GESSVDD-Sw-GR-min	0.90	0.95	0.92	0.92	0.73	0.01	0.37	0.34	0.26	0.30
GESSVDD-Sw-SR-min	0.76	0.87	0.92	0.85	0.96	0.83	0.90	0.55	0.31	0.43
GESSVDD-kNN-$\mathcal{S}$-max	0.82	0.90	0.86	0.86	0.63	0.54	0.59	0.42	0.23	0.32
GESSVDD-kNN-GR-max	0.81	0.93	0.88	0.87	0.85	0.26	0.55	0.45	0.36	0.40
GESSVDD-kNN-SR-max	0.82	0.96	0.92	0.90	0.72	0.74	0.73	0.41	0.40	0.41
GESSVDD-kNN-$\mathcal{S}$-min	0.82	0.86	0.92	0.87	0.80	0.85	0.82	0.37	0.44	0.41
GESSVDD-kNN-GR-min	0.87	0.94	0.87	0.90	0.83	0.38	0.60	0.49	0.36	0.43
GESSVDD-kNN-SR-min	0.82	0.88	0.94	0.88	0.72	0.90	0.81	0.49	0.37	0.43
GESSVDD-PCA-$\mathcal{S}$-max	0.80	0.90	0.93	0.88	0.84	0.73	0.79	0.39	0.30	0.35
GESSVDD-PCA-GR-max	0.75	0.87	0.87	0.83	0.90	0.07	0.48	0.47	0.21	0.34
GESSVDD-PCA-SR-max	0.78	0.95	0.93	0.89	0.94	0.47	0.71	0.36	0.34	0.35
GESSVDD-PCA-$\mathcal{S}$-min	0.77	0.94	0.93	0.88	0.88	0.70	0.79	0.57	0.27	0.42
GESSVDD-PCA-GR-min	0.86	0.92	0.92	0.90	0.91	0.06	0.49	0.43	0.24	0.34
GESSVDD-PCA-SR-min	0.77	0.86	0.93	0.85	0.92	0.59	0.76	0.39	0.32	0.36
GESSVDD-I-$\mathcal{S}$-max	0.78	0.91	0.93	0.87	0.81	0.60	0.70	0.26	0.27	0.27
GESSVDD-I-GR-max	0.83	0.93	0.90	0.89	0.87	0.04	0.45	0.39	0.27	0.33
GESSVDD-I-SR-max	0.78	0.91	0.90	0.86	0.88	0.46	0.67	0.37	0.39	0.38
GESSVDD-I-$\mathcal{S}$-min	0.78	0.96	0.90	0.88	0.93	0.57	0.75	0.44	0.28	0.36
GESSVDD-I-GR-min (ESSVDD)	0.84	0.91	0.88	0.88	0.88	0.10	0.49	0.49	0.28	0.39
GESSVDD-I-SR-min	0.78	0.95	0.90	0.87	0.77	0.60	0.68	0.49	0.23	0.36
GESSVDD-0-$\mathcal{S}$-max	0.83	0.94	0.93	0.90	0.79	0.45	0.62	0.29	0.25	0.27
GESSVDD-0-GR-max	0.84	0.96	0.93	0.91	0.92	0.57	0.74	0.33	0.37	0.35
GESSVDD-0-$\mathcal{S}$-min	0.85	0.92	0.93	0.90	0.79	0.54	0.67	0.46	0.44	0.45
GESSVDD-0-GR-min (SSVDD)	0.87	0.96	0.92	0.92	0.92	0.53	0.72	0.41	0.50	0.46
ESVDD	**0.93**	**0.98**	**0.98**	**0.96**	**1.00**	**1.00**	**1.00**	**1.00**	0.43	**0.71**
SVDD	0.85	0.94	0.93	0.91	0.51	0.78	0.65	0.24	0.60	0.42
OCSVM	0.67	0.77	0.54	0.66	0.30	0.68	0.49	0.47	0.45	0.46
GESVDD-PCA	0.89	0.97	0.95	0.94	0.98	0.17	0.57	0.43	0.41	0.42
GESVDD-Sw	0.91	0.97	0.96	0.95	0.97	0.17	0.57	0.61	0.57	0.59
GESVDD-kNN	0.91	0.97	0.95	0.95	0.83	0.16	0.49	0.42	**0.69**	0.55
GESVM-PCA	0.87	0.97	0.97	0.93	1.00	0.19	0.59	0.47	0.49	0.48
GESVM-Sw	0.85	0.96	0.95	0.92	**1.00**	0.15	0.57	0.55	0.36	0.45
GESVM-kNN	0.88	**0.98**	0.97	0.94	1.00	0.15	0.57	0.49	0.50	0.50

Table 27: *TNR* results for **non-linear** data description

Dataset	Iris				Ionosphere			Sonar		
Target class	I-S	S-VC	S-V	Av.	I-B	I-G	Av.	S-R	S-M	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.96	0.94	0.73	0.88	0.61	0.57	0.59	0.34	0.41	0.37
GESSVDD-Sb-GR-max	**1.00**	0.84	0.95	0.93	0.37	0.80	0.58	0.33	0.32	0.32
GESSVDD-Sb-SR-max	0.75	0.85	0.89	0.83	0.36	0.66	0.51	0.27	0.41	0.34
GESSVDD-Sb-$\mathcal{S}$-min	0.93	0.93	0.85	0.90	0.61	0.71	0.66	0.42	0.56	0.49
GESSVDD-Sb-GR-min	**1.00**	0.88	0.89	0.92	0.32	0.80	0.56	0.33	0.58	0.45
GESSVDD-Sb-SR-min	0.75	0.87	0.91	0.84	0.48	0.66	0.57	0.24	0.53	0.38
GESSVDD-Sw-$\mathcal{S}$-max	0.84	0.96	0.84	0.88	0.68	0.77	0.72	0.44	0.21	0.32
GESSVDD-Sw-GR-max	**1.00**	0.97	0.91	0.96	0.30	0.84	0.57	0.35	0.29	0.32
GESSVDD-Sw-SR-max	**1.00**	0.86	0.85	0.90	0.60	0.80	0.70	0.25	0.66	0.46
GESSVDD-Sw-$\mathcal{S}$-min	0.70	0.90	0.87	0.82	0.59	0.77	0.68	0.47	0.41	0.44
GESSVDD-Sw-GR-min	**1.00**	0.88	0.85	0.91	0.39	0.81	0.60	0.36	0.24	0.30
GESSVDD-Sw-SR-min	**1.00**	0.86	0.84	0.90	**0.94**	0.82	**0.88**	0.25	0.28	0.27
GESSVDD-kNN-$\mathcal{S}$-max	**1.00**	0.89	0.89	0.93	0.59	0.78	0.68	0.38	0.39	0.39
GESSVDD-kNN-GR-max	**1.00**	0.92	0.93	0.95	0.60	0.88	0.74	0.42	0.36	0.39
GESSVDD-kNN-SR-max	**1.00**	0.84	0.90	0.91	0.59	0.76	0.68	0.38	0.49	0.43
GESSVDD-kNN-$\mathcal{S}$-min	**1.00**	0.88	0.91	0.93	0.26	0.78	0.52	0.59	0.39	0.49
GESSVDD-kNN-GR-min	0.93	0.97	0.94	0.95	0.38	0.89	0.64	0.40	0.48	0.44
GESSVDD-kNN-SR-min	**1.00**	0.83	0.93	0.92	0.50	0.76	0.63	0.48	0.57	0.53
GESSVDD-PCA-$\mathcal{S}$-max	**1.00**	0.97	0.93	0.96	0.28	0.81	0.55	0.25	0.26	0.25
GESSVDD-PCA-GR-max	0.96	0.91	0.91	0.93	0.41	0.85	0.63	0.38	0.24	0.31
GESSVDD-PCA-SR-max	**1.00**	0.91	0.95	0.95	0.51	0.81	0.66	0.32	0.59	0.45
GESSVDD-PCA-$\mathcal{S}$-min	**1.00**	0.91	0.75	0.88	0.36	0.83	0.60	0.56	0.54	0.55
GESSVDD-PCA-GR-min	0.97	0.66	0.89	0.84	0.38	0.89	0.64	0.42	0.37	0.40
GESSVDD-PCA-SR-min	**1.00**	0.91	0.85	0.92	0.36	0.86	0.61	0.59	0.48	0.53
GESSVDD-I-$\mathcal{S}$-max	**1.00**	0.96	0.87	0.94	0.12	0.79	0.45	0.36	0.28	0.32
GESSVDD-I-GR-max	0.99	0.91	0.87	0.92	0.44	0.83	0.64	0.47	0.37	0.42
GESSVDD-I-SR-max	**1.00**	0.96	0.87	0.94	0.10	0.79	0.45	0.37	0.57	0.47
GESSVDD-I-$\mathcal{S}$-min	**1.00**	0.96	0.87	0.94	0.21	0.79	0.50	0.45	0.57	0.51
GESSVDD-I-GR-min (ESSVDD)	0.97	0.93	0.91	0.94	0.43	0.86	0.64	0.46	0.38	0.42
GESSVDD-I-SR-min	**1.00**	0.96	0.85	0.94	0.17	0.79	0.48	0.46	0.50	0.48
GESSVDD-0-$\mathcal{S}$-max	**1.00**	0.91	0.93	0.95	0.46	0.58	0.52	0.28	0.70	0.49
GESSVDD-0-GR-max	**1.00**	0.93	0.89	0.94	0.55	0.73	0.64	0.35	0.40	0.38
GESSVDD-0-$\mathcal{S}$-min	0.97	0.89	0.94	0.93	0.32	0.59	0.46	0.45	0.46	0.46
GESSVDD-0-GR-min (SSVDD)	**1.00**	0.94	0.95	0.96	0.28	0.87	0.57	0.31	0.36	0.33
ESVDD	**1.00**	**0.99**	0.95	**0.98**	0.21	**0.94**	0.57	**0.72**	**0.93**	**0.83**
SVDD	**1.00**	0.95	0.91	0.95	0.71	0.79	0.75	0.38	0.46	0.42
OCSVM	0.80	0.41	**0.98**	0.73	0.65	0.69	0.67	0.47	0.53	0.50
GESVDD-PCA	**1.00**	0.95	0.95	0.97	0.51	0.88	0.69	0.46	0.50	0.48
GESVDD-Sw	**1.00**	0.96	0.96	0.97	0.59	0.89	0.74	0.44	0.53	0.49
GESVDD-kNN	**1.00**	0.96	0.95	0.97	0.75	0.90	0.83	0.48	0.56	0.52
GESVM-PCA	**1.00**	0.97	0.95	0.97	0.33	0.93	0.63	0.44	0.58	0.51
GESVM-Sw	**1.00**	0.93	0.95	0.96	0.50	0.92	0.71	0.46	0.68	0.57
GESVM-kNN	**1.00**	0.95	0.95	0.97	0.33	0.91	0.62	0.44	0.62	0.53

Table 28: *TNR* results for **non-linear** data description

Dataset	MNIST											Liver		
Target class	0	1	2	3	4	5	6	7	8	9	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.16	0.16	0.06	0.08	0.39	0.29	0.31	0.00	0.12	0.09	0.17	0.37	0.35	0.36
GESSVDD-Sb-GR-max	0.62	0.32	0.48	0.41	0.20	0.33	0.21	0.38	0.11	0.58	0.36	0.29	0.35	0.32
GESSVDD-Sb-SR-max	0.01	0.59	0.26	0.67	0.05	0.04	0.10	0.26	0.11	0.47	0.26	0.29	0.27	0.28
GESSVDD-Sb-$\mathcal{S}$-min	0.79	0.08	0.00	0.10	0.69	0.82	0.80	0.00	0.21	0.20	0.37	0.34	0.33	0.33
GESSVDD-Sb-GR-min	0.62	0.18	0.67	0.24	0.04	0.61	0.71	0.18	**1.00**	0.38	0.46	0.35	0.24	0.30
GESSVDD-Sb-SR-min	0.01	0.59	0.26	0.67	0.05	0.04	0.10	0.26	0.11	0.47	0.26	0.34	0.29	0.31
GESSVDD-Sw-$\mathcal{S}$-max	0.07	0.05	0.25	0.14	0.28	0.11	0.18	0.33	0.05	0.10	0.16	0.45	0.32	0.38
GESSVDD-Sw-GR-max	0.25	0.22	0.09	0.44	0.12	0.52	0.23	0.43	0.20	0.17	0.27	0.36	0.40	0.38
GESSVDD-Sw-SR-max	0.44	0.80	0.20	0.55	0.79	0.32	0.84	0.79	0.15	0.48	0.53	0.39	0.47	0.43
GESSVDD-Sw-$\mathcal{S}$-min	0.00	0.01	0.86	0.22	0.48	0.04	0.21	0.44	0.00	0.01	0.23	0.24	0.33	0.28
GESSVDD-Sw-GR-min	0.52	0.48	0.55	0.27	0.02	0.39	0.45	0.34	0.85	0.36	0.42	0.33	0.28	0.30
GESSVDD-Sw-SR-min	0.44	0.80	0.20	0.55	0.79	0.32	0.84	0.79	0.15	0.48	0.53	0.31	0.40	0.36
GESSVDD-kNN-$\mathcal{S}$-max	0.08	0.23	0.53	0.06	0.55	0.31	0.31	0.26	0.27	0.25	0.28	0.39	0.32	0.36
GESSVDD-kNN-GR-max	0.62	0.32	0.49	0.49	0.24	0.47	0.18	0.40	0.37	0.55	0.41	0.25	0.45	0.35
GESSVDD-kNN-SR-max	0.89	0.80	0.04	0.81	0.78	0.52	0.79	0.79	0.66	0.76	0.68	0.41	0.42	0.41
GESSVDD-kNN-$\mathcal{S}$-min	0.01	0.40	0.70	0.00	0.77	0.75	0.39	0.43	0.24	0.12	0.38	0.33	0.31	0.32
GESSVDD-kNN-GR-min	0.40	0.18	0.67	0.60	0.26	0.34	0.43	0.22	0.31	0.35	0.38	0.39	0.39	0.39
GESSVDD-kNN-SR-min	0.89	0.80	0.04	0.81	0.78	0.52	0.79	0.79	0.66	0.76	0.68	0.40	0.44	0.42
GESSVDD-PCA-$\mathcal{S}$-max	0.19	0.33	0.15	0.17	0.18	0.31	0.05	0.54	0.24	0.17	0.23	0.28	0.37	0.32
GESSVDD-PCA-GR-max	0.45	0.25	0.19	0.29	0.27	0.27	0.41	0.08	0.28	0.28	0.28	0.39	0.34	0.37
GESSVDD-PCA-SR-max	0.96	0.80	0.95	0.67	0.78	0.80	0.84	0.09	0.42	0.79	0.71	0.36	0.27	0.32
GESSVDD-PCA-$\mathcal{S}$-min	0.41	0.99	0.21	0.18	0.60	0.32	0.01	0.88	0.14	0.05	0.38	0.34	0.28	0.31
GESSVDD-PCA-GR-min	0.60	0.68	0.29	0.50	0.37	0.30	0.51	0.01	0.77	0.21	0.42	0.31	0.34	0.32
GESSVDD-PCA-SR-min	0.96	0.80	0.95	0.67	0.78	0.80	0.84	0.09	0.42	0.79	0.71	0.32	0.26	0.29
GESSVDD-I-$\mathcal{S}$-max	0.30	0.33	0.17	0.17	0.30	0.32	0.42	0.54	0.21	0.17	0.29	0.42	0.27	0.34
GESSVDD-I-GR-max	0.36	0.34	0.18	0.09	0.19	0.52	0.46	0.43	0.36	0.21	0.31	0.35	0.42	0.39
GESSVDD-I-SR-max	0.96	0.80	0.39	0.78	0.78	0.27	0.84	0.90	0.79	0.79	0.73	0.35	0.42	0.39
GESSVDD-I-$\mathcal{S}$-min	0.32	0.99	0.24	0.13	0.30	0.33	0.37	0.88	0.16	0.05	0.38	0.39	0.26	0.32
GESSVDD-I-GR-min (ESSVDD)	**0.99**	0.95	0.21	0.97	0.22	0.42	0.24	0.33	0.63	0.26	0.52	0.40	0.49	0.45
GESSVDD-I-SR-min	0.96	0.80	0.39	0.78	0.78	0.27	0.84	0.90	0.79	0.79	0.73	0.30	0.37	0.33
GESSVDD-0-$\mathcal{S}$-max	0.22	0.09	0.06	0.28	0.23	0.40	0.08	0.28	0.11	0.05	0.18	0.36	0.37	0.36
GESSVDD-0-GR-max	0.60	0.34	0.48	0.39	0.43	0.49	0.43	0.35	0.44	0.17	0.41	0.36	0.44	0.40
GESSVDD-0-$\mathcal{S}$-min	0.39	**1.00**	**0.97**	0.13	0.07	0.23	0.01	0.58	0.02	0.00	0.34	0.37	0.37	0.37
GESSVDD-0-GR-min (SSVDD)	0.89	0.29	0.67	0.24	0.27	0.32	0.24	0.41	0.30	0.06	0.37	0.37	0.39	0.38
ESVDD	0.95	0.96	**0.97**	**0.99**	**0.98**	**0.99**	**0.97**	**0.98**	0.99	**0.98**	**0.98**	0.43	0.54	0.49
SVDD	0.03	0.02	0.46	0.03	0.02	0.03	0.03	0.04	0.02	0.03	0.07	0.49	0.40	0.45
OCSVM	0.38	0.48	0.46	0.25	0.39	0.54	0.47	0.34	0.41	0.52	0.42	0.27	0.08	0.17
GESVDD-PCA	0.90	0.95	0.77	0.74	0.79	0.67	0.88	0.84	0.68	0.87	0.81	0.51	0.49	0.50
GESVDD-Sw	0.89	0.95	0.78	0.74	0.79	0.66	0.00	0.86	0.63	0.87	0.72	0.51	0.52	0.51
GESVDD-kNN	0.89	0.95	0.77	0.73	0.79	0.64	0.85	0.86	0.71	0.87	0.81	0.51	0.52	0.52
GESVM-PCA	0.90	0.90	0.81	0.71	0.89	0.58	0.83	0.85	0.75	0.90	0.81	0.50	0.55	0.52
GESVM-Sw	0.92	0.90	0.72	0.71	0.81	0.72	0.80	0.89	0.68	0.82	0.80	0.50	0.51	0.51
GESVM-kNN	0.90	0.99	0.67	0.83	0.84	0.86	0.95	0.86	0.80	0.83	0.85	**0.61**	**0.66**	**0.64**

Table 29: *TNR* results for **non-linear** data description in the proposed framework with added noise

Dataset	Heart Clean train set Clean test set			Heart Clean train set Corrupted test set			Heart Corrupted train set Clean test set			Heart Corrupted train set Corrupted test set		
Target class	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.29	0.29	0.29	0.42	0.29	0.36	0.13	0.05	0.09	0.35	0.40	0.37
GESSVDD-Sb-GR-max	0.34	0.34	0.34	0.35	0.23	0.29	0.33	0.31	**0.32**	0.27	0.45	0.36
GESSVDD-Sb-SR-max	0.32	0.32	0.32	0.22	0.24	0.23	0.30	0.22	0.26	0.43	0.44	0.44
GESSVDD-Sb-$\mathcal{S}$-min	0.41	0.41	0.41	0.44	0.19	0.32	0.19	0.19	0.19	0.35	0.20	0.28
GESSVDD-Sb-GR-min	0.45	0.45	0.45	0.32	0.18	0.25	0.21	0.32	0.26	0.33	0.33	0.33
GESSVDD-Sb-SR-min	0.46	0.46	0.46	0.25	0.13	0.19	**0.36**	0.16	0.26	0.43	0.44	0.43
GESSVDD-Sw-$\mathcal{S}$-max	0.43	0.43	0.43	0.00	0.12	0.06	0.13	0.00	0.06	0.36	0.34	0.35
GESSVDD-Sw-GR-max	0.16	0.16	0.16	0.36	0.39	0.38	0.07	0.11	0.09	0.42	0.39	0.41
GESSVDD-Sw-SR-max	0.43	0.43	0.43	0.11	0.06	0.08	0.00	0.06	0.03	0.43	0.41	0.42
GESSVDD-Sw-$\mathcal{S}$-min	0.34	0.34	0.34	0.00	0.16	0.08	0.00	0.15	0.08	0.44	0.43	0.43
GESSVDD-Sw-GR-min	0.40	0.40	0.40	0.18	0.29	0.24	0.04	0.05	0.05	0.45	0.44	0.45
GESSVDD-Sw-SR-min	0.44	0.44	0.44	0.03	0.19	0.11	0.00	0.00	0.00	0.41	0.33	0.37
GESSVDD-kNN-$\mathcal{S}$-max	0.44	0.44	0.44	0.00	0.00	0.00	0.14	0.04	0.09	0.45	0.45	0.45
GESSVDD-kNN-GR-max	0.43	0.43	0.43	0.28	0.19	0.24	0.12	0.00	0.06	0.44	0.38	0.41
GESSVDD-kNN-SR-max	0.42	0.42	0.42	0.04	0.03	0.03	0.00	0.00	0.00	0.47	0.40	0.44
GESSVDD-kNN-$\mathcal{S}$-min	0.49	0.49	0.49	0.03	0.00	0.02	0.03	0.00	0.02	0.46	0.43	0.44
GESSVDD-kNN-GR-min	0.47	0.47	0.47	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.43	0.46
GESSVDD-kNN-SR-min	0.47	0.47	0.47	0.04	0.07	0.06	0.06	0.00	0.03	0.45	0.38	0.41
GESSVDD-PCA-$\mathcal{S}$-max	0.37	0.37	0.37	0.00	0.03	0.01	0.00	0.00	0.00	0.49	0.36	0.42
GESSVDD-PCA-GR-max	0.47	0.47	0.47	0.30	0.34	0.32	0.13	0.16	0.14	0.45	0.36	0.40
GESSVDD-PCA-SR-max	0.46	0.46	0.46	0.19	0.05	0.12	0.06	0.09	0.08	0.34	0.41	0.37
GESSVDD-PCA-$\mathcal{S}$-min	0.24	0.24	0.24	0.06	0.11	0.08	0.00	0.00	0.00	0.43	0.39	0.41
GESSVDD-PCA-GR-min	0.30	0.30	0.30	0.28	0.22	0.25	0.00	0.04	0.02	0.46	0.47	0.47
GESSVDD-PCA-SR-min	0.44	0.44	0.44	0.10	0.07	0.09	0.03	0.00	0.02	0.37	0.41	0.39
GESSVDD-I-$\mathcal{S}$-max	0.38	0.38	0.38	0.06	0.03	0.04	0.04	0.00	0.02	0.45	0.38	0.42
GESSVDD-I-GR-max	0.42	0.42	0.42	0.37	0.35	0.36	0.03	0.05	0.04	0.48	0.31	0.39
GESSVDD-I-SR-max	0.49	0.49	0.49	0.13	0.03	0.08	0.00	0.00	0.00	0.43	0.44	0.43
GESSVDD-I-$\mathcal{S}$-min	0.44	0.44	0.44	0.00	0.25	0.12	0.08	0.00	0.04	0.36	0.39	0.38
GESSVDD-I-GR-min (ESSVDD)	0.38	0.38	0.38	0.37	0.15	0.26	0.00	0.11	0.06	0.38	0.46	0.42
GESSVDD-I-SR-min	0.46	0.46	0.46	0.15	0.14	0.14	0.00	0.08	0.04	0.42	0.45	0.44
GESSVDD-0-$\mathcal{S}$-max	0.35	0.35	0.35	0.24	0.21	0.22	0.26	0.32	0.29	0.39	0.36	0.38
GESSVDD-0-GR-max	0.45	0.45	0.45	0.06	0.22	0.14	0.06	**0.44**	0.25	0.46	0.36	0.41
GESSVDD-0-$\mathcal{S}$-min	0.32	0.32	0.32	0.14	0.16	0.15	**0.36**	0.09	0.23	0.34	0.40	0.37
GESSVDD-0-GR-min (SSVDD)	0.53	0.53	0.53	0.12	0.24	0.18	0.00	0.09	0.04	0.37	0.47	0.42
ESVDD	0.34	0.34	0.34	0.00	0.00	0.00	0.00	0.03	0.02	0.49	0.52	0.51
SVDD	0.53	0.53	0.53	0.07	0.15	0.11	0.22	0.07	0.15	0.47	0.45	0.46
OCSVM	0.20	0.20	0.20	0.00	0.00	0.00	0.00	0.07	0.03	0.50	0.49	0.50
GESVDD-PCA	0.68	0.68	0.68	0.00	0.00	0.00	0.00	0.00	0.00	0.51	0.53	0.52
GESVDD-Sw	0.68	0.68	0.68	0.00	0.00	0.00	0.00	0.00	0.00	0.53	0.52	0.52
GESVDD-kNN	0.70	0.70	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.52	0.50
GESVM-PCA	0.66	0.66	0.66	0.00	0.00	0.00	0.00	0.00	0.00	0.49	0.49	0.49
GESVM-Sw	0.67	0.67	0.67	0.00	0.00	0.00	0.00	0.05	0.03	0.50	0.48	0.49
GESVM-kNN	**0.76**	**0.76**	**0.76**	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	**0.56**	**0.56**	**0.56**

*S3.2.3. FPR results **non-linear** data description*

Table 30: *FPR* results for **non-linear** data description

Dataset	Seeds				Qualitative bankruptcy			Somerville happiness		
Target class	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.31	0.11	0.12	0.18	0.50	0.30	0.40	**0.81**	**0.84**	**0.83**
GESSVDD-Sb-GR-max	0.11	0.05	0.14	0.10	0.26	0.47	0.37	0.75	0.66	0.70
GESSVDD-Sb-SR-max	0.22	0.10	0.11	0.15	0.52	0.59	0.55	0.39	0.76	0.57
GESSVDD-Sb-$\mathcal{S}$-min	0.26	0.08	0.12	0.15	0.47	0.39	0.43	0.55	0.74	0.64
GESSVDD-Sb-GR-min	0.11	0.06	0.12	0.10	0.31	0.38	0.34	0.63	0.64	0.64
GESSVDD-Sb-SR-min	**0.35**	0.04	0.11	0.17	0.52	0.39	0.45	0.59	0.81	0.70
GESSVDD-Sw-$\mathcal{S}$-max	0.17	0.16	0.14	0.16	0.19	0.83	0.51	0.59	0.77	0.68
GESSVDD-Sw-GR-max	0.10	0.17	0.11	0.13	0.19	0.95	0.57	0.67	0.73	0.70
GESSVDD-Sw-SR-max	0.21	0.14	0.10	0.15	0.20	0.35	0.27	0.56	0.73	0.64
GESSVDD-Sw-$\mathcal{S}$-min	0.19	0.16	0.14	0.16	0.19	0.83	0.51	0.48	0.56	0.52
GESSVDD-Sw-GR-min	0.10	0.05	0.08	0.08	0.27	**0.99**	**0.63**	0.66	0.74	0.70
GESSVDD-Sw-SR-min	0.24	0.13	0.08	0.15	0.04	0.17	0.10	0.45	0.69	0.57
GESSVDD-kNN-$\mathcal{S}$-max	0.18	0.10	0.14	0.14	0.37	0.46	0.41	0.58	0.77	0.68
GESSVDD-kNN-GR-max	0.19	0.07	0.12	0.13	0.15	0.74	0.45	0.55	0.64	0.60
GESSVDD-kNN-SR-max	0.18	0.04	0.08	0.10	0.28	0.26	0.27	0.59	0.60	0.59
GESSVDD-kNN-$\mathcal{S}$-min	0.18	0.14	0.08	0.13	0.20	0.15	0.18	0.63	0.56	0.59
GESSVDD-kNN-GR-min	0.13	0.06	0.13	0.10	0.17	0.62	0.40	0.51	0.64	0.57
GESSVDD-kNN-SR-min	0.18	0.12	0.06	0.12	0.28	0.10	0.19	0.51	0.63	0.57
GESSVDD-PCA-$\mathcal{S}$-max	0.20	0.10	0.07	0.12	0.16	0.27	0.21	0.61	0.70	0.65
GESSVDD-PCA-GR-max	0.25	0.13	0.13	0.17	0.10	0.93	0.52	0.53	0.79	0.66
GESSVDD-PCA-SR-max	0.22	0.05	0.07	0.11	0.06	0.53	0.29	0.64	0.66	0.65
GESSVDD-PCA-$\mathcal{S}$-min	0.23	0.06	0.07	0.12	0.12	0.30	0.21	0.43	0.73	0.58
GESSVDD-PCA-GR-min	0.14	0.08	0.08	0.10	0.09	0.94	0.51	0.57	0.76	0.66
GESSVDD-PCA-SR-min	0.23	0.14	0.07	0.15	0.08	0.41	0.24	0.61	0.68	0.64
GESSVDD-I-$\mathcal{S}$-max	0.22	0.09	0.07	0.13	0.19	0.40	0.30	0.74	0.73	0.73
GESSVDD-I-GR-max	0.17	0.07	0.10	0.11	0.13	0.96	0.55	0.61	0.73	0.67
GESSVDD-I-SR-max	0.22	0.09	0.10	0.14	0.12	0.54	0.33	0.63	0.61	0.62
GESSVDD-I-$\mathcal{S}$-min	0.22	0.04	0.10	0.12	0.07	0.43	0.25	0.56	0.72	0.64
GESSVDD-I-GR-min (ESSVDD)	0.16	0.09	0.12	0.12	0.12	0.90	0.51	0.51	0.72	0.61
GESSVDD-I-SR-min	0.22	0.05	0.10	0.13	0.23	0.40	0.32	0.51	0.77	0.64
GESSVDD-0-$\mathcal{S}$-max	0.17	0.06	0.07	0.10	0.21	0.55	0.38	0.71	0.75	0.73
GESSVDD-0-GR-max	0.16	0.04	0.07	0.09	0.08	0.43	0.26	0.67	0.63	0.65
GESSVDD-0-$\mathcal{S}$-min	0.15	0.08	0.07	0.10	0.21	0.46	0.33	0.54	0.56	0.55
GESSVDD-0-GR-min (SSVDD)	0.13	0.04	0.08	0.08	0.08	0.47	0.28	0.59	0.50	0.54
ESVDD	0.07	0.02	0.02	0.04	0.00	0.00	0.00	0.00	0.57	0.29
SVDD	0.15	0.06	0.07	0.09	0.49	0.22	0.35	0.76	0.40	0.58
OCSVM	0.33	**0.23**	**0.46**	**0.34**	**0.70**	0.32	0.51	0.53	0.55	0.54
GESVDD-PCA	0.11	0.03	0.05	0.06	0.02	0.83	0.43	0.57	0.59	0.58
GESVDD-Sw	0.09	0.03	0.04	0.05	0.03	0.83	0.43	0.39	0.43	0.41
GESVDD-kNN	0.09	0.03	0.05	0.05	0.17	0.84	0.51	0.58	0.31	0.45
GESVM-PCA	0.13	0.03	0.03	0.07	0.00	0.81	0.41	0.53	0.51	0.52
GESVM-Sw	0.15	0.04	0.05	0.08	0.00	0.85	0.43	0.45	0.64	0.55
GESVM-kNN	0.12	0.02	0.03	0.06	0.00	0.85	0.43	0.51	0.50	0.50

Table 31: *FPR* results for **non-linear** data description

Dataset	Iris				Ionosphere			Sonar		
Target class	I-S	S-VC	S-V	Av.	I-B	I-G	Av.	S-R	S-M	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.04	0.06	**0.27**	0.12	0.39	**0.43**	0.41	0.66	0.59	0.63
GESSVDD-Sb-GR-max	0.00	0.16	0.05	0.07	0.63	0.20	0.42	0.67	0.68	0.68
GESSVDD-Sb-SR-max	0.25	0.15	0.11	0.17	0.64	0.34	0.49	0.73	0.59	0.66
GESSVDD-Sb-$\mathcal{S}$-min	0.07	0.07	0.15	0.10	0.39	0.29	0.34	0.58	0.44	0.51
GESSVDD-Sb-GR-min	0.00	0.12	0.11	0.08	0.68	0.20	0.44	0.67	0.42	0.55
GESSVDD-Sb-SR-min	0.25	0.13	0.09	0.16	0.52	0.34	0.43	**0.76**	0.47	0.62
GESSVDD-Sw-$\mathcal{S}$-max	0.16	0.04	0.16	0.12	0.32	0.23	0.28	0.56	**0.79**	0.68
GESSVDD-Sw-GR-max	0.00	0.03	0.09	0.04	0.70	0.16	0.43	0.65	0.71	0.68
GESSVDD-Sw-SR-max	0.00	0.14	0.15	0.10	0.40	0.20	0.30	0.75	0.34	0.54
GESSVDD-Sw-$\mathcal{S}$-min	**0.30**	0.10	0.13	0.18	0.41	0.23	0.32	0.53	0.59	0.56
GESSVDD-Sw-GR-min	0.00	0.12	0.15	0.09	0.61	0.19	0.40	0.64	0.76	0.70
GESSVDD-Sw-SR-min	0.00	0.14	0.16	0.10	0.06	0.18	0.12	0.75	0.72	0.73
GESSVDD-kNN-$\mathcal{S}$-max	0.00	0.11	0.11	0.07	0.41	0.22	0.32	0.62	0.61	0.61
GESSVDD-kNN-GR-max	0.00	0.08	0.07	0.05	0.40	0.12	0.26	0.58	0.64	0.61
GESSVDD-kNN-SR-max	0.00	0.16	0.10	0.09	0.41	0.24	0.32	0.62	0.51	0.57
GESSVDD-kNN-$\mathcal{S}$-min	0.00	0.12	0.09	0.07	0.74	0.22	0.48	0.41	0.61	0.51
GESSVDD-kNN-GR-min	0.07	0.03	0.06	0.05	0.62	0.11	0.36	0.60	0.52	0.56
GESSVDD-kNN-SR-min	0.00	0.17	0.07	0.08	0.50	0.24	0.37	0.52	0.43	0.47
GESSVDD-PCA-$\mathcal{S}$-max	0.00	0.03	0.07	0.04	0.72	0.19	0.45	0.75	0.74	**0.75**
GESSVDD-PCA-GR-max	0.04	0.09	0.09	0.07	0.59	0.15	0.37	0.62	0.76	0.69
GESSVDD-PCA-SR-max	0.00	0.09	0.05	0.05	0.49	0.19	0.34	0.68	0.41	0.55
GESSVDD-PCA-$\mathcal{S}$-min	0.00	0.09	0.25	0.12	0.64	0.17	0.40	0.44	0.46	0.45
GESSVDD-PCA-GR-min	0.03	0.34	0.11	0.16	0.62	0.11	0.36	0.58	0.63	0.60
GESSVDD-PCA-SR-min	0.00	0.09	0.15	0.08	0.64	0.14	0.39	0.41	0.52	0.47
GESSVDD-I-$\mathcal{S}$-max	0.00	0.04	0.13	0.06	0.88	0.21	0.55	0.64	0.72	0.68
GESSVDD-I-GR-max	0.01	0.09	0.13	0.08	0.56	0.17	0.36	0.53	0.63	0.58
GESSVDD-I-SR-max	0.00	0.04	0.13	0.06	**0.90**	0.21	**0.55**	0.63	0.43	0.53
GESSVDD-I-$\mathcal{S}$-min	0.00	0.04	0.13	0.06	0.79	0.21	0.50	0.55	0.43	0.49
GESSVDD-I-GR-min (ESSVDD)	0.03	0.07	0.09	0.06	0.57	0.14	0.36	0.54	0.62	0.58
GESSVDD-I-SR-min	0.00	0.04	0.15	0.06	0.83	0.21	0.52	0.54	0.50	0.52
GESSVDD-0-$\mathcal{S}$-max	0.00	0.09	0.07	0.05	0.54	0.42	0.48	0.72	0.30	0.51
GESSVDD-0-GR-max	0.00	0.07	0.11	0.06	0.45	0.27	0.36	0.65	0.60	0.62
GESSVDD-0-$\mathcal{S}$-min	0.03	0.11	0.06	0.07	0.68	0.41	0.54	0.55	0.54	0.54
GESSVDD-0-GR-min (SSVDD)	0.00	0.06	0.05	0.04	0.72	0.13	0.43	0.69	0.64	0.67
ESVDD	0.00	0.01	0.05	0.02	0.79	0.06	0.43	0.28	0.07	0.17
SVDD	0.00	0.05	0.09	0.05	0.29	0.21	0.25	0.62	0.54	0.58
OCSVM	0.20	**0.59**	0.02	**0.27**	0.35	0.31	0.33	0.53	0.47	0.50
GESVDD-PCA	0.00	0.05	0.05	0.03	0.49	0.12	0.31	0.54	0.50	0.52
GESVDD-Sw	0.00	0.04	0.04	0.03	0.41	0.11	0.26	0.56	0.47	0.51
GESVDD-kNN	0.00	0.04	0.05	0.03	0.25	0.10	0.17	0.52	0.44	0.48
GESVM-PCA	0.00	0.03	0.05	0.03	0.67	0.07	0.37	0.56	0.42	0.49
GESVM-Sw	0.00	0.07	0.05	0.04	0.50	0.08	0.29	0.54	0.32	0.43
GESVM-kNN	0.00	0.05	0.05	0.03	0.67	0.09	0.38	0.56	0.38	0.47

Table 32: *FPR* results for **non-linear** data description

Dataset	MNIST											Liver		
Target class	0	1	2	3	4	5	6	7	8	9	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.84	0.84	0.94	0.92	0.61	0.71	0.69	**1.00**	0.88	0.91	0.83	0.63	0.65	0.64
GESSVDD-Sb-GR-max	0.38	0.68	0.52	0.59	0.80	0.67	0.79	0.62	0.89	0.42	0.64	0.71	0.65	0.68
GESSVDD-Sb-SR-max	0.99	0.41	0.74	0.33	0.95	0.96	0.90	0.74	0.89	0.53	0.74	0.71	0.73	0.72
GESSVDD-Sb-$\mathcal{S}$-min	0.21	0.92	**1.00**	0.90	0.31	0.18	0.20	**1.00**	0.79	0.80	0.63	0.66	0.67	0.67
GESSVDD-Sb-GR-min	0.38	0.82	0.33	0.76	0.96	0.39	0.29	0.82	0.00	0.62	0.54	0.65	0.76	0.70
GESSVDD-Sb-SR-min	0.99	0.41	0.74	0.33	0.95	0.96	0.90	0.74	0.89	0.53	0.74	0.66	0.71	0.69
GESSVDD-Sw-$\mathcal{S}$-max	0.93	0.95	0.75	0.86	0.72	0.89	0.82	0.67	0.95	0.90	0.84	0.55	0.68	0.62
GESSVDD-Sw-GR-max	0.75	0.78	0.91	0.56	0.88	0.48	0.77	0.57	0.80	0.83	0.73	0.64	0.60	0.62
GESSVDD-Sw-SR-max	0.56	0.20	0.80	0.45	0.21	0.68	0.16	0.21	0.85	0.52	0.47	0.61	0.53	0.57
GESSVDD-Sw-$\mathcal{S}$-min	**1.00**	**0.99**	0.14	0.78	0.52	0.96	0.79	0.56	**1.00**	0.99	0.77	**0.76**	0.67	0.72
GESSVDD-Sw-GR-min	0.48	0.52	0.45	0.73	**0.98**	0.61	0.55	0.66	0.15	0.64	0.58	0.67	0.72	0.70
GESSVDD-Sw-SR-min	0.56	0.20	0.80	0.45	0.21	0.68	0.16	0.21	0.85	0.52	0.47	0.69	0.60	0.64
GESSVDD-kNN-$\mathcal{S}$-max	0.92	0.77	0.47	0.94	0.45	0.69	0.69	0.74	0.73	0.75	0.72	0.61	0.68	0.64
GESSVDD-kNN-GR-max	0.38	0.68	0.51	0.51	0.76	0.53	0.82	0.60	0.63	0.45	0.59	0.75	0.55	0.65
GESSVDD-kNN-SR-max	0.11	0.20	0.96	0.19	0.22	0.48	0.21	0.21	0.34	0.24	0.32	0.59	0.58	0.59
GESSVDD-kNN-$\mathcal{S}$-min	0.99	0.60	0.30	**1.00**	0.23	0.25	0.61	0.57	0.76	0.88	0.62	0.67	0.69	0.68
GESSVDD-kNN-GR-min	0.60	0.82	0.33	0.40	0.74	0.66	0.57	0.78	0.69	0.65	0.62	0.61	0.61	0.61
GESSVDD-kNN-SR-min	0.11	0.20	0.96	0.19	0.22	0.48	0.21	0.21	0.34	0.24	0.32	0.60	0.56	0.58
GESSVDD-PCA-$\mathcal{S}$-max	0.81	0.67	0.85	0.83	0.82	0.69	0.95	0.46	0.76	0.83	0.77	0.72	0.63	0.68
GESSVDD-PCA-GR-max	0.55	0.75	0.81	0.71	0.73	0.73	0.59	0.92	0.72	0.72	0.72	0.61	0.66	0.63
GESSVDD-PCA-SR-max	0.04	0.20	0.05	0.33	0.22	0.20	0.16	0.91	0.58	0.21	0.29	0.64	0.73	0.68
GESSVDD-PCA-$\mathcal{S}$-min	0.59	0.01	0.79	0.82	0.40	0.68	0.99	0.12	0.86	0.95	0.62	0.66	0.72	0.69
GESSVDD-PCA-GR-min	0.40	0.32	0.71	0.50	0.63	0.70	0.49	0.99	0.23	0.79	0.58	0.69	0.66	0.68
GESSVDD-PCA-SR-min	0.04	0.20	0.05	0.33	0.22	0.20	0.16	0.91	0.58	0.21	0.29	0.68	0.74	0.71
GESSVDD-I-$\mathcal{S}$-max	0.70	0.67	0.83	0.83	0.70	0.68	0.58	0.46	0.79	0.83	0.71	0.58	0.73	0.66
GESSVDD-I-GR-max	0.64	0.66	0.82	0.91	0.81	0.48	0.54	0.57	0.64	0.79	0.69	0.65	0.58	0.61
GESSVDD-I-SR-max	0.04	0.20	0.61	0.22	0.22	0.73	0.16	0.10	0.21	0.21	0.27	0.65	0.58	0.61
GESSVDD-I-$\mathcal{S}$-min	0.68	0.01	0.76	0.87	0.70	0.67	0.63	0.12	0.84	0.95	0.62	0.61	0.74	0.68
GESSVDD-I-GR-min (ESSVDD)	0.01	0.05	0.79	0.03	0.78	0.58	0.76	0.67	0.37	0.74	0.48	0.60	0.51	0.55
GESSVDD-I-SR-min	0.04	0.20	0.61	0.22	0.22	0.73	0.16	0.10	0.21	0.21	0.27	0.70	0.63	0.67
GESSVDD-0-$\mathcal{S}$-max	0.78	0.91	0.94	0.72	0.77	0.60	0.92	0.72	0.89	0.95	0.82	0.64	0.63	0.64
GESSVDD-0-GR-max	0.40	0.66	0.52	0.61	0.57	0.51	0.57	0.65	0.56	0.83	0.59	0.64	0.56	0.60
GESSVDD-0-$\mathcal{S}$-min	0.61	0.00	0.03	0.87	0.93	0.77	0.99	0.42	0.98	**1.00**	0.66	0.63	0.63	0.63
GESSVDD-0-GR-min (SSVDD)	0.11	0.71	0.33	0.76	0.73	0.68	0.76	0.59	0.70	0.94	0.63	0.63	0.61	0.62
ESVDD	0.05	0.04	0.03	0.01	0.02	0.01	0.03	0.02	0.01	0.02	0.02	0.57	0.46	0.51
SVDD	0.97	0.98	0.54	0.97	**0.98**	**0.97**	0.97	0.96	0.98	0.97	**0.93**	0.51	0.60	0.55
OCSVM	0.62	0.52	0.54	0.75	0.61	0.46	0.53	0.66	0.59	0.48	0.58	0.73	**0.92**	**0.83**
GESVDD-PCA	0.10	0.05	0.23	0.26	0.21	0.33	0.12	0.16	0.32	0.13	0.19	0.49	0.51	0.50
GESVDD-Sw	0.11	0.05	0.22	0.26	0.21	0.34	**1.00**	0.14	0.37	0.13	0.28	0.49	0.48	0.49
GESVDD-kNN	0.11	0.05	0.23	0.27	0.21	0.36	0.15	0.14	0.29	0.13	0.19	0.49	0.48	0.48
GESVM-PCA	0.10	0.10	0.19	0.29	0.11	0.42	0.17	0.15	0.25	0.10	0.19	0.50	0.45	0.48
GESVM-Sw	0.08	0.10	0.28	0.29	0.19	0.28	0.20	0.11	0.32	0.18	0.20	0.50	0.49	0.49
GESVM-kNN	0.10	0.01	0.33	0.17	0.16	0.14	0.05	0.14	0.20	0.17	0.15	0.39	0.34	0.36

Table 33: *FPR* results for **non-linear** data description in the proposed framework with added noise

Dataset	Heart Clean train set Clean test set			Heart Clean train set Corrupted test set			Heart Corrupted train set Clean test set			Heart Corrupted train set Corrupted test set		
Target class	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.71	0.71	0.71	0.58	0.71	0.64	0.87	0.95	0.91	0.65	0.60	0.63
GESSVDD-Sb-GR-max	0.66	0.66	0.66	0.65	0.77	0.71	0.67	0.69	0.68	**0.73**	0.55	0.64
GESSVDD-Sb-SR-max	0.68	0.68	0.68	0.78	0.76	0.77	0.70	0.78	0.74	0.57	0.56	0.56
GESSVDD-Sb-$\mathcal{S}$-min	0.59	0.59	0.59	0.56	0.81	0.68	0.81	0.81	0.81	0.65	**0.80**	**0.72**
GESSVDD-Sb-GR-min	0.55	0.55	0.55	0.68	0.82	0.75	0.79	0.68	0.74	0.67	0.67	0.67
GESSVDD-Sb-SR-min	0.54	0.54	0.54	0.75	0.87	0.81	0.64	0.84	0.74	0.57	0.56	0.57
GESSVDD-Sw-$\mathcal{S}$-max	0.57	0.57	0.57	**1.00**	0.88	0.94	0.87	**1.00**	0.94	0.64	0.66	0.65
GESSVDD-Sw-GR-max	**0.84**	**0.84**	**0.84**	0.64	0.61	0.62	0.93	0.89	0.91	0.58	0.61	0.59
GESSVDD-Sw-SR-max	0.57	0.57	0.57	0.89	0.94	0.92	**1.00**	0.94	0.97	0.57	0.59	0.58
GESSVDD-Sw-$\mathcal{S}$-min	0.66	0.66	0.66	**1.00**	0.84	0.92	**1.00**	0.85	0.92	0.56	0.57	0.57
GESSVDD-Sw-GR-min	0.60	0.60	0.60	0.82	0.71	0.76	0.96	0.95	0.95	0.55	0.56	0.55
GESSVDD-Sw-SR-min	0.56	0.56	0.56	0.97	0.81	0.89	**1.00**	**1.00**	**1.00**	0.59	0.67	0.63
GESSVDD-kNN-$\mathcal{S}$-max	0.56	0.56	0.56	**1.00**	**1.00**	**1.00**	0.86	0.96	0.91	0.55	0.55	0.55
GESSVDD-kNN-GR-max	0.57	0.57	0.57	0.72	0.81	0.76	0.88	**1.00**	0.94	0.56	0.62	0.59
GESSVDD-kNN-SR-max	0.58	0.58	0.58	0.96	0.97	0.97	**1.00**	**1.00**	**1.00**	0.53	0.60	0.56
GESSVDD-kNN-$\mathcal{S}$-min	0.51	0.51	0.51	0.97	**1.00**	0.98	0.97	**1.00**	0.98	0.54	0.57	0.56
GESSVDD-kNN-GR-min	0.53	0.53	0.53	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.52	0.57	0.54
GESSVDD-kNN-SR-min	0.53	0.53	0.53	0.96	0.93	0.94	0.94	**1.00**	0.97	0.55	0.62	0.59
GESSVDD-PCA-$\mathcal{S}$-max	0.63	0.63	0.63	**1.00**	0.97	0.99	**1.00**	**1.00**	**1.00**	0.51	0.64	0.58
GESSVDD-PCA-GR-max	0.53	0.53	0.53	0.70	0.66	0.68	0.87	0.84	0.86	0.55	0.64	0.60
GESSVDD-PCA-SR-max	0.54	0.54	0.54	0.81	0.95	0.88	0.94	0.91	0.92	0.66	0.59	0.63
GESSVDD-PCA-$\mathcal{S}$-min	0.76	0.76	0.76	0.94	0.89	0.92	**1.00**	**1.00**	**1.00**	0.57	0.61	0.59
GESSVDD-PCA-GR-min	0.70	0.70	0.70	0.72	0.78	0.75	**1.00**	0.96	0.98	0.54	0.53	0.53
GESSVDD-PCA-SR-min	0.56	0.56	0.56	0.90	0.93	0.91	0.97	**1.00**	0.98	0.63	0.59	0.61
GESSVDD-I-$\mathcal{S}$-max	0.62	0.62	0.62	0.94	0.97	0.96	0.96	**1.00**	0.98	0.55	0.62	0.58
GESSVDD-I-GR-max	0.58	0.58	0.58	0.63	0.65	0.64	0.97	0.95	0.96	0.52	0.69	0.61
GESSVDD-I-SR-max	0.51	0.51	0.51	0.87	0.97	0.92	**1.00**	**1.00**	**1.00**	0.57	0.56	0.57
GESSVDD-I-$\mathcal{S}$-min	0.56	0.56	0.56	**1.00**	0.75	0.88	0.92	**1.00**	0.96	0.64	0.61	0.62
GESSVDD-I-GR-min (ESSVDD)	0.62	0.62	0.62	0.63	0.85	0.74	**1.00**	0.89	0.94	0.62	0.54	0.58
GESSVDD-I-SR-min	0.54	0.54	0.54	0.85	0.86	0.86	**1.00**	0.92	0.96	0.58	0.55	0.56
GESSVDD-0-$\mathcal{S}$-max	0.65	0.65	0.65	0.76	0.79	0.78	0.74	0.68	0.71	0.61	0.64	0.62
GESSVDD-0-GR-max	0.55	0.55	0.55	0.94	0.78	0.86	0.94	0.56	0.75	0.54	0.64	0.59
GESSVDD-0-$\mathcal{S}$-min	0.68	0.68	0.68	0.86	0.84	0.85	0.64	0.91	0.77	0.66	0.60	0.63
GESSVDD-0-GR-min (SSVDD)	0.47	0.47	0.47	0.88	0.76	0.82	**1.00**	0.91	0.96	0.63	0.53	0.58
ESVDD	0.66	0.66	0.66	**1.00**	**1.00**	**1.00**	**1.00**	0.97	0.98	0.51	0.48	0.49
SVDD	0.47	0.47	0.47	0.93	0.85	0.89	0.78	0.93	0.85	0.53	0.55	0.54
OCSVM	0.80	0.80	0.80	**1.00**	**1.00**	**1.00**	**1.00**	0.93	0.97	0.50	0.51	0.50
GESVDD-PCA	0.32	0.32	0.32	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.49	0.47	0.48
GESVDD-Sw	0.32	0.32	0.32	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.47	0.48	0.48
GESVDD-kNN	0.30	0.30	0.30	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.52	0.48	0.50
GESVM-PCA	0.34	0.34	0.34	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.51	0.51	0.51
GESVM-Sw	0.33	0.33	0.33	**1.00**	**1.00**	**1.00**	**1.00**	0.95	0.97	0.50	0.52	0.51
GESVM-kNN	0.24	0.24	0.24	0.00	0.00	0.00	**1.00**	**1.00**	**1.00**	0.44	0.44	0.44

*S3.2.4. FNR results **non-linear** data description*

Table 34: *FNR* results for **non-linear** data description

Dataset	Seeds				Qualitative bankruptcy			Somerville happiness		
Target class	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.14	0.06	0.11	0.10	0.24	0.54	0.39	0.22	0.23	0.22
GESSVDD-Sb-GR-max	0.22	0.36	0.08	0.22	0.07	0.41	0.24	0.26	0.34	0.30
GESSVDD-Sb-SR-max	0.22	0.26	0.03	0.17	0.30	0.55	0.42	0.53	0.51	0.52
GESSVDD-Sb-$\mathcal{S}$-min	0.14	0.10	0.19	0.15	0.24	0.49	0.37	0.38	0.28	0.33
GESSVDD-Sb-GR-min	0.22	0.42	0.06	0.23	0.06	0.46	0.26	0.26	0.36	0.31
GESSVDD-Sb-SR-min	0.19	0.22	0.05	0.15	0.30	0.60	0.45	0.40	0.45	0.43
GESSVDD-Sw-$\mathcal{S}$-max	0.11	0.05	0.01	0.06	0.12	0.11	0.11	0.29	0.29	0.29
GESSVDD-Sw-GR-max	0.12	0.10	0.06	0.10	0.20	0.11	0.15	0.29	0.29	0.29
GESSVDD-Sw-SR-max	0.14	0.07	0.12	0.11	0.09	0.42	0.25	0.32	0.39	0.36
GESSVDD-Sw-$\mathcal{S}$-min	0.22	0.05	0.01	0.09	0.08	0.14	0.11	0.37	0.43	0.40
GESSVDD-Sw-GR-min	0.11	0.07	0.09	0.09	0.17	0.11	0.14	0.26	0.28	0.27
GESSVDD-Sw-SR-min	0.12	0.11	0.10	0.11	0.09	0.38	0.24	0.43	0.37	0.40
GESSVDD-kNN-$\mathcal{S}$-max	0.07	0.10	0.06	0.07	0.10	0.22	0.16	0.36	0.42	0.39
GESSVDD-kNN-GR-max	0.16	0.10	0.09	0.12	0.09	0.23	0.16	0.29	0.42	0.35
GESSVDD-kNN-SR-max	0.09	0.11	0.01	0.07	0.08	0.30	0.19	0.40	0.43	0.42
GESSVDD-kNN-$\mathcal{S}$-min	0.07	0.10	0.05	0.07	0.17	0.28	0.22	0.33	0.43	0.38
GESSVDD-kNN-GR-min	0.19	0.07	0.04	0.10	0.13	0.38	0.25	0.25	0.48	0.37
GESSVDD-kNN-SR-min	0.08	0.09	0.07	0.08	0.13	0.37	0.25	0.46	0.37	0.41
GESSVDD-PCA-$\mathcal{S}$-max	0.11	0.08	0.05	0.08	0.44	0.33	0.39	0.29	0.27	0.28
GESSVDD-PCA-GR-max	0.15	0.07	0.06	0.09	0.10	0.14	0.12	0.34	0.32	0.33
GESSVDD-PCA-SR-max	0.10	0.27	0.05	0.14	0.13	0.40	0.27	0.31	0.35	0.33
GESSVDD-PCA-$\mathcal{S}$-min	0.07	0.05	0.05	0.05	0.17	0.36	0.27	0.45	0.40	0.43
GESSVDD-PCA-GR-min	0.13	0.19	0.05	0.12	0.06	0.16	0.11	0.32	0.34	0.33
GESSVDD-PCA-SR-min	0.07	0.05	0.06	0.06	0.07	0.28	0.17	0.30	0.33	0.32
GESSVDD-I-$\mathcal{S}$-max	0.06	0.06	0.22	0.11	0.13	0.37	0.25	0.28	0.38	0.33
GESSVDD-I-GR-max	0.17	0.10	0.08	0.12	0.06	0.15	0.10	0.29	0.29	0.29
GESSVDD-I-SR-max	0.06	0.06	0.02	0.04	0.15	0.17	0.16	0.31	0.44	0.38
GESSVDD-I-$\mathcal{S}$-min	0.06	0.07	0.04	0.05	0.21	0.38	0.30	0.41	0.37	0.39
GESSVDD-I-GR-min (ESSVDD)	0.16	0.10	0.07	0.11	0.03	0.19	0.11	0.29	0.42	0.35
GESSVDD-I-SR-min	0.06	0.04	0.03	0.04	0.07	0.35	0.21	0.34	0.32	0.33
GESSVDD-0-$\mathcal{S}$-max	0.12	0.10	0.07	0.10	0.32	0.48	0.40	0.27	0.16	0.21
GESSVDD-0-GR-max	0.14	0.10	0.06	0.10	0.05	0.57	0.31	0.36	0.27	0.32
GESSVDD-0-$\mathcal{S}$-min	0.13	0.10	0.10	0.11	0.34	0.46	0.40	0.43	0.46	0.44
GESSVDD-0-GR-min (SSVDD)	0.13	0.08	0.05	0.09	0.04	0.43	0.24	0.39	0.55	0.47
ESVDD	0.29	0.21	0.22	0.24	**1.00**	**1.00**	**1.00**	**1.00**	0.49	**0.75**
SVDD	0.14	0.11	0.04	0.10	0.32	0.73	0.53	0.24	0.66	0.45
OCSVM	**0.59**	**0.48**	**0.52**	**0.53**	0.28	0.50	0.39	0.51	0.45	0.48
GESVDD-PCA	0.18	0.11	0.09	0.13	0.09	0.36	0.22	0.37	0.41	0.39
GESVDD-Sw	0.25	0.11	0.10	0.15	0.09	0.38	0.23	0.56	0.55	0.55
GESVDD-kNN	0.22	0.13	0.08	0.14	0.13	0.35	0.24	0.34	**0.67**	0.51
GESVM-PCA	0.16	0.16	0.10	0.14	0.09	0.63	0.36	0.39	0.52	0.45
GESVM-Sw	0.15	0.15	0.13	0.15	0.13	0.58	0.35	0.43	0.46	0.45
GESVM-kNN	0.18	0.17	0.15	0.17	0.14	0.69	0.42	0.37	0.47	0.42

Table 35: *FNR* results for **non-linear** data description

Dataset	Iris				Ionosphere			Sonar		
Target class	I-S	S-VC	S-V	Av.	I-B	I-G	Av.	S-R	S-M	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.08	0.17	0.13	0.13	0.42	0.27	0.34	0.37	0.36	0.36
GESSVDD-Sb-GR-max	0.09	0.17	0.15	0.14	0.20	0.04	0.12	0.14	0.31	0.22
GESSVDD-Sb-SR-max	0.09	0.15	0.19	0.14	0.19	**0.73**	0.46	0.24	0.58	0.41
GESSVDD-Sb-$\mathcal{S}$-min	0.08	0.13	0.11	0.11	0.43	0.47	0.45	0.41	0.52	0.46
GESSVDD-Sb-GR-min	0.09	0.16	0.09	0.12	0.24	0.04	0.14	0.14	0.50	0.32
GESSVDD-Sb-SR-min	0.05	0.21	0.31	0.19	0.36	**0.73**	0.54	0.22	**0.76**	0.49
GESSVDD-Sw-$\mathcal{S}$-max	0.11	0.13	0.15	0.13	0.46	0.05	0.25	0.32	0.23	0.28
GESSVDD-Sw-GR-max	0.12	0.17	0.12	0.14	0.02	0.06	0.04	0.18	0.19	0.19
GESSVDD-Sw-SR-max	0.05	0.13	0.15	0.11	0.56	0.05	0.31	0.19	0.55	0.37
GESSVDD-Sw-$\mathcal{S}$-min	0.16	0.12	0.16	0.15	0.45	0.05	0.25	0.41	0.33	0.37
GESSVDD-Sw-GR-min	0.09	0.13	0.17	0.13	0.03	0.07	0.05	0.24	0.15	0.20
GESSVDD-Sw-SR-min	0.07	0.13	0.12	0.11	0.81	0.08	0.45	0.19	0.27	0.23
GESSVDD-kNN-$\mathcal{S}$-max	0.12	0.13	0.21	0.16	0.21	0.03	0.12	0.29	0.28	0.29
GESSVDD-kNN-GR-max	0.11	0.08	0.16	0.12	0.19	0.08	0.14	0.31	0.27	0.29
GESSVDD-kNN-SR-max	0.11	0.08	0.23	0.14	0.21	0.04	0.12	0.27	0.41	0.34
GESSVDD-kNN-$\mathcal{S}$-min	0.11	0.12	0.23	0.15	0.32	0.03	0.18	0.39	0.32	0.36
GESSVDD-kNN-GR-min	0.09	0.20	0.12	0.14	0.03	0.06	0.04	0.25	0.25	0.25
GESSVDD-kNN-SR-min	0.11	0.08	0.27	0.15	0.34	0.04	0.19	0.34	0.40	0.37
GESSVDD-PCA-$\mathcal{S}$-max	0.12	0.07	0.17	0.12	0.38	0.07	0.23	0.18	0.24	0.21
GESSVDD-PCA-GR-max	0.12	0.16	0.20	0.16	0.09	0.05	0.07	0.21	0.18	0.19
GESSVDD-PCA-SR-max	0.12	0.05	0.19	0.12	0.64	0.11	0.37	0.26	0.52	0.39
GESSVDD-PCA-$\mathcal{S}$-min	0.12	0.05	0.12	0.10	0.56	0.08	0.32	0.43	0.52	0.47
GESSVDD-PCA-GR-min	0.12	0.13	0.17	0.14	0.09	0.07	0.08	0.28	0.21	0.24
GESSVDD-PCA-SR-min	0.12	0.05	0.19	0.12	0.48	0.13	0.30	0.45	0.36	0.40
GESSVDD-I-$\mathcal{S}$-max	0.04	0.11	0.17	0.11	0.30	0.02	0.16	0.31	0.30	0.30
GESSVDD-I-GR-max	0.08	0.17	0.12	0.12	0.07	0.06	0.07	0.30	0.21	0.25
GESSVDD-I-SR-max	0.04	0.09	0.17	0.10	0.32	0.02	0.17	0.30	0.53	0.42
GESSVDD-I-$\mathcal{S}$-min	0.11	0.09	0.17	0.12	0.34	0.02	0.18	0.41	0.51	0.46
GESSVDD-I-GR-min (ESSVDD)	0.08	0.15	0.13	0.12	0.03	0.07	0.05	0.34	0.18	0.26
GESSVDD-I-SR-min	0.09	0.09	0.15	0.11	0.37	0.02	0.19	0.39	0.41	0.40
GESSVDD-0-$\mathcal{S}$-max	0.12	0.09	0.13	0.12	0.14	0.21	0.17	0.13	0.49	0.31
GESSVDD-0-GR-max	0.08	0.11	0.12	0.10	0.32	0.07	0.19	0.29	0.24	0.27
GESSVDD-0-$\mathcal{S}$-min	0.08	0.12	0.13	0.11	0.22	0.09	0.15	0.33	0.37	0.35
GESSVDD-0-GR-min (SSVDD)	0.11	0.09	0.15	0.12	0.04	0.08	0.06	0.18	0.28	0.23
ESVDD	0.43	0.29	0.27	0.33	0.31	0.17	0.24	**0.58**	0.71	**0.64**
SVDD	0.16	0.11	0.15	0.14	0.76	0.07	0.42	0.24	0.22	0.23
OCSVM	**0.47**	**0.57**	0.65	**0.56**	0.56	0.59	**0.58**	0.51	0.40	0.46
GESVDD-PCA	0.29	0.12	0.17	0.20	0.69	0.11	0.40	0.32	0.27	0.30
GESVDD-Sw	0.20	0.20	0.16	0.19	0.75	0.10	0.43	0.32	0.35	0.33
GESVDD-kNN	0.29	0.13	0.16	0.20	**0.82**	0.11	0.46	0.39	0.36	0.37
GESVM-PCA	0.19	0.16	0.13	0.16	0.50	0.12	0.31	0.36	0.36	0.36
GESVM-Sw	0.21	0.07	0.17	0.15	0.53	0.12	0.33	0.35	0.48	0.42
GESVM-kNN	0.20	0.16	0.16	0.17	0.47	0.14	0.31	0.34	0.44	0.39

Table 36: *FNR* results for **non-linear** data description

Dataset	MNIST											Liver		
Target class	0	1	2	3	4	5	6	7	8	9	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.84	0.84	0.94	0.92	0.61	0.71	0.69	**1.00**	0.88	0.91	0.83	0.63	0.65	0.64
GESSVDD-Sb-GR-max	0.38	0.68	0.52	0.59	0.80	0.67	0.79	0.62	0.89	0.42	0.64	0.71	0.65	0.68
GESSVDD-Sb-SR-max	0.99	0.41	0.74	0.33	0.95	0.96	0.90	0.74	0.89	0.53	0.74	0.71	0.73	0.72
GESSVDD-Sb-$\mathcal{S}$-min	0.21	0.92	**1.00**	0.90	0.31	0.18	0.20	**1.00**	0.79	0.80	0.63	0.66	0.67	0.67
GESSVDD-Sb-GR-min	0.38	0.82	0.33	0.76	0.96	0.39	0.29	0.82	0.00	0.62	0.54	0.65	0.76	0.70
GESSVDD-Sb-SR-min	0.99	0.41	0.74	0.33	0.95	0.96	0.90	0.74	0.89	0.53	0.74	0.66	0.71	0.69
GESSVDD-Sw-$\mathcal{S}$-max	0.93	0.95	0.75	0.86	0.72	0.89	0.82	0.67	0.95	0.90	0.84	0.55	0.68	0.62
GESSVDD-Sw-GR-max	0.75	0.78	0.91	0.56	0.88	0.48	0.77	0.57	0.80	0.83	0.73	0.64	0.60	0.62
GESSVDD-Sw-SR-max	0.56	0.20	0.80	0.45	0.21	0.68	0.16	0.21	0.85	0.52	0.47	0.61	0.53	0.57
GESSVDD-Sw-$\mathcal{S}$-min	**1.00**	**0.99**	0.14	0.78	0.52	0.96	0.79	0.56	**1.00**	0.99	0.77	**0.76**	0.67	0.72
GESSVDD-Sw-GR-min	0.48	0.52	0.45	0.73	**0.98**	0.61	0.55	0.66	0.15	0.64	0.58	0.67	0.72	0.70
GESSVDD-Sw-SR-min	0.56	0.20	0.80	0.45	0.21	0.68	0.16	0.21	0.85	0.52	0.47	0.69	0.60	0.64
GESSVDD-kNN-$\mathcal{S}$-max	0.92	0.77	0.47	0.94	0.45	0.69	0.69	0.74	0.73	0.75	0.72	0.61	0.68	0.64
GESSVDD-kNN-GR-max	0.38	0.68	0.51	0.51	0.76	0.53	0.82	0.60	0.63	0.45	0.59	0.75	0.55	0.65
GESSVDD-kNN-SR-max	0.11	0.20	0.96	0.19	0.22	0.48	0.21	0.21	0.34	0.24	0.32	0.59	0.58	0.59
GESSVDD-kNN-$\mathcal{S}$-min	0.99	0.60	0.30	**1.00**	0.23	0.25	0.61	0.57	0.76	0.88	0.62	0.67	0.69	0.68
GESSVDD-kNN-GR-min	0.60	0.82	0.33	0.40	0.74	0.66	0.57	0.78	0.69	0.65	0.62	0.61	0.61	0.61
GESSVDD-kNN-SR-min	0.11	0.20	0.96	0.19	0.22	0.48	0.21	0.21	0.34	0.24	0.32	0.60	0.56	0.58
GESSVDD-PCA-$\mathcal{S}$-max	0.81	0.67	0.85	0.83	0.82	0.69	0.95	0.46	0.76	0.83	0.77	0.72	0.63	0.68
GESSVDD-PCA-GR-max	0.55	0.75	0.81	0.71	0.73	0.73	0.59	0.92	0.72	0.72	0.72	0.61	0.66	0.63
GESSVDD-PCA-SR-max	0.04	0.20	0.05	0.33	0.22	0.20	0.16	0.91	0.58	0.21	0.29	0.64	0.73	0.68
GESSVDD-PCA-$\mathcal{S}$-min	0.59	0.01	0.79	0.82	0.40	0.68	0.99	0.12	0.86	0.95	0.62	0.66	0.72	0.69
GESSVDD-PCA-GR-min	0.40	0.32	0.71	0.50	0.63	0.70	0.49	0.99	0.23	0.79	0.58	0.69	0.66	0.68
GESSVDD-PCA-SR-min	0.04	0.20	0.05	0.33	0.22	0.20	0.16	0.91	0.58	0.21	0.29	0.68	0.74	0.71
GESSVDD-I-$\mathcal{S}$-max	0.70	0.67	0.83	0.83	0.70	0.68	0.58	0.46	0.79	0.83	0.71	0.58	0.73	0.66
GESSVDD-I-GR-max	0.64	0.66	0.82	0.91	0.81	0.48	0.54	0.57	0.64	0.79	0.69	0.65	0.58	0.61
GESSVDD-I-SR-max	0.04	0.20	0.61	0.22	0.22	0.73	0.16	0.10	0.21	0.21	0.27	0.65	0.58	0.61
GESSVDD-I-$\mathcal{S}$-min	0.68	0.01	0.76	0.87	0.70	0.67	0.63	0.12	0.84	0.95	0.62	0.61	0.74	0.68
GESSVDD-I-GR-min (ESSVDD)	0.01	0.05	0.79	0.03	0.78	0.58	0.76	0.67	0.37	0.74	0.48	0.60	0.51	0.55
GESSVDD-I-SR-min	0.04	0.20	0.61	0.22	0.22	0.73	0.16	0.10	0.21	0.21	0.27	0.70	0.63	0.67
GESSVDD-0-$\mathcal{S}$-max	0.78	0.91	0.94	0.72	0.77	0.60	0.92	0.72	0.89	0.95	0.82	0.64	0.63	0.64
GESSVDD-0-GR-max	0.40	0.66	0.52	0.61	0.57	0.51	0.57	0.65	0.56	0.83	0.59	0.64	0.56	0.60
GESSVDD-0-$\mathcal{S}$-min	0.61	0.00	0.03	0.87	0.93	0.77	0.99	0.42	0.98	**1.00**	0.66	0.63	0.63	0.63
GESSVDD-0-GR-min (SSVDD)	0.11	0.71	0.33	0.76	0.73	0.68	0.76	0.59	0.70	0.94	0.63	0.63	0.61	0.62
ESVDD	0.05	0.04	0.03	0.01	0.02	0.01	0.03	0.02	0.01	0.02	0.02	0.57	0.46	0.51
SVDD	0.97	0.98	0.54	0.97	**0.98**	**0.97**	0.97	0.96	0.98	0.97	**0.93**	0.51	0.60	0.55
OCSVM	0.62	0.52	0.54	0.75	0.61	0.46	0.53	0.66	0.59	0.48	0.58	0.73	**0.92**	**0.83**
GESVDD-PCA	0.10	0.05	0.23	0.26	0.21	0.33	0.12	0.16	0.32	0.13	0.19	0.49	0.51	0.50
GESVDD-Sw	0.11	0.05	0.22	0.26	0.21	0.34	**1.00**	0.14	0.37	0.13	0.28	0.49	0.48	0.49
GESVDD-kNN	0.11	0.05	0.23	0.27	0.21	0.36	0.15	0.14	0.29	0.13	0.19	0.49	0.48	0.48
GESVM-PCA	0.10	0.10	0.19	0.29	0.11	0.42	0.17	0.15	0.25	0.10	0.19	0.50	0.45	0.48
GESVM-Sw	0.08	0.10	0.28	0.29	0.19	0.28	0.20	0.11	0.32	0.18	0.20	0.50	0.49	0.49
GESVM-kNN	0.09	0.10	0.19	0.30	0.10	0.42	0.17	0.16	0.34	0.13	0.20	0.56	0.53	0.54

Table 37: *FNR* results for **non-linear** data description in the proposed framework with added noise

Dataset	Heart Clean train set Clean test set			Heart Clean train set Corrupted test set			Heart Corrupted train set Clean test set			Heart Corrupted train set Corrupted test set		
Target class	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.71	0.71	0.71	0.58	0.71	0.64	0.87	0.95	0.91	0.65	0.60	0.63
GESSVDD-Sb-GR-max	0.66	0.66	0.66	0.65	0.77	0.71	0.67	0.69	0.68	**0.73**	0.55	0.64
GESSVDD-Sb-SR-max	0.68	0.68	0.68	0.78	0.76	0.77	0.70	0.78	0.74	0.57	0.56	0.56
GESSVDD-Sb-$\mathcal{S}$-min	0.59	0.59	0.59	0.56	0.81	0.68	0.81	0.81	0.81	0.65	**0.80**	**0.72**
GESSVDD-Sb-GR-min	0.55	0.55	0.55	0.68	0.82	0.75	0.79	0.68	0.74	0.67	0.67	0.67
GESSVDD-Sb-SR-min	0.54	0.54	0.54	0.75	0.87	0.81	0.64	0.84	0.74	0.57	0.56	0.57
GESSVDD-Sw-$\mathcal{S}$-max	0.57	0.57	0.57	**1.00**	0.88	0.94	0.87	**1.00**	0.94	0.64	0.66	0.65
GESSVDD-Sw-GR-max	**0.84**	**0.84**	**0.84**	0.64	0.61	0.62	0.93	0.89	0.91	0.58	0.61	0.59
GESSVDD-Sw-SR-max	0.57	0.57	0.57	0.89	0.94	0.92	**1.00**	0.94	0.97	0.57	0.59	0.58
GESSVDD-Sw-$\mathcal{S}$-min	0.66	0.66	0.66	**1.00**	0.84	0.92	**1.00**	0.85	0.92	0.56	0.57	0.57
GESSVDD-Sw-GR-min	0.60	0.60	0.60	0.82	0.71	0.76	0.96	0.95	0.95	0.55	0.56	0.55
GESSVDD-Sw-SR-min	0.56	0.56	0.56	0.97	0.81	0.89	**1.00**	**1.00**	**1.00**	0.59	0.67	0.63
GESSVDD-kNN-$\mathcal{S}$-max	0.56	0.56	0.56	**1.00**	**1.00**	**1.00**	0.86	0.96	0.91	0.55	0.55	0.55
GESSVDD-kNN-GR-max	0.57	0.57	0.57	0.72	0.81	0.76	0.88	**1.00**	0.94	0.56	0.62	0.59
GESSVDD-kNN-SR-max	0.58	0.58	0.58	0.96	0.97	0.97	**1.00**	**1.00**	**1.00**	0.53	0.60	0.56
GESSVDD-kNN-$\mathcal{S}$-min	0.51	0.51	0.51	0.97	**1.00**	0.98	0.97	**1.00**	0.98	0.54	0.57	0.56
GESSVDD-kNN-GR-min	0.53	0.53	0.53	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.52	0.57	0.54
GESSVDD-kNN-SR-min	0.53	0.53	0.53	0.96	0.93	0.94	0.94	**1.00**	0.97	0.55	0.62	0.59
GESSVDD-PCA-$\mathcal{S}$-max	0.63	0.63	0.63	**1.00**	0.97	0.99	**1.00**	**1.00**	**1.00**	0.51	0.64	0.58
GESSVDD-PCA-GR-max	0.53	0.53	0.53	0.70	0.66	0.68	0.87	0.84	0.86	0.55	0.64	0.60
GESSVDD-PCA-SR-max	0.54	0.54	0.54	0.81	0.95	0.88	0.94	0.91	0.92	0.66	0.59	0.63
GESSVDD-PCA-$\mathcal{S}$-min	0.76	0.76	0.76	0.94	0.89	0.92	**1.00**	**1.00**	**1.00**	0.57	0.61	0.59
GESSVDD-PCA-GR-min	0.70	0.70	0.70	0.72	0.78	0.75	**1.00**	0.96	0.98	0.54	0.53	0.53
GESSVDD-PCA-SR-min	0.56	0.56	0.56	0.90	0.93	0.91	0.97	**1.00**	0.98	0.63	0.59	0.61
GESSVDD-I-$\mathcal{S}$-max	0.62	0.62	0.62	0.94	0.97	0.96	0.96	**1.00**	0.98	0.55	0.62	0.58
GESSVDD-I-GR-max	0.58	0.58	0.58	0.63	0.65	0.64	0.97	0.95	0.96	0.52	0.69	0.61
GESSVDD-I-SR-max	0.51	0.51	0.51	0.87	0.97	0.92	**1.00**	**1.00**	**1.00**	0.57	0.56	0.57
GESSVDD-I-$\mathcal{S}$-min	0.56	0.56	0.56	**1.00**	0.75	0.88	0.92	**1.00**	0.96	0.64	0.61	0.62
GESSVDD-I-GR-min (ESSVDD)	0.62	0.62	0.62	0.63	0.85	0.74	**1.00**	0.89	0.94	0.62	0.54	0.58
GESSVDD-I-SR-min	0.54	0.54	0.54	0.85	0.86	0.86	**1.00**	0.92	0.96	0.58	0.55	0.56
GESSVDD-0-$\mathcal{S}$-max	0.65	0.65	0.65	0.76	0.79	0.78	0.74	0.68	0.71	0.61	0.64	0.62
GESSVDD-0-GR-max	0.55	0.55	0.55	0.94	0.78	0.86	0.94	0.56	0.75	0.54	0.64	0.59
GESSVDD-0-$\mathcal{S}$-min	0.68	0.68	0.68	0.86	0.84	0.85	0.64	0.91	0.77	0.66	0.60	0.63
GESSVDD-0-GR-min (SSVDD)	0.47	0.47	0.47	0.88	0.76	0.82	**1.00**	0.91	0.96	0.63	0.53	0.58
ESVDD	0.66	0.66	0.66	**1.00**	**1.00**	**1.00**	**1.00**	0.97	0.98	0.51	0.48	0.49
SVDD	0.47	0.47	0.47	0.93	0.85	0.89	0.78	0.93	0.85	0.53	0.55	0.54
OCSVM	0.80	0.80	0.80	**1.00**	**1.00**	**1.00**	**1.00**	0.93	0.97	0.50	0.51	0.50
GESVDD-PCA	0.32	0.32	0.32	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.49	0.47	0.48
GESVDD-Sw	0.32	0.32	0.32	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.47	0.48	0.48
GESVDD-kNN	0.30	0.30	0.30	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.52	0.48	0.50
GESVM-PCA	0.34	0.34	0.34	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.51	0.51	0.51
GESVM-Sw	0.33	0.33	0.33	**1.00**	**1.00**	**1.00**	**1.00**	0.95	0.97	0.50	0.52	0.51
GESVM-kNN	0.40	0.40	0.40	**1.00**	**1.00**	**1.00**	0.00	0.00	0.00	0.55	0.55	0.55

S4. *Standard deviation* of the *Gmean* results for linear methods

Table 38: *Standard deviation* of the *Gmean* results for **linear** data description over Seeds, Qualitative bankruptcy, Somerville happiness and Liver datasets

Dataset	Seeds				Qualitative bankruptcy			Somerville happiness			Liver		
Target class	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.03	0.09	0.08	0.07	0.15	0.24	0.19	0.12	0.06	0.09	0.04	0.10	0.07
GESSVDD-Sb-GR-max	0.12	0.36	0.38	0.29	0.11	0.18	0.15	0.17	0.15	0.16	0.07	0.07	0.07
GESSVDD-Sb-SR-max	0.09	0.03	0.03	0.05	0.36	0.09	0.22	0.03	0.14	0.09	0.05	0.12	0.08
GESSVDD-Sb-$\mathcal{S}$-min	0.08	0.05	0.07	0.07	0.18	0.25	0.21	0.17	0.08	0.12	0.17	0.04	0.11
GESSVDD-Sb-GR-min	0.04	0.32	0.05	0.14	0.14	0.14	0.14	0.12	0.21	0.17	0.06	0.07	0.06
GESSVDD-Sb-SR-min	0.10	0.02	0.03	0.05	0.36	0.21	0.28	0.13	0.13	0.13	0.05	0.06	0.06
GESSVDD-Sw-$\mathcal{S}$-max	0.04	0.04	0.05	0.04	0.06	0.00	0.03	0.18	0.12	0.15	0.08	0.05	0.06
GESSVDD-Sw-GR-max	0.03	0.04	0.03	0.03	0.14	0.08	0.11	0.25	0.12	0.19	0.04	0.04	0.04
GESSVDD-Sw-SR-max	0.10	0.05	0.15	0.10	0.17	0.15	0.16	0.14	0.06	0.10	0.06	0.04	0.05
GESSVDD-Sw-$\mathcal{S}$-min	0.06	0.03	0.03	0.04	0.02	0.16	0.09	0.07	0.12	0.09	0.10	0.07	0.08
GESSVDD-Sw-GR-min	0.06	0.19	0.04	0.10	0.08	0.15	0.11	0.14	0.11	0.12	0.06	0.03	0.04
GESSVDD-Sw-SR-min	0.14	0.02	0.03	0.06	0.08	0.13	0.11	0.08	0.07	0.08	0.07	0.04	0.06
GESSVDD-kNN-$\mathcal{S}$-max	0.03	0.04	0.05	0.04	0.07	0.15	0.11	0.12	0.07	0.09	0.09	0.05	0.07
GESSVDD-kNN-GR-max	0.04	0.05	0.05	0.05	0.12	0.26	0.19	0.07	0.17	0.12	0.09	0.13	0.11
GESSVDD-kNN-SR-max	0.09	0.03	0.02	0.04	0.07	0.19	0.13	0.08	0.12	0.10	0.05	0.07	0.06
GESSVDD-kNN-$\mathcal{S}$-min	0.07	0.04	0.05	0.05	0.07	0.18	0.12	0.08	0.08	0.08	0.22	0.06	0.14
GESSVDD-kNN-GR-min	0.04	0.03	0.03	0.03	0.04	0.12	0.08	0.08	0.26	0.17	0.15	0.15	0.15
GESSVDD-kNN-SR-min	0.04	0.03	0.03	0.03	0.06	0.18	0.12	0.05	0.08	0.06	0.16	0.08	0.12
GESSVDD-PCA-$\mathcal{S}$-max	0.05	0.04	0.04	0.04	0.04	0.20	0.12	0.09	0.08	0.08	0.07	0.04	0.06
GESSVDD-PCA-GR-max	0.03	0.04	0.04	0.04	0.03	0.14	0.09	0.13	0.08	0.11	0.07	0.06	0.07
GESSVDD-PCA-SR-max	0.04	0.04	0.03	0.04	0.05	0.18	0.12	0.06	0.09	0.08	0.03	0.05	0.04
GESSVDD-PCA-$\mathcal{S}$-min	0.05	0.03	0.06	0.05	0.06	0.08	0.07	0.10	0.08	0.09	0.09	0.05	0.07
GESSVDD-PCA-GR-min	0.04	0.05	0.19	0.09	0.05	0.12	0.08	0.12	0.09	0.10	0.09	0.05	0.07
GESSVDD-PCA-SR-min	0.08	0.04	0.03	0.05	0.07	0.21	0.14	0.13	0.03	0.08	0.06	0.03	0.04
GESSVDD-I-$\mathcal{S}$-max	0.04	0.04	0.04	0.04	0.05	0.20	0.13	0.08	0.06	0.07	0.10	0.03	0.06
GESSVDD-I-GR-max	0.05	0.04	0.02	0.04	0.02	0.13	0.07	0.06	0.14	0.10	0.09	0.04	0.06
GESSVDD-I-SR-max	0.05	0.05	0.04	0.05	0.05	0.09	0.07	0.10	0.06	0.08	0.10	0.06	0.08
GESSVDD-I-$\mathcal{S}$-min	0.06	0.03	0.05	0.05	0.06	0.19	0.13	0.11	0.08	0.09	0.10	0.05	0.07
GESSVDD-I-GR-min (ESSVDD)	0.04	0.05	0.06	0.05	0.06	0.11	0.08	0.06	0.06	0.06	0.08	0.04	0.06
GESSVDD-I-SR-min	0.04	0.05	0.04	0.04	0.01	0.20	0.11	0.11	0.06	0.09	0.08	0.06	0.07
GESSVDD-0-$\mathcal{S}$-max	0.04	0.04	0.03	0.04	0.03	0.00	0.01	0.22	0.09	0.16	0.03	0.04	0.03
GESSVDD-0-GR-max	0.03	0.04	0.03	0.03	0.05	0.00	0.03	0.24	0.10	0.17	0.03	0.03	0.03
GESSVDD-0-$\mathcal{S}$-min	0.14	0.04	0.06	0.08	0.05	0.15	0.10	0.08	0.07	0.08	0.08	0.09	0.08
GESSVDD-0-GR-min (SSVDD)	0.03	0.04	0.02	0.03	0.06	0.18	0.12	0.11	0.10	0.10	0.02	0.04	0.03
ESVDD	0.07	0.05	0.10	0.08	0.02	0.15	0.09	0.07	0.04	0.05	0.07	0.04	0.06
SVDD	0.03	0.05	0.04	0.04	0.03	0.00	0.02	0.15	0.07	0.11	0.09	0.09	0.09
OCSVM	0.30	0.16	0.35	0.27	0.05	0.17	0.11	0.09	0.12	0.11	0.09	0.03	0.06

Table 39: *Standard deviation* of the *Gmean* results for **linear** data description over Iris, ionosphere, and Sonar datasets

Dataset	Iris				Ionosphere			Sonar		
Target class	I-S	S-VC	S-V	Av.	I-B	I-G	Av.	S-R	S-M	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.04	0.10	0.06	0.07	0.21	0.27	0.24	0.13	0.14	0.13
GESSVDD-Sb-GR-max	0.09	0.04	0.04	0.06	0.29	0.07	0.18	0.10	0.11	0.11
GESSVDD-Sb-SR-max	0.23	0.06	0.14	0.14	0.12	0.35	0.23	0.20	0.12	0.16
GESSVDD-Sb-$\mathcal{S}$-min	0.04	0.10	0.03	0.06	0.23	0.28	0.26	0.11	0.13	0.12
GESSVDD-Sb-GR-min	0.03	0.07	0.05	0.05	0.23	0.07	0.15	0.09	0.11	0.10
GESSVDD-Sb-SR-min	0.23	0.39	0.10	0.24	0.25	0.41	0.33	0.22	0.27	0.25
GESSVDD-Sw-$\mathcal{S}$-max	0.06	0.04	0.06	0.05	0.14	0.01	0.08	0.06	0.13	0.09
GESSVDD-Sw-GR-max	0.08	0.06	0.07	0.07	0.11	0.03	0.07	0.06	0.16	0.11
GESSVDD-Sw-SR-max	0.02	0.09	0.05	0.05	0.08	0.02	0.05	0.06	0.14	0.10
GESSVDD-Sw-$\mathcal{S}$-min	0.06	0.06	0.09	0.07	0.10	0.03	0.06	0.07	0.11	0.09
GESSVDD-Sw-GR-min	0.02	0.06	0.04	0.04	0.09	0.03	0.06	0.08	0.13	0.10
GESSVDD-Sw-SR-min	0.02	0.10	0.07	0.06	0.17	0.01	0.09	0.05	0.14	0.09
GESSVDD-kNN-$\mathcal{S}$-max	0.03	0.06	0.06	0.05	0.16	0.04	0.10	0.03	0.03	0.03
GESSVDD-kNN-GR-max	0.04	0.27	0.21	0.17	0.10	0.03	0.07	0.06	0.07	0.06
GESSVDD-kNN-SR-max	0.02	0.04	0.04	0.03	0.18	0.04	0.11	0.02	0.06	0.04
GESSVDD-kNN-$\mathcal{S}$-min	0.04	0.05	0.05	0.05	0.16	0.04	0.10	0.03	0.03	0.03
GESSVDD-kNN-GR-min	0.04	0.06	0.04	0.05	0.18	0.02	0.10	0.09	0.08	0.08
GESSVDD-kNN-SR-min	0.02	0.06	0.05	0.04	0.10	0.04	0.07	0.02	0.07	0.05
GESSVDD-PCA-$\mathcal{S}$-max	0.02	0.03	0.07	0.04	0.13	0.02	0.07	0.04	0.08	0.06
GESSVDD-PCA-GR-max	0.05	0.05	0.06	0.05	0.13	0.02	0.08	0.08	0.06	0.07
GESSVDD-PCA-SR-max	0.02	0.03	0.07	0.04	0.07	0.02	0.04	0.23	0.15	0.19
GESSVDD-PCA-$\mathcal{S}$-min	0.04	0.03	0.06	0.05	0.09	0.03	0.06	0.04	0.12	0.08
GESSVDD-PCA-GR-min	0.04	0.15	0.07	0.09	0.10	0.01	0.06	0.04	0.03	0.03
GESSVDD-PCA-SR-min	0.02	0.04	0.05	0.04	0.12	0.03	0.07	0.06	0.15	0.10
GESSVDD-I-$\mathcal{S}$-max	0.02	0.03	0.07	0.04	0.07	0.03	0.05	0.04	0.15	0.10
GESSVDD-I-GR-max	0.04	0.06	0.10	0.06	0.08	0.02	0.05	0.05	0.06	0.05
GESSVDD-I-SR-max	0.02	0.03	0.20	0.08	0.16	0.03	0.09	0.05	0.15	0.10
GESSVDD-I-$\mathcal{S}$-min	0.04	0.03	0.10	0.06	0.05	0.03	0.04	0.04	0.15	0.09
GESSVDD-I-GR-min (ESSVDD)	0.04	0.14	0.12	0.10	0.08	0.02	0.05	0.04	0.05	0.04
GESSVDD-I-SR-min	0.03	0.03	0.07	0.05	0.07	0.02	0.05	0.02	0.15	0.09
GESSVDD-0-$\mathcal{S}$-max	0.06	0.06	0.03	0.05	0.12	0.08	0.10	0.12	0.10	0.11
GESSVDD-0-GR-max	0.04	0.04	0.03	0.04	0.10	0.04	0.07	0.10	0.12	0.11
GESSVDD-0-$\mathcal{S}$-min	0.08	0.06	0.10	0.08	0.20	0.06	0.13	0.07	0.12	0.09
GESSVDD-0-GR-min (SSVDD)	0.05	0.04	0.03	0.04	0.12	0.04	0.08	0.06	0.09	0.08
ESVDD	0.03	0.08	0.07	0.06	0.05	0.04	0.04	0.00	0.08	0.04
SVDD	0.03	0.03	0.03	0.03	0.05	0.02	0.03	0.08	0.11	0.09
OCSVM	0.33	0.11	0.28	0.24	0.12	0.04	0.08	0.08	0.06	0.07

Table 40: *Standard deviation* of the *Gmean* results for **linear** data description over Heart dataset and its manually created corrupted versions

Dataset	Heart Clean train set Clean test set			Heart Clean train set Corrupted test set			Heart Corrupted train set Clean test set			Heart Corrupted train set Corrupted test set		
Target class	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.12	0.25	0.19	0.14	0.14	0.14	0.16	0.09	0.13	0.12	0.12	0.12
GESSVDD-Sb-GR-max	0.17	0.32	0.25	0.18	0.16	0.17	0.23	0.22	0.23	0.10	0.14	0.12
GESSVDD-Sb-SR-max	0.18	0.15	0.16	0.14	0.17	0.15	0.24	0.21	0.22	0.06	0.14	0.10
GESSVDD-Sb-$\mathcal{S}$-min	0.18	0.20	0.19	0.11	0.18	0.15	0.11	0.13	0.12	0.14	0.11	0.12
GESSVDD-Sb-GR-min	0.16	0.04	0.10	0.16	0.17	0.16	0.21	0.21	0.21	0.13	0.09	0.11
GESSVDD-Sb-SR-min	0.11	0.15	0.13	0.18	0.17	0.17	0.27	0.08	0.18	0.08	0.11	0.09
GESSVDD-Sw-$\mathcal{S}$-max	0.08	0.09	0.09	0.20	0.14	0.17	0.00	0.00	0.00	0.12	0.08	0.10
GESSVDD-Sw-GR-max	0.15	0.04	0.10	0.28	0.26	0.27	0.00	0.00	0.00	0.17	0.05	0.11
GESSVDD-Sw-SR-max	0.11	0.12	0.12	0.06	0.00	0.03	0.00	0.00	0.00	0.08	0.05	0.06
GESSVDD-Sw-$\mathcal{S}$-min	0.16	0.08	0.12	0.14	0.17	0.16	0.00	0.00	0.00	0.05	0.07	0.06
GESSVDD-Sw-GR-min	0.09	0.09	0.09	0.28	0.27	0.28	0.00	0.00	0.00	0.08	0.07	0.08
GESSVDD-Sw-SR-min	0.06	0.06	0.06	0.00	0.06	0.03	0.06	0.00	0.03	0.04	0.07	0.05
GESSVDD-kNN-$\mathcal{S}$-max	0.06	0.13	0.10	0.09	0.13	0.11	0.08	0.00	0.04	0.15	0.07	0.11
GESSVDD-kNN-GR-max	0.09	0.09	0.09	0.19	0.19	0.19	0.24	0.19	0.21	0.13	0.07	0.10
GESSVDD-kNN-SR-max	0.15	0.11	0.13	0.00	0.08	0.04	0.00	0.00	0.00	0.11	0.09	0.10
GESSVDD-kNN-$\mathcal{S}$-min	0.11	0.16	0.14	0.11	0.06	0.09	0.17	0.00	0.08	0.10	0.09	0.09
GESSVDD-kNN-GR-min	0.09	0.09	0.09	0.13	0.00	0.06	0.30	0.19	0.24	0.05	0.07	0.06
GESSVDD-kNN-SR-min	0.06	0.12	0.09	0.00	0.19	0.10	0.00	0.00	0.00	0.10	0.06	0.08
GESSVDD-PCA-$\mathcal{S}$-max	0.12	0.10	0.11	0.11	0.00	0.05	0.00	0.00	0.00	0.08	0.07	0.07
GESSVDD-PCA-GR-max	0.07	0.06	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.07	0.09
GESSVDD-PCA-SR-max	0.10	0.06	0.08	0.00	0.06	0.03	0.00	0.00	0.00	0.13	0.08	0.10
GESSVDD-PCA-$\mathcal{S}$-min	0.13	0.08	0.11	0.00	0.19	0.09	0.00	0.00	0.00	0.05	0.06	0.05
GESSVDD-PCA-GR-min	0.05	0.15	0.10	0.27	0.00	0.14	0.00	0.00	0.00	0.09	0.03	0.06
GESSVDD-PCA-SR-min	0.10	0.09	0.09	0.16	0.12	0.14	0.00	0.00	0.00	0.11	0.10	0.10
GESSVDD-I-$\mathcal{S}$-max	0.11	0.10	0.11	0.13	0.12	0.13	0.00	0.00	0.00	0.11	0.07	0.09
GESSVDD-I-GR-max	0.11	0.15	0.13	0.28	0.21	0.25	0.00	0.00	0.00	0.12	0.04	0.08
GESSVDD-I-SR-max	0.09	0.08	0.09	0.06	0.06	0.06	0.00	0.00	0.00	0.12	0.08	0.10
GESSVDD-I-$\mathcal{S}$-min	0.06	0.14	0.10	0.15	0.06	0.11	0.00	0.00	0.00	0.04	0.06	0.05
GESSVDD-I-GR-min (ESSVDD)	0.11	0.08	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.08	0.08
GESSVDD-I-SR-min	0.10	0.12	0.11	0.06	0.00	0.03	0.00	0.00	0.00	0.04	0.15	0.09
GESSVDD-0-$\mathcal{S}$-max	0.11	0.14	0.13	0.15	0.21	0.18	0.00	0.00	0.00	0.19	0.10	0.14
GESSVDD-0-GR-max	0.12	0.09	0.10	0.21	0.24	0.23	0.00	0.00	0.00	0.09	0.08	0.09
GESSVDD-0-$\mathcal{S}$-min	0.07	0.04	0.05	0.20	0.00	0.10	0.00	0.00	0.00	0.11	0.05	0.08
GESSVDD-0-GR-min (SSVDD)	0.10	0.11	0.10	0.00	0.24	0.12	0.00	0.00	0.00	0.09	0.05	0.07
ESVDD	0.07	0.05	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.03	0.06
SVDD	0.17	0.10	0.14	0.11	0.08	0.10	0.31	0.09	0.20	0.09	0.09	0.09
OCSVM	0.08	0.06	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.06	0.07

S5. *Standard deviation* of the *Gmean* results for non-linear methods

Table 41: *Standard deviation* of the *Gmean* results for **non-linear** data description over Seeds, Qualitative bankruptcy, Somerville happiness and Liver datasets

Dataset	Seeds				Qualitative bankruptcy			Somerville happiness			Liver		
Target class	S-K	S-R	S-C	Av.	QB-B	QB-N	Av.	SH-H	SH-U	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.10	0.03	0.01	0.05	0.10	0.19	0.15	0.12	0.09	0.11	0.06	0.10	0.08
GESSVDD-Sb-GR-max	0.06	0.32	0.09	0.16	0.08	0.34	0.21	0.13	0.17	0.15	0.20	0.07	0.13
GESSVDD-Sb-SR-max	0.08	0.11	0.05	0.08	0.34	0.28	0.31	0.16	0.18	0.17	0.21	0.13	0.17
GESSVDD-Sb-$\mathcal{S}$-min	0.09	0.04	0.08	0.07	0.09	0.27	0.18	0.25	0.19	0.22	0.10	0.10	0.10
GESSVDD-Sb-GR-min	0.06	0.43	0.10	0.20	0.08	0.14	0.11	0.12	0.14	0.13	0.08	0.22	0.15
GESSVDD-Sb-SR-min	0.05	0.09	0.04	0.06	0.34	0.05	0.20	0.09	0.18	0.13	0.21	0.24	0.22
GESSVDD-Sw-$\mathcal{S}$-max	0.10	0.08	0.05	0.07	0.10	0.33	0.21	0.05	0.09	0.07	0.08	0.11	0.10
GESSVDD-Sw-GR-max	0.06	0.17	0.07	0.10	0.08	0.20	0.14	0.12	0.07	0.10	0.21	0.13	0.17
GESSVDD-Sw-SR-max	0.03	0.05	0.08	0.06	0.09	0.23	0.16	0.15	0.10	0.12	0.10	0.15	0.13
GESSVDD-Sw-$\mathcal{S}$-min	0.07	0.08	0.05	0.07	0.10	0.21	0.15	0.11	0.06	0.09	0.23	0.21	0.22
GESSVDD-Sw-GR-min	0.04	0.05	0.04	0.04	0.12	0.07	0.09	0.12	0.07	0.09	0.19	0.18	0.19
GESSVDD-Sw-SR-min	0.03	0.06	0.08	0.06	0.04	0.10	0.07	0.14	0.10	0.12	0.18	0.18	0.18
GESSVDD-kNN-$\mathcal{S}$-max	0.05	0.03	0.09	0.06	0.14	0.13	0.14	0.12	0.09	0.11	0.16	0.23	0.20
GESSVDD-kNN-GR-max	0.06	0.06	0.06	0.06	0.05	0.26	0.15	0.07	0.05	0.06	0.18	0.11	0.15
GESSVDD-kNN-SR-max	0.06	0.02	0.03	0.04	0.06	0.08	0.07	0.09	0.10	0.09	0.11	0.12	0.12
GESSVDD-kNN-$\mathcal{S}$-min	0.05	0.07	0.03	0.05	0.08	0.07	0.07	0.10	0.15	0.13	0.22	0.19	0.20
GESSVDD-kNN-GR-min	0.05	0.03	0.05	0.04	0.08	0.17	0.12	0.08	0.07	0.07	0.13	0.15	0.14
GESSVDD-kNN-SR-min	0.06	0.06	0.05	0.06	0.17	0.09	0.13	0.08	0.18	0.13	0.15	0.11	0.13
GESSVDD-PCA-$\mathcal{S}$-max	0.06	0.06	0.04	0.05	0.26	0.22	0.24	0.10	0.07	0.08	0.20	0.10	0.15
GESSVDD-PCA-GR-max	0.09	0.06	0.03	0.06	0.04	0.22	0.13	0.11	0.07	0.09	0.23	0.21	0.22
GESSVDD-PCA-SR-max	0.04	0.42	0.04	0.16	0.02	0.23	0.13	0.10	0.06	0.08	0.22	0.17	0.19
GESSVDD-PCA-$\mathcal{S}$-min	0.05	0.04	0.04	0.04	0.12	0.44	0.28	0.06	0.03	0.04	0.22	0.05	0.13
GESSVDD-PCA-GR-min	0.04	0.06	0.05	0.05	0.04	0.16	0.10	0.12	0.05	0.09	0.20	0.12	0.16
GESSVDD-PCA-SR-min	0.04	0.09	0.04	0.06	0.02	0.18	0.10	0.07	0.12	0.10	0.18	0.15	0.17
GESSVDD-I-$\mathcal{S}$-max	0.06	0.03	0.42	0.17	0.08	0.38	0.23	0.07	0.08	0.08	0.12	0.18	0.15
GESSVDD-I-GR-max	0.09	0.05	0.02	0.06	0.05	0.12	0.09	0.12	0.07	0.10	0.20	0.12	0.16
GESSVDD-I-SR-max	0.06	0.03	0.03	0.04	0.09	0.31	0.20	0.11	0.16	0.14	0.21	0.15	0.18
GESSVDD-I-$\mathcal{S}$-min	0.06	0.05	0.04	0.05	0.06	0.32	0.19	0.13	0.06	0.09	0.12	0.17	0.15
GESSVDD-I-GR-min (ESSVDD)	0.03	0.04	0.07	0.05	0.07	0.06	0.06	0.09	0.05	0.07	0.08	0.11	0.09
GESSVDD-I-SR-min	0.06	0.03	0.04	0.04	0.09	0.30	0.19	0.10	0.13	0.11	0.19	0.07	0.13
GESSVDD-0-$\mathcal{S}$-max	0.03	0.04	0.03	0.03	0.16	0.07	0.12	0.23	0.14	0.19	0.10	0.21	0.16
GESSVDD-0-GR-max	0.03	0.02	0.04	0.03	0.04	0.12	0.08	0.22	0.16	0.19	0.21	0.05	0.13
GESSVDD-0-$\mathcal{S}$-min	0.02	0.04	0.01	0.03	0.16	0.15	0.15	0.22	0.22	0.22	0.22	0.13	0.18
GESSVDD-0-GR-min (SSVDD)	0.02	0.03	0.04	0.03	0.03	0.27	0.15	0.07	0.11	0.09	0.21	0.13	0.17
ESVDD	0.06	0.06	0.12	0.08	0.00	0.00	0.00	0.00	0.19	0.10	0.24	0.07	0.16
SVDD	0.02	0.06	0.03	0.04	0.32	0.26	0.29	0.15	0.13	0.14	0.05	0.24	0.14
OCSVM	0.20	0.28	0.29	0.26	0.26	0.11	0.18	0.02	0.08	0.05	0.16	0.18	0.17
GESVDD-PCA	0.04	0.04	0.03	0.04	0.03	0.17	0.10	0.12	0.05	0.09	0.05	0.07	0.06
GESVDD-Sw	0.05	0.04	0.03	0.04	0.02	0.17	0.09	0.10	0.10	0.10	0.05	0.04	0.04
GESVDD-kNN	0.06	0.04	0.03	0.04	0.06	0.10	0.08	0.13	0.07	0.10	0.04	0.04	0.04
GESVM-PCA	0.02	0.04	0.04	0.04	0.05	0.06	0.05	0.07	0.07	0.07	0.07	0.04	0.05
GESVM-Sw	0.03	0.05	0.05	0.04	0.06	0.13	0.10	0.02	0.10	0.06	0.07	0.04	0.05
GESVM-kNN	0.03	0.05	0.08	0.06	0.05	0.07	0.06	0.08	0.07	0.08	0.08	0.05	0.07

Table 42: *Standard deviation* of the *Gmean* results for **non-linear** data description over Iris, ionosphere, and Sonar datasets

Dataset	Iris				Ionosphere			Sonar		
Target class	I-S	S-VC	S-V	Av.	I-B	I-G	Av.	S-R	S-M	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.06	0.05	0.16	0.09	0.12	0.11	0.12	0.11	0.24	0.17
GESSVDD-Sb-GR-max	0.03	0.07	0.07	0.06	0.26	0.08	0.17	0.12	0.17	0.14
GESSVDD-Sb-SR-max	0.18	0.09	0.08	0.12	0.14	0.24	0.19	0.10	0.23	0.17
GESSVDD-Sb-$\mathcal{S}$-min	0.07	0.04	0.06	0.06	0.22	0.30	0.26	0.10	0.12	0.11
GESSVDD-Sb-GR-min	0.03	0.04	0.04	0.04	0.29	0.08	0.19	0.12	0.16	0.14
GESSVDD-Sb-SR-min	0.20	0.06	0.28	0.18	0.18	0.24	0.21	0.10	0.22	0.16
GESSVDD-Sw-$\mathcal{S}$-max	0.25	0.07	0.09	0.14	0.33	0.03	0.18	0.06	0.11	0.09
GESSVDD-Sw-GR-max	0.03	0.04	0.03	0.03	0.28	0.03	0.15	0.12	0.15	0.13
GESSVDD-Sw-SR-max	0.02	0.04	0.05	0.04	0.23	0.03	0.13	0.21	0.06	0.14
GESSVDD-Sw-$\mathcal{S}$-min	0.29	0.05	0.06	0.13	0.27	0.03	0.15	0.06	0.11	0.08
GESSVDD-Sw-GR-min	0.04	0.09	0.05	0.06	0.10	0.03	0.06	0.06	0.10	0.08
GESSVDD-Sw-SR-min	0.02	0.04	0.05	0.04	0.28	0.03	0.15	0.21	0.16	0.19
GESSVDD-kNN-$\mathcal{S}$-max	0.03	0.06	0.04	0.04	0.16	0.05	0.10	0.07	0.13	0.10
GESSVDD-kNN-GR-max	0.03	0.05	0.09	0.06	0.15	0.02	0.09	0.06	0.15	0.11
GESSVDD-kNN-SR-max	0.05	0.11	0.05	0.07	0.16	0.02	0.09	0.07	0.12	0.09
GESSVDD-kNN-$\mathcal{S}$-min	0.04	0.06	0.06	0.05	0.10	0.05	0.07	0.08	0.24	0.16
GESSVDD-kNN-GR-min	0.08	0.06	0.03	0.06	0.18	0.02	0.10	0.07	0.07	0.07
GESSVDD-kNN-SR-min	0.05	0.11	0.04	0.07	0.16	0.02	0.09	0.08	0.10	0.09
GESSVDD-PCA-$\mathcal{S}$-max	0.03	0.02	0.09	0.04	0.09	0.04	0.06	0.09	0.21	0.15
GESSVDD-PCA-GR-max	0.06	0.05	0.06	0.05	0.21	0.02	0.12	0.07	0.12	0.10
GESSVDD-PCA-SR-max	0.03	0.05	0.06	0.05	0.06	0.09	0.08	0.17	0.05	0.11
GESSVDD-PCA-$\mathcal{S}$-min	0.03	0.05	0.06	0.05	0.06	0.03	0.05	0.12	0.07	0.09
GESSVDD-PCA-GR-min	0.06	0.15	0.07	0.09	0.03	0.03	0.03	0.07	0.11	0.09
GESSVDD-PCA-SR-min	0.03	0.05	0.10	0.06	0.18	0.03	0.10	0.02	0.19	0.11
GESSVDD-I-$\mathcal{S}$-max	0.02	0.04	0.06	0.04	0.15	0.02	0.09	0.07	0.20	0.14
GESSVDD-I-GR-max	0.04	0.08	0.04	0.06	0.17	0.02	0.09	0.03	0.04	0.03
GESSVDD-I-SR-max	0.02	0.03	0.06	0.04	0.12	0.02	0.07	0.06	0.09	0.07
GESSVDD-I-$\mathcal{S}$-min	0.02	0.03	0.06	0.04	0.21	0.02	0.12	0.04	0.06	0.05
GESSVDD-I-GR-min (ESSVDD)	0.07	0.09	0.05	0.07	0.07	0.03	0.05	0.02	0.09	0.05
GESSVDD-I-SR-min	0.02	0.03	0.06	0.04	0.14	0.02	0.08	0.05	0.09	0.07
GESSVDD-0-$\mathcal{S}$-max	0.06	0.05	0.04	0.05	0.08	0.12	0.10	0.14	0.05	0.10
GESSVDD-0-GR-max	0.04	0.04	0.03	0.04	0.26	0.05	0.16	0.09	0.10	0.09
GESSVDD-0-$\mathcal{S}$-min	0.07	0.06	0.03	0.05	0.14	0.09	0.11	0.12	0.25	0.19
GESSVDD-0-GR-min (SSVDD)	0.03	0.06	0.04	0.04	0.37	0.05	0.21	0.13	0.09	0.11
ESVDD	0.38	0.07	0.07	0.17	0.06	0.04	0.05	0.04	0.08	0.06
SVDD	0.04	0.05	0.02	0.04	0.19	0.03	0.11	0.07	0.06	0.07
OCSVM	0.33	0.32	0.18	0.28	0.13	0.16	0.14	0.08	0.11	0.09
GESVDD-PCA	0.12	0.05	0.05	0.07	0.06	0.02	0.04	0.07	0.06	0.06
GESVDD-Sw	0.04	0.07	0.04	0.05	0.04	0.02	0.03	0.07	0.10	0.08
GESVDD-kNN	0.12	0.05	0.04	0.07	0.11	0.02	0.07	0.06	0.06	0.06
GESVM-PCA	0.05	0.05	0.03	0.04	0.07	0.02	0.05	0.09	0.04	0.06
GESVM-Sw	0.06	0.04	0.02	0.04	0.12	0.03	0.07	0.10	0.05	0.08
GESVM-kNN	0.04	0.07	0.06	0.06	0.07	0.05	0.06	0.05	0.05	0.05

Table 43: *Standard deviation* of the *Gmean* results for **non-linear** data description over Heart dataset and its manually created corrupted versions

Dataset	Heart Clean train set Clean test set			Heart Clean train set Corrupted test set			Heart Corrupted train set Clean test set			Heart Corrupted train set Corrupted test set		
Target class	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.	DP	DA	Av.
GESSVDD-Sb-$\mathcal{S}$-max	0.21	0.20	0.20	0.10	0.16	0.13	0.21	0.11	0.16	0.13	0.08	0.10
GESSVDD-Sb-GR-max	0.20	0.17	0.19	0.07	0.15	0.11	0.26	0.27	0.26	0.20	0.06	0.13
GESSVDD-Sb-SR-max	0.20	0.25	0.23	0.21	0.23	0.22	0.22	0.18	0.20	0.08	0.06	0.07
GESSVDD-Sb-$\mathcal{S}$-min	0.12	0.09	0.10	0.11	0.12	0.12	0.20	0.30	0.25	0.12	0.14	0.13
GESSVDD-Sb-GR-min	0.15	0.25	0.20	0.20	0.18	0.19	0.23	0.31	0.27	0.19	0.11	0.15
GESSVDD-Sb-SR-min	0.18	0.23	0.21	0.24	0.18	0.21	0.18	0.16	0.17	0.08	0.07	0.08
GESSVDD-Sw-$\mathcal{S}$-max	0.17	0.17	0.17	0.00	0.18	0.09	0.29	0.00	0.14	0.16	0.12	0.14
GESSVDD-Sw-GR-max	0.23	0.12	0.17	0.07	0.23	0.15	0.16	0.17	0.16	0.10	0.14	0.12
GESSVDD-Sw-SR-max	0.08	0.05	0.06	0.17	0.08	0.12	0.00	0.13	0.06	0.12	0.03	0.07
GESSVDD-Sw-$\mathcal{S}$-min	0.10	0.14	0.12	0.00	0.15	0.07	0.00	0.23	0.11	0.10	0.10	0.10
GESSVDD-Sw-GR-min	0.13	0.13	0.13	0.13	0.21	0.17	0.09	0.12	0.11	0.05	0.13	0.09
GESSVDD-Sw-SR-min	0.10	0.10	0.10	0.06	0.21	0.14	0.00	0.00	0.00	0.06	0.20	0.13
GESSVDD-kNN-$\mathcal{S}$-max	0.13	0.10	0.11	0.00	0.00	0.00	0.21	0.10	0.15	0.09	0.06	0.08
GESSVDD-kNN-GR-max	0.09	0.20	0.14	0.20	0.27	0.23	0.26	0.00	0.13	0.08	0.11	0.09
GESSVDD-kNN-SR-max	0.11	0.07	0.09	0.09	0.06	0.08	0.00	0.00	0.00	0.11	0.05	0.08
GESSVDD-kNN-$\mathcal{S}$-min	0.11	0.10	0.11	0.07	0.00	0.03	0.07	0.00	0.03	0.10	0.09	0.10
GESSVDD-kNN-GR-min	0.07	0.07	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.14	0.10
GESSVDD-kNN-SR-min	0.09	0.12	0.10	0.09	0.16	0.12	0.08	0.00	0.04	0.07	0.10	0.08
GESSVDD-PCA-$\mathcal{S}$-max	0.16	0.10	0.13	0.00	0.06	0.03	0.00	0.00	0.00	0.09	0.21	0.15
GESSVDD-PCA-GR-max	0.05	0.18	0.12	0.22	0.22	0.22	0.29	0.23	0.26	0.12	0.14	0.13
GESSVDD-PCA-SR-max	0.11	0.23	0.17	0.18	0.11	0.15	0.08	0.21	0.15	0.11	0.05	0.08
GESSVDD-PCA-$\mathcal{S}$-min	0.20	0.16	0.18	0.13	0.10	0.11	0.00	0.00	0.00	0.07	0.23	0.15
GESSVDD-PCA-GR-min	0.09	0.30	0.20	0.12	0.15	0.14	0.00	0.08	0.04	0.07	0.09	0.08
GESSVDD-PCA-SR-min	0.13	0.13	0.13	0.10	0.10	0.10	0.07	0.00	0.03	0.16	0.03	0.09
GESSVDD-I-$\mathcal{S}$-max	0.11	0.09	0.10	0.08	0.06	0.07	0.09	0.00	0.05	0.09	0.13	0.11
GESSVDD-I-GR-max	0.12	0.17	0.15	0.15	0.23	0.19	0.07	0.11	0.09	0.08	0.10	0.09
GESSVDD-I-SR-max	0.09	0.26	0.18	0.09	0.07	0.08	0.00	0.00	0.00	0.11	0.04	0.08
GESSVDD-I-$\mathcal{S}$-min	0.07	0.22	0.15	0.00	0.22	0.11	0.18	0.00	0.09	0.14	0.08	0.11
GESSVDD-I-GR-min (ESSVDD)	0.12	0.12	0.12	0.11	0.15	0.13	0.00	0.18	0.09	0.09	0.05	0.07
GESSVDD-I-SR-min	0.09	0.18	0.13	0.22	0.17	0.20	0.00	0.19	0.09	0.14	0.04	0.09
GESSVDD-0-$\mathcal{S}$-max	0.11	0.13	0.12	0.10	0.22	0.16	0.29	0.24	0.26	0.09	0.17	0.13
GESSVDD-0-GR-max	0.10	0.14	0.12	0.13	0.13	0.13	0.14	0.21	0.17	0.11	0.20	0.16
GESSVDD-0-$\mathcal{S}$-min	0.10	0.21	0.15	0.18	0.16	0.17	0.15	0.13	0.14	0.20	0.17	0.19
GESSVDD-0-GR-min (SSVDD)	0.18	0.09	0.14	0.19	0.26	0.22	0.00	0.20	0.10	0.21	0.06	0.14
ESVDD	0.04	0.07	0.05	0.00	0.00	0.00	0.00	0.07	0.03	0.12	0.04	0.08
SVDD	0.07	0.04	0.06	0.15	0.21	0.18	0.35	0.10	0.22	0.06	0.18	0.12
OCSVM	0.19	0.16	0.18	0.00	0.00	0.00	0.00	0.16	0.08	0.08	0.14	0.11
GESVDD-PCA	0.03	0.07	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.07	0.06
GESVDD-Sw	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.07	0.07
GESVDD-kNN	0.03	0.05	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.08	0.05
GESVM-PCA	0.02	0.05	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.07	0.06
GESVM-Sw	0.02	0.06	0.04	0.00	0.00	0.00	0.00	0.12	0.06	0.07	0.06	0.07
GESVM-kNN	0.03	0.04	0.03	0.00	0.00	0.00	0.00	0.07	0.03	0.06	0.07	0.07

S6. Sensitivity analysis

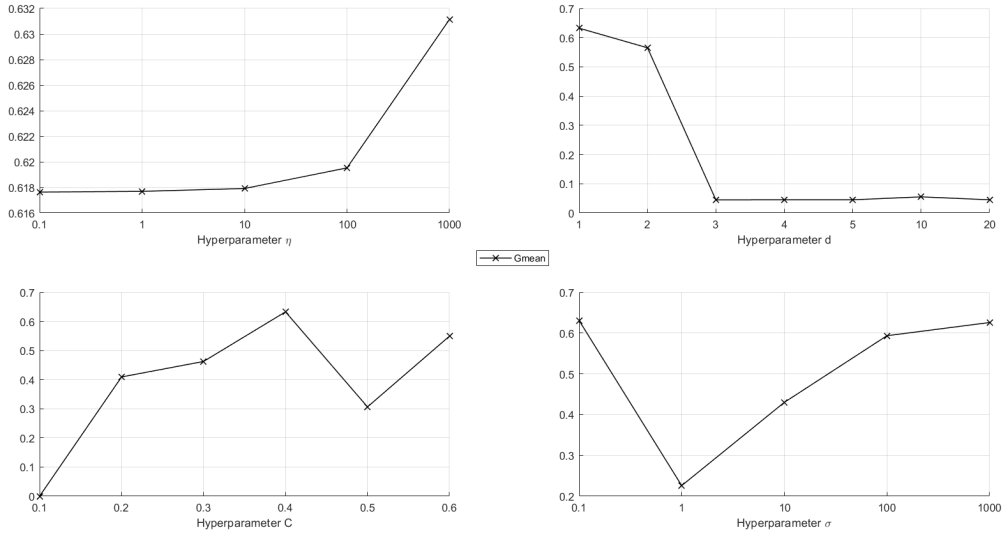

Figure 1: Hyperparameters sensitivity analysis for GESSVDD-Sb-GR-min on MNIST dataset (target class=0)

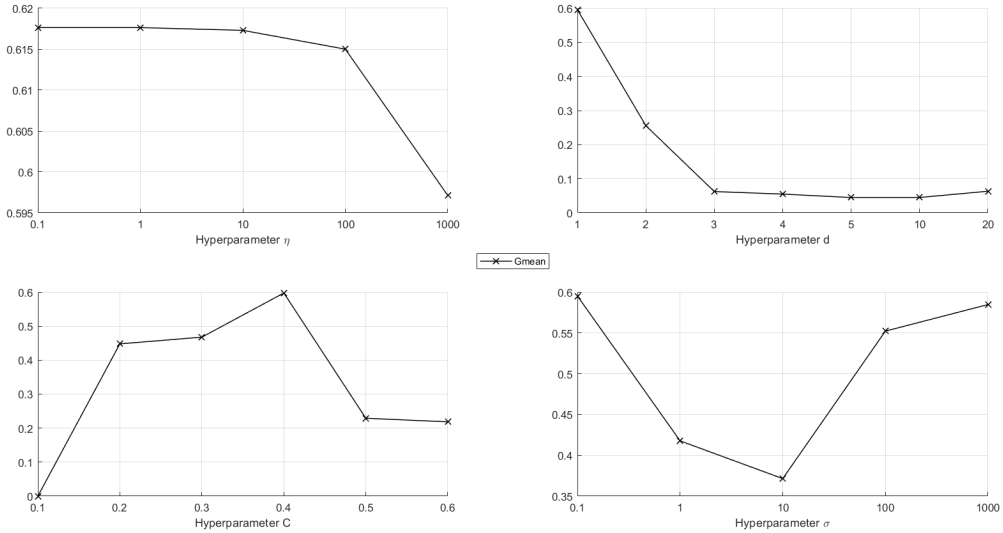

Figure 2: Hyperparameters sensitivity analysis for GESSVDD-Sb-GR-max on MNIST dataset (target class=0)

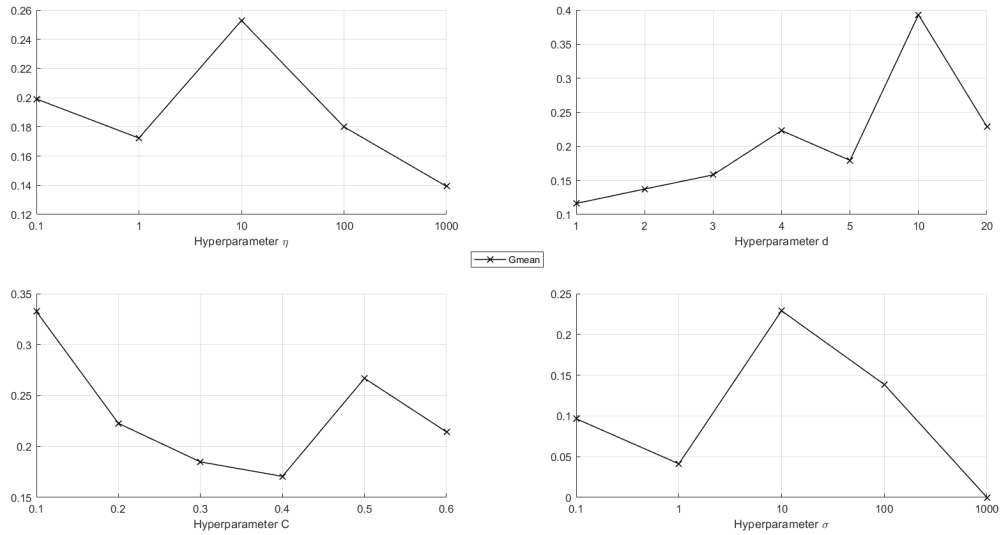

Figure 3: Hyperparameters sensitivity analysis for GESSVDD-Sb-$\mathcal{S}$-min on MNIST dataset (target class=0)

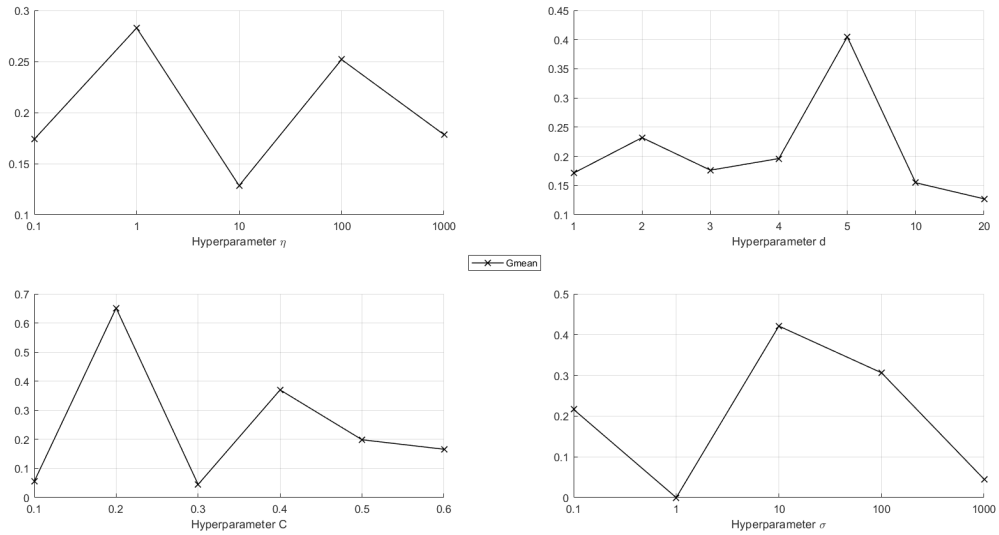

Figure 4: Hyperparameters sensitivity analysis for GESSVDD-Sb-$\mathcal{S}$-max on MNIST dataset (target class=0)

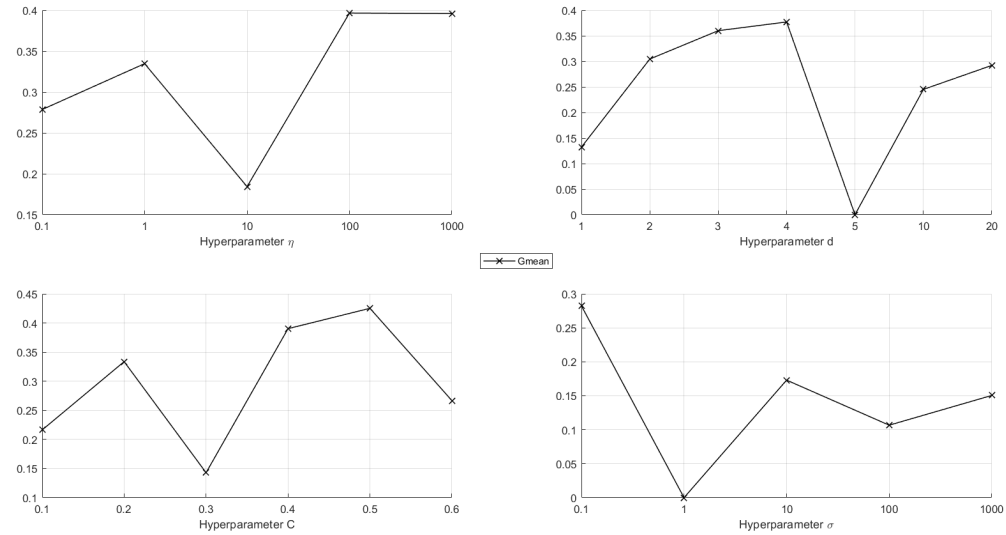

Figure 5: Hyperparameters sensitivity analysis for GESSVDD-Sb-SR-min on MNIST dataset (target class=0)

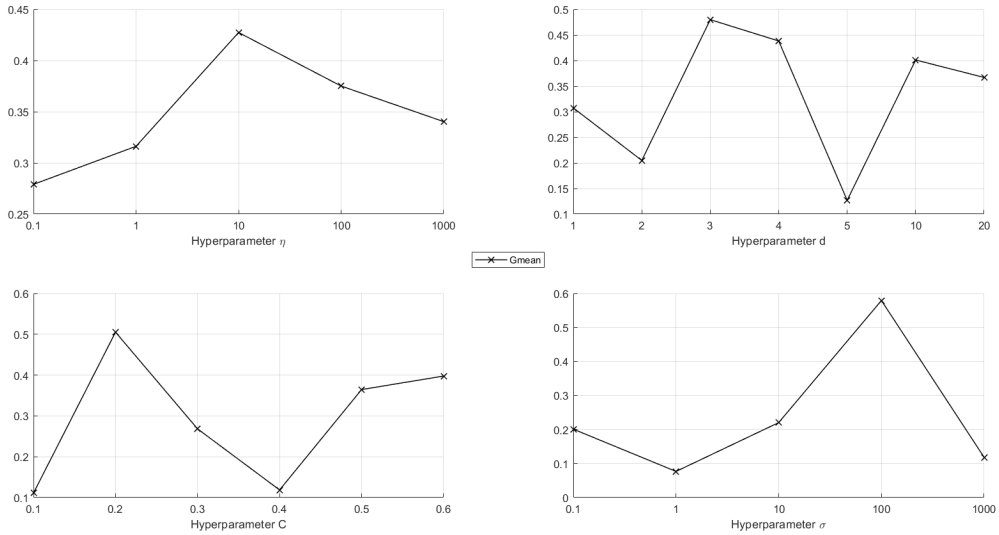

Figure 6: Hyperparameters sensitivity analysis for GESSVDD-Sb-SR-max on MNIST dataset (target class=0)

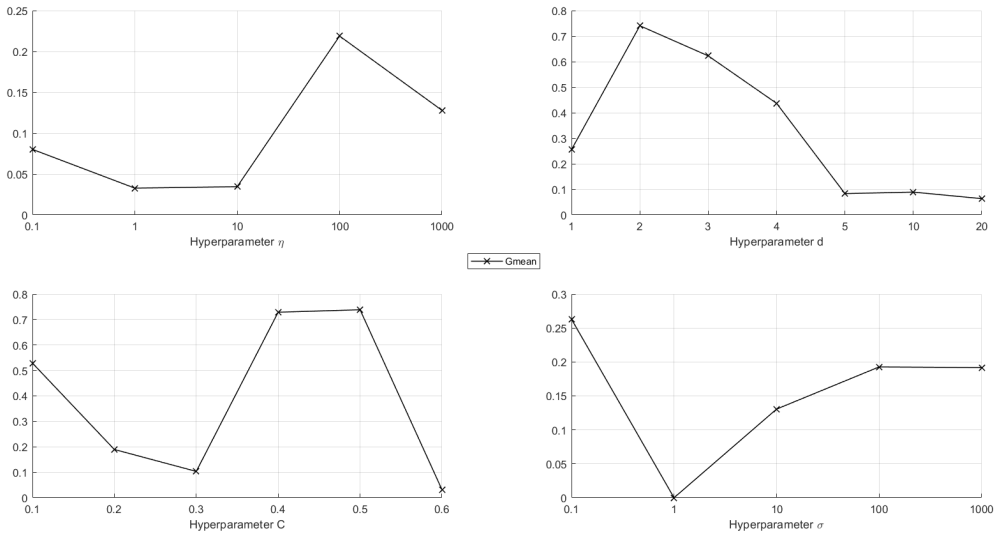

Figure 7: Hyperparameters sensitivity analysis for GESSVDD-Sw-GR-min on MNIST dataset (target class=0)

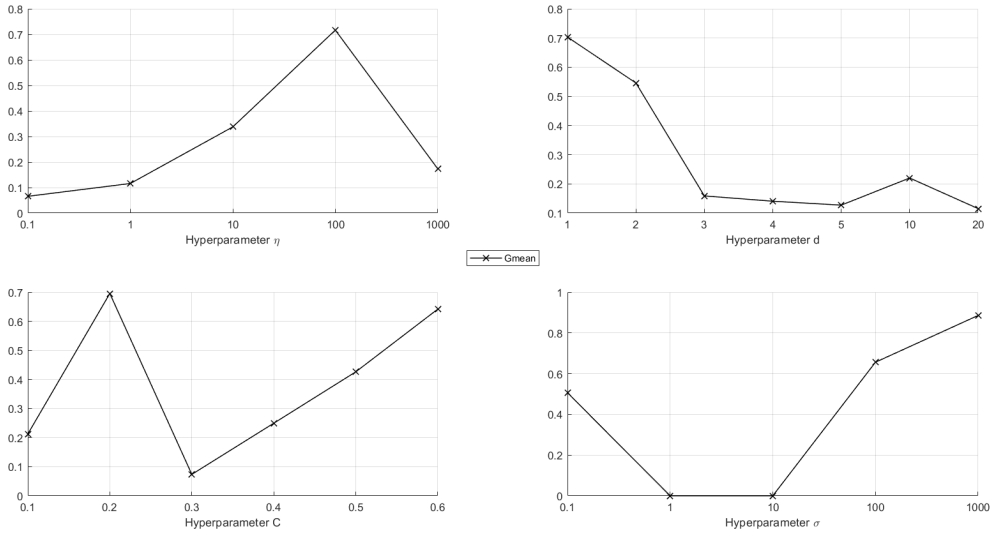

Figure 8: Hyperparameters sensitivity analysis for GESSVDD-Sw-GR-max on MNIST dataset (target class=0)

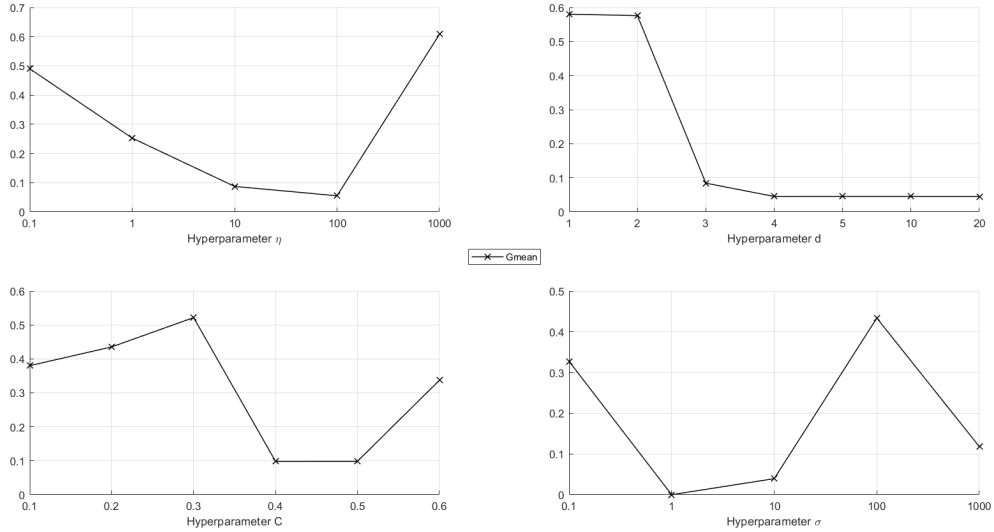

Figure 9: Hyperparameters sensitivity analysis for GESSVDD-Sw-$\mathcal{S}$-min on MNIST dataset (target class=0)

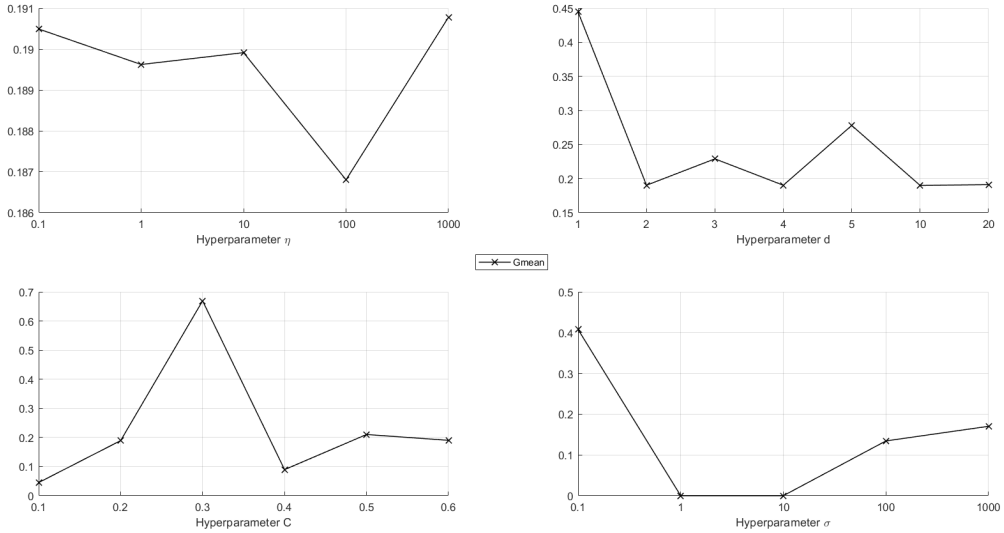

Figure 10: Hyperparameters sensitivity analysis for GESSVDD-Sw-$\mathcal{S}$-max on MNIST dataset (target class=0)

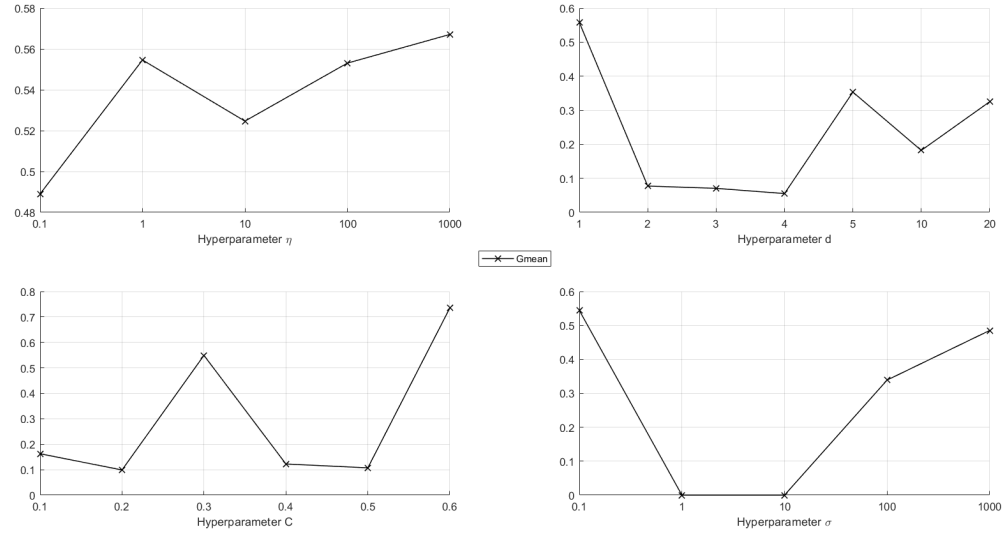

Figure 11: Hyperparameters sensitivity analysis for GESSVDD-Sw-SR-min on MNIST dataset (target class=0)

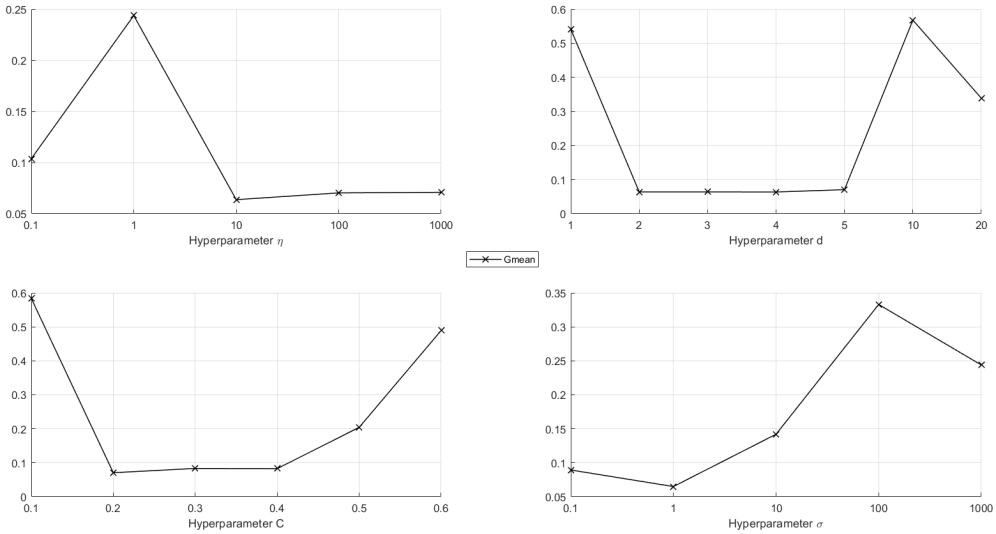

Figure 12: Hyperparameters sensitivity analysis for GESSVDD-Sw-SR-max on MNIST dataset (target class=0)

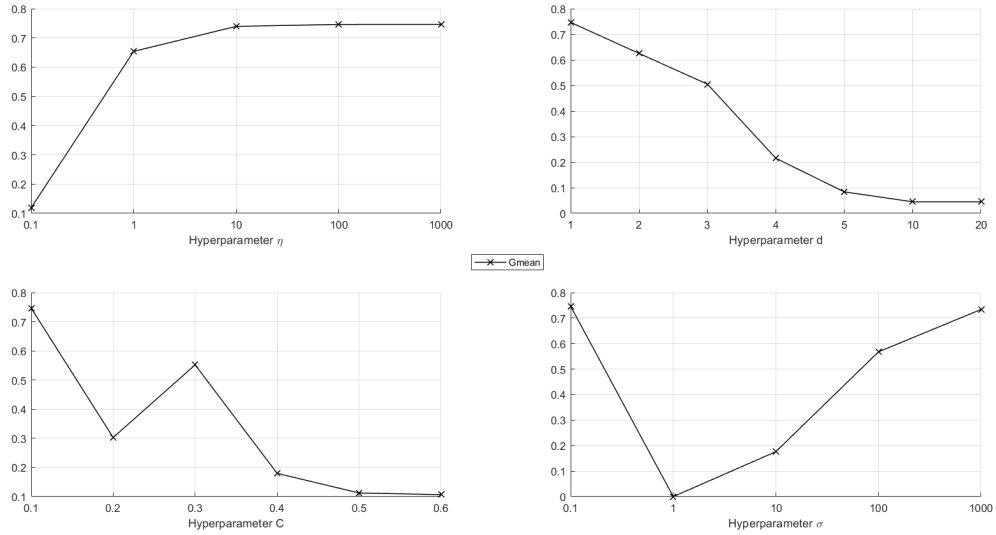

Figure 13: Hyperparameters sensitivity analysis for GESSVDD-PCA-GR-min

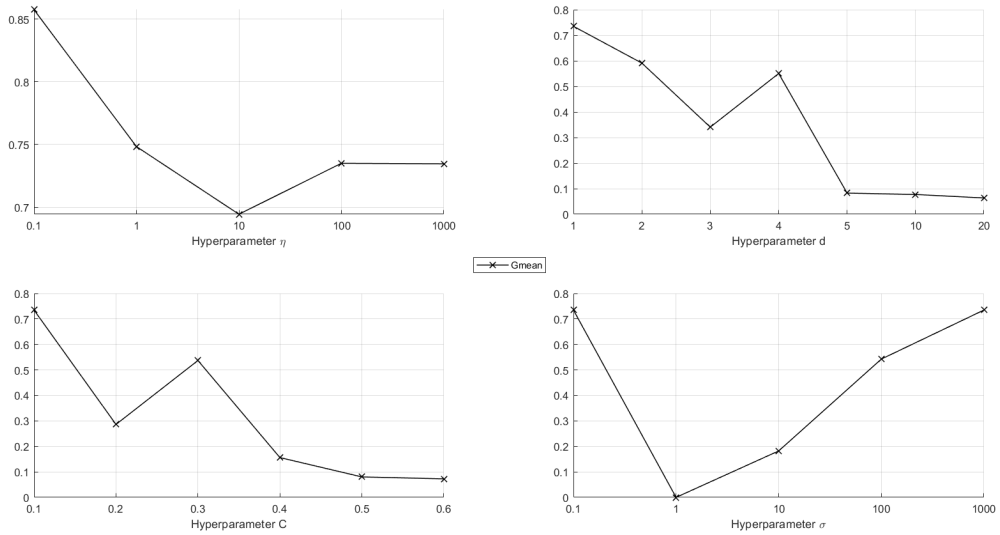

Figure 14: Hyperparameters sensitivity analysis for GESSVDD-PCA-GR-max on MNIST dataset (target class=0)

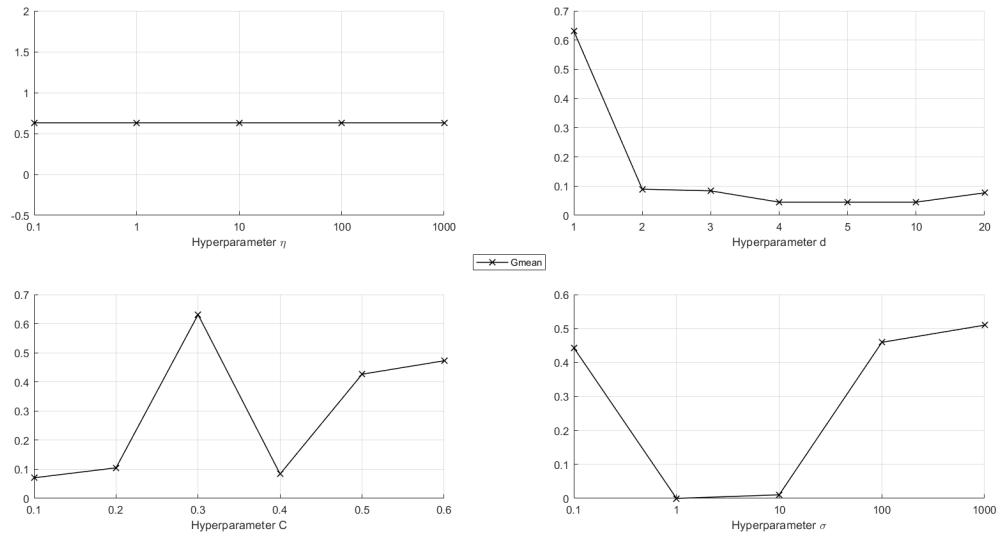

Figure 15: Hyperparameters sensitivity analysis for GESSVDD-PCA-$\mathcal{S}$-min on MNIST dataset (target class=0)

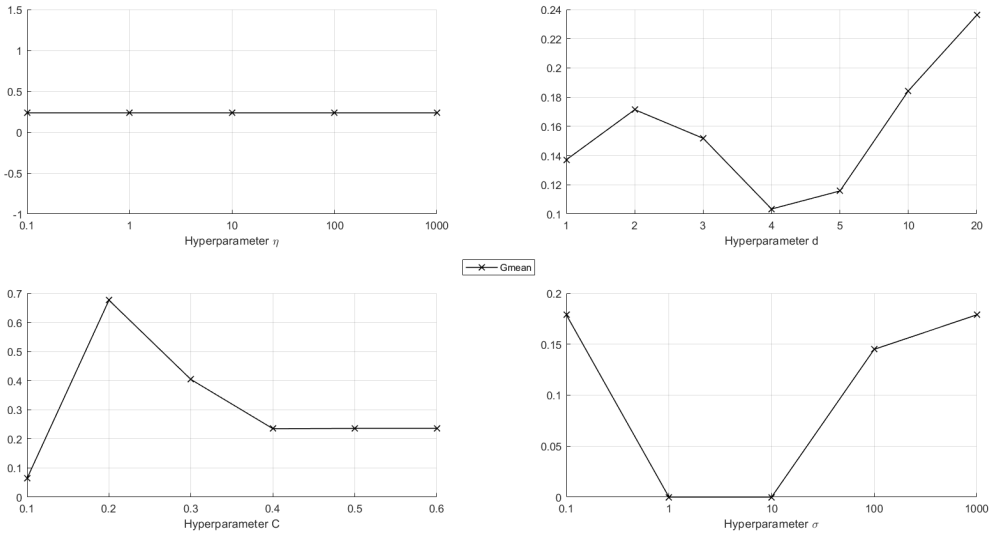

Figure 16: Hyperparameters sensitivity analysis for GESSVDD-PCA-$\mathcal{S}$-max on MNIST dataset (target class=0)

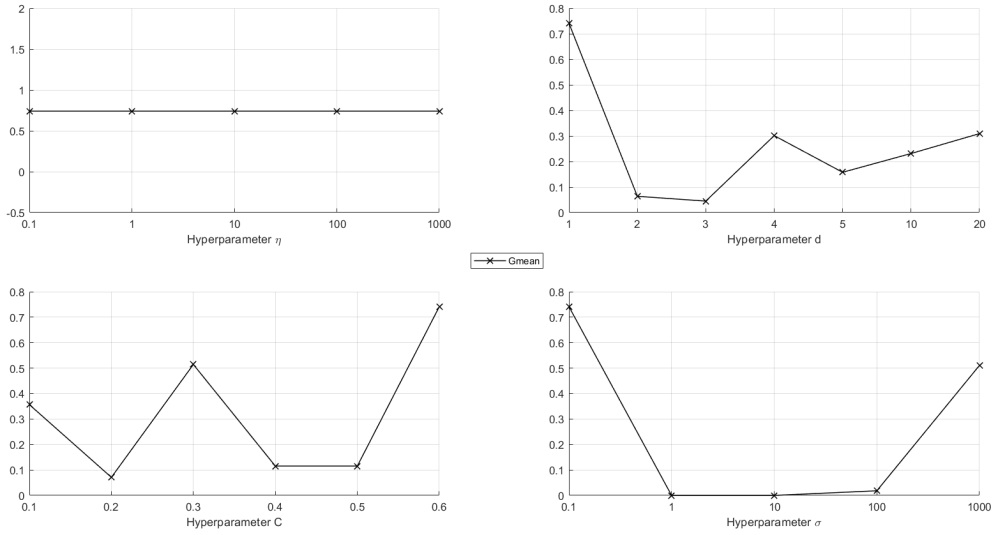

Figure 17: Hyperparameters sensitivity analysis for GESSVDD-SR-PCA-min on MNIST dataset (target class=0)

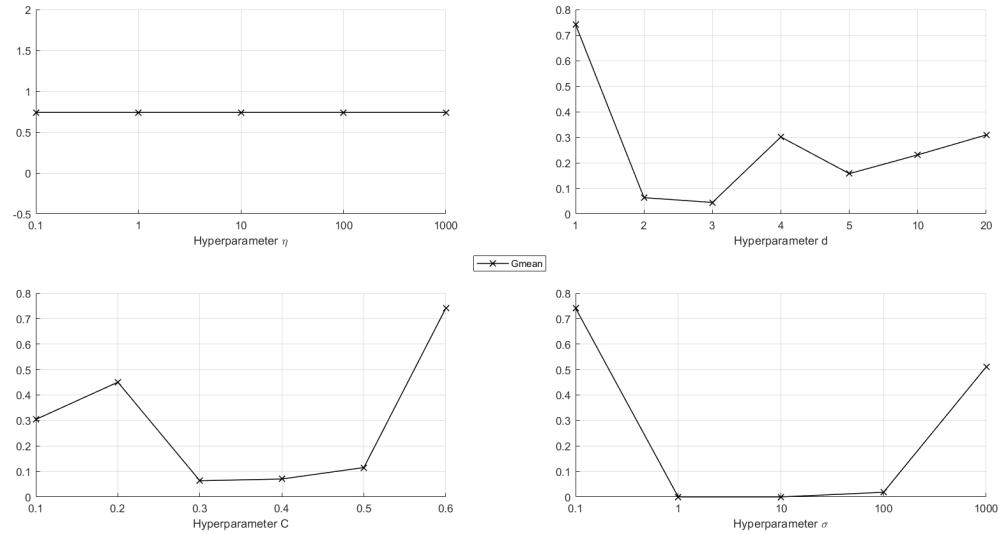

Figure 18: Hyperparameters sensitivity analysis for GESSVDD-SR-PCA-max on MNIST dataset (target class=0)

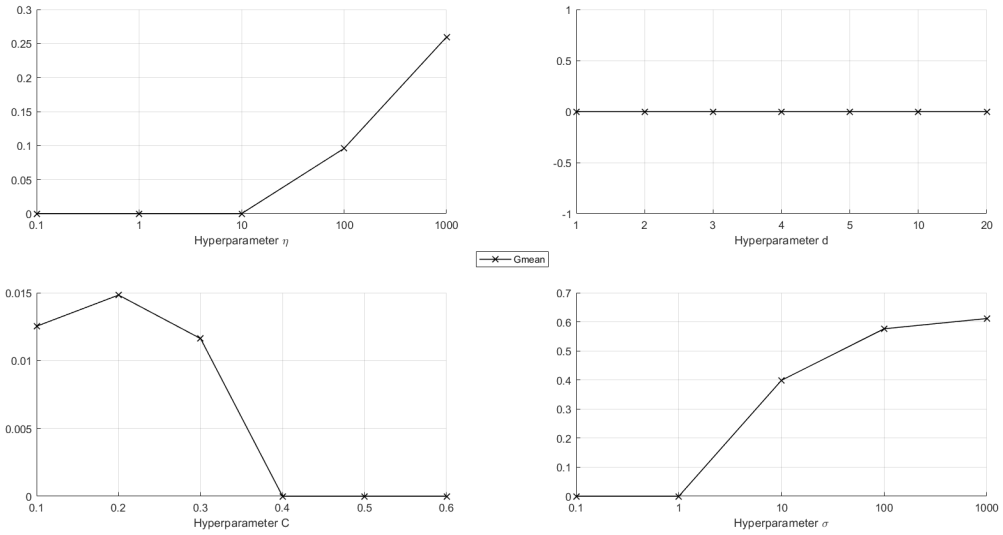

Figure 19: Hyperparameters sensitivity analysis for GESSVDD-GR-kNN-min on MNIST dataset (target class=0)

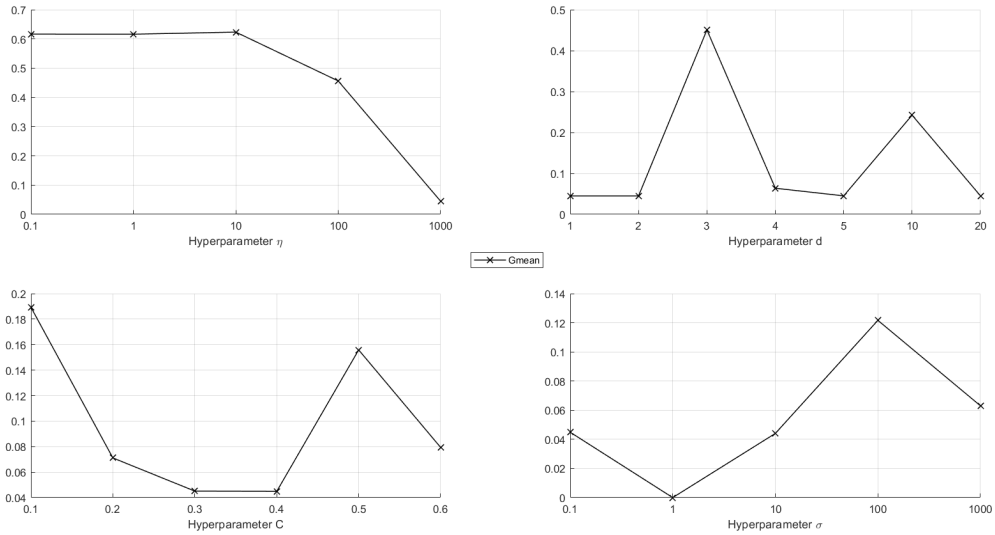

Figure 20: Hyperparameters sensitivity analysis for GESSVDD-GR-max on MNIST dataset (target class=0)

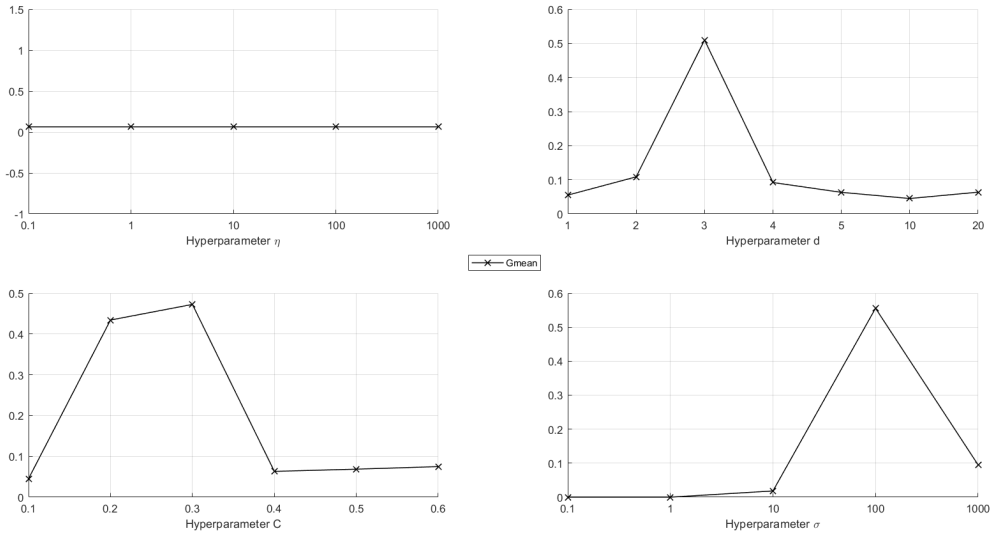

Figure 21: Hyperparameters sensitivity analysis for GESSVDD-$\mathcal{S}$-kNN-min on MNIST dataset (target class=0)

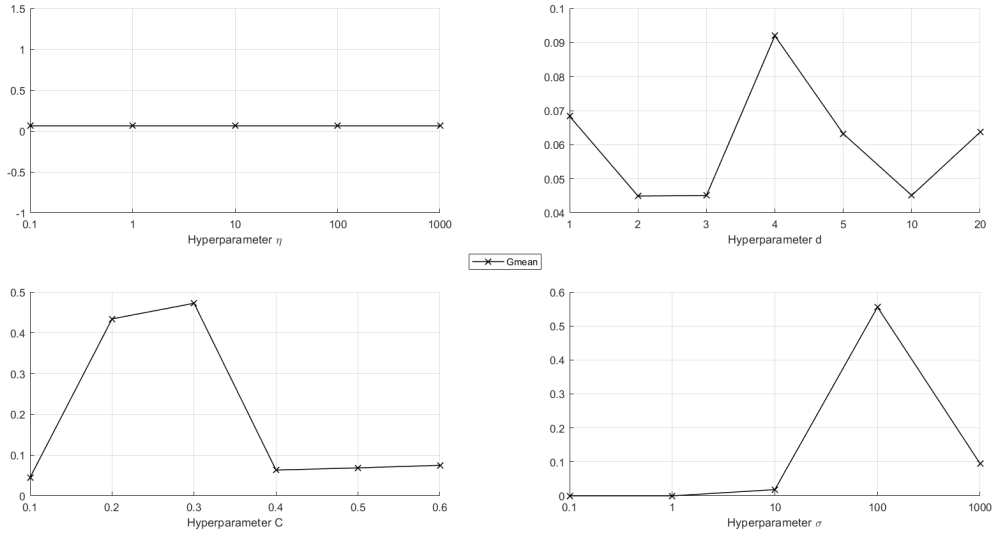

Figure 22: Hyperparameters sensitivity analysis for GESSVDD-$\mathcal{S}$-kNN-max on MNIST dataset (target class=0)

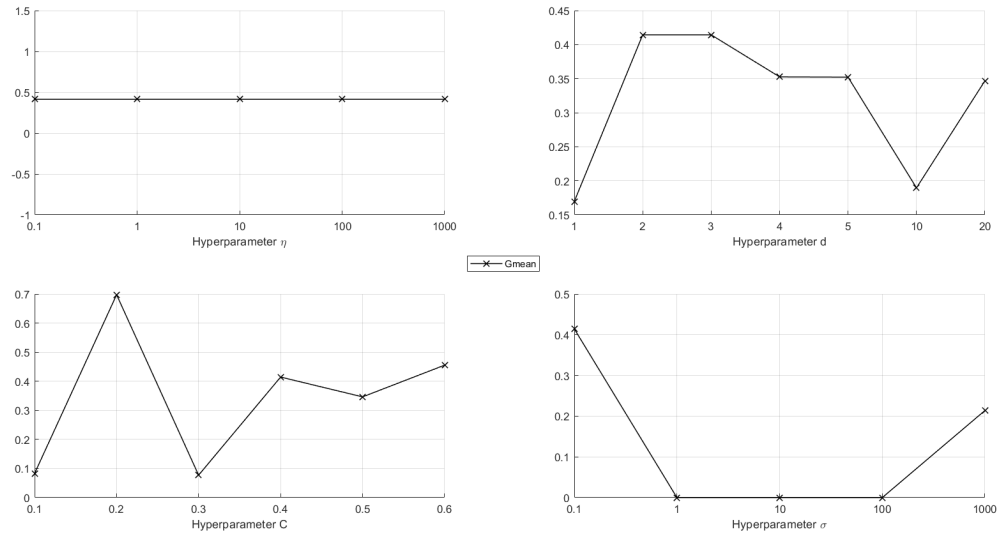

Figure 23: Hyperparameters sensitivity analysis for GESSVDD-SR-kNN-min on MNIST dataset (target class=0)

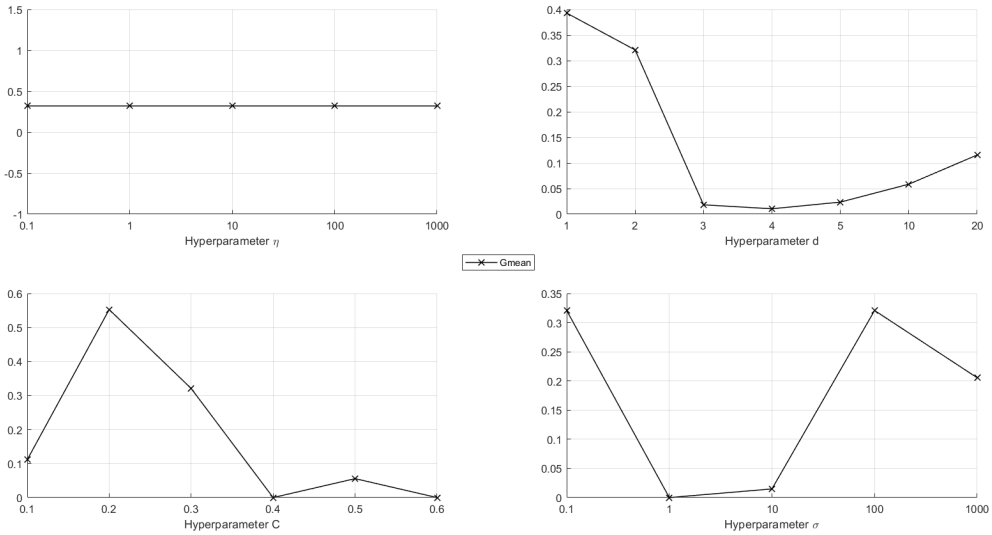

Figure 24: Hyperparameters sensitivity analysis for GESSVDD-SR-kNN-max on MNIST dataset (target class=0)

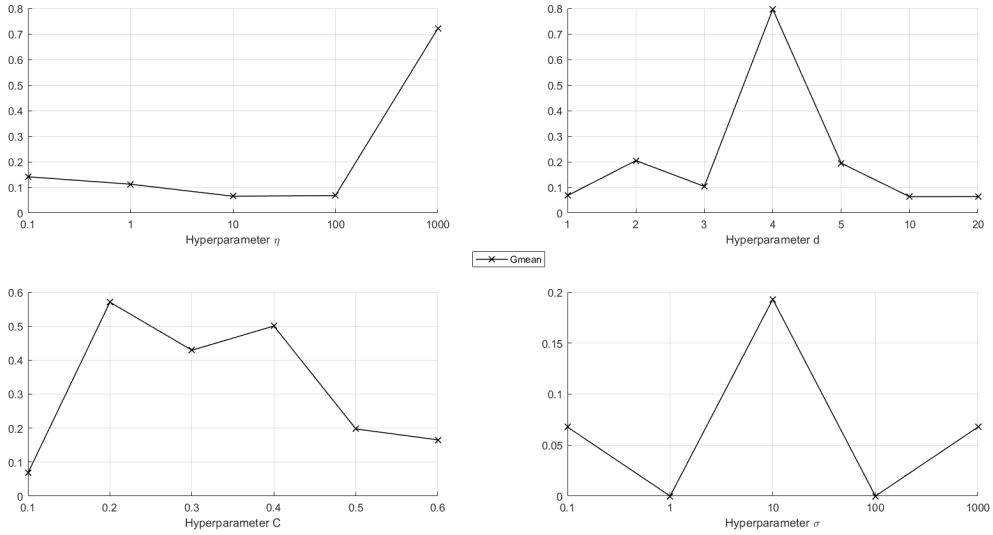

Figure 25: Hyperparameters sensitivity analysis for GESSVDD-GR-I-min on MNIST dataset (target class=0)

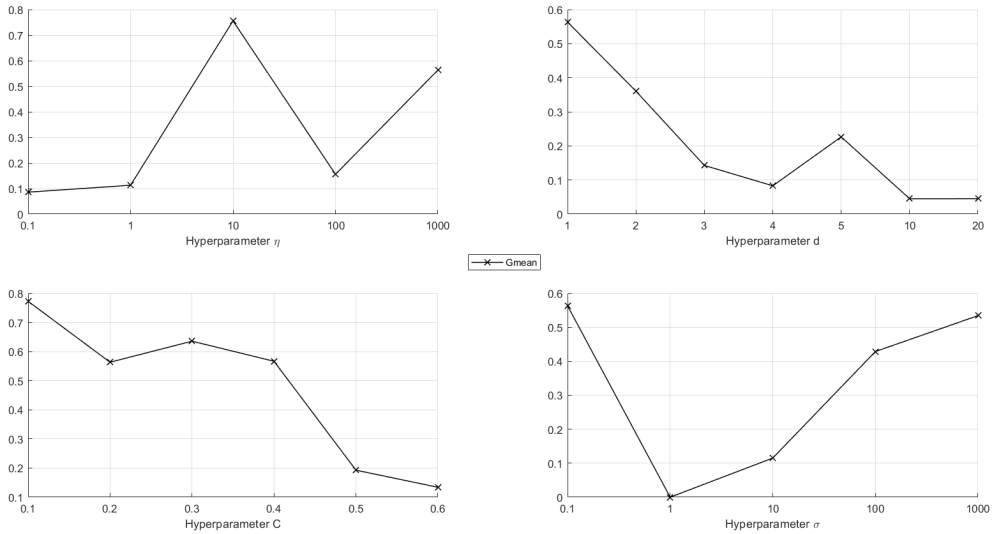

Figure 26: Hyperparameters sensitivity analysis for GESSVDD-GR-I-max on MNIST dataset (target class=0)

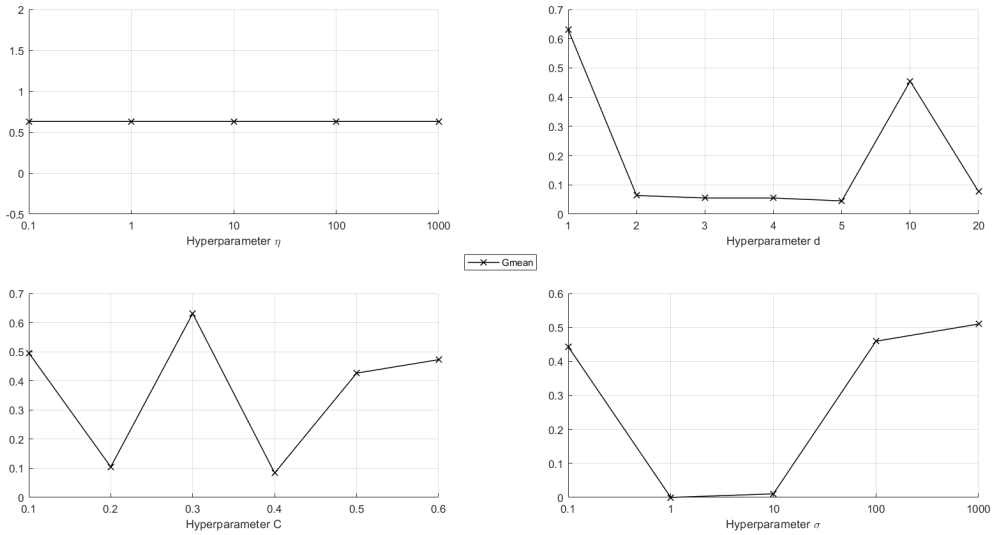

Figure 27: Hyperparameters sensitivity analysis for GESSVDD-$\mathcal{S}$-I-min on MNIST dataset (target class=0)

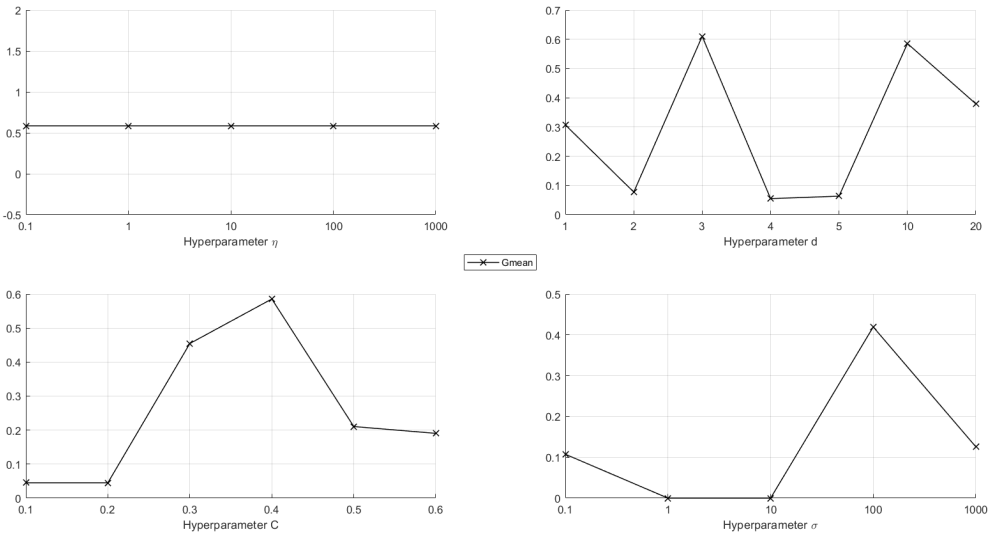

Figure 28: Hyperparameters sensitivity analysis for GESSVDD-$\mathcal{S}$-I-max on MNIST dataset (target class=0)

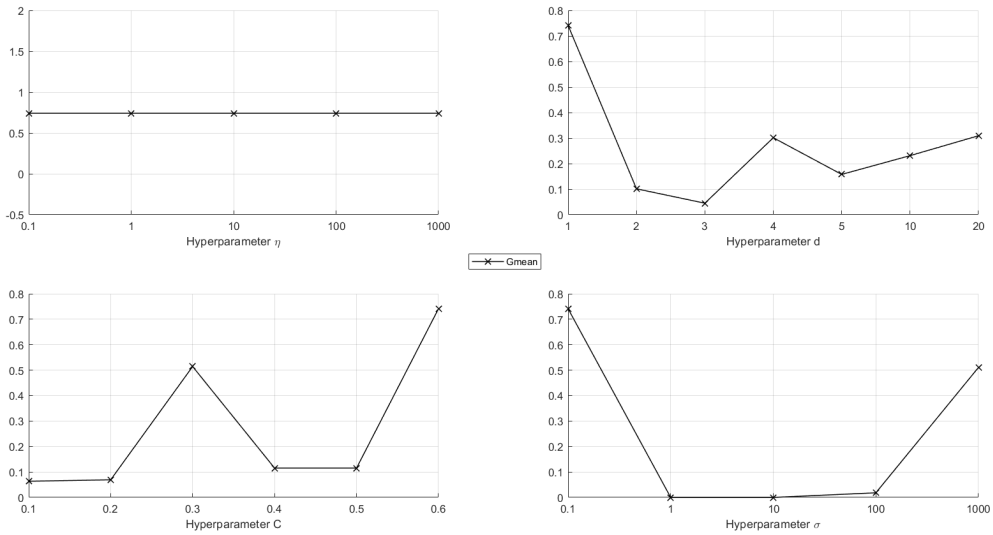

Figure 29: Hyperparameters sensitivity analysis for GESSVDD-SR-I-min on MNIST dataset (target class=0)

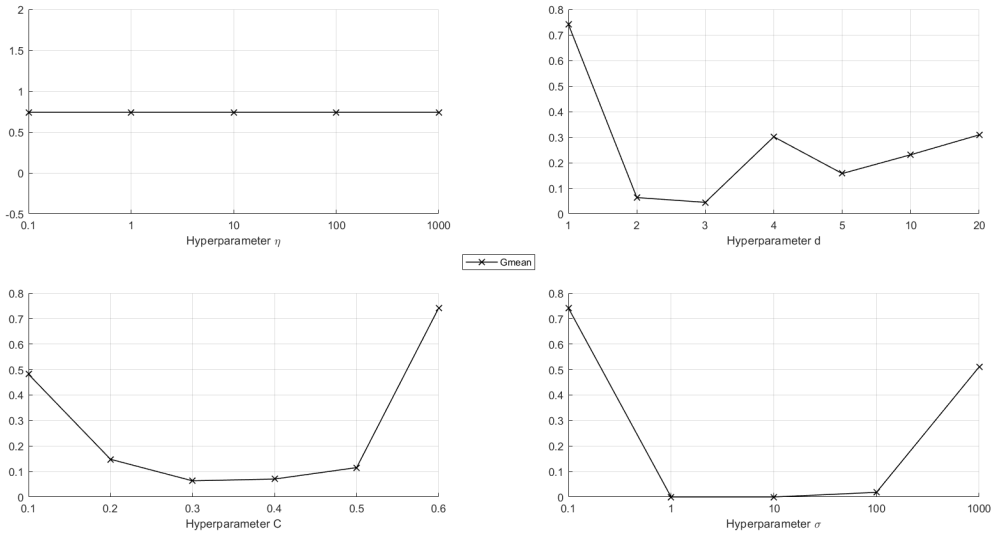

Figure 30: Hyperparameters sensitivity analysis for GESSVDD-SR-I-max on MNIST dataset (target class=0)

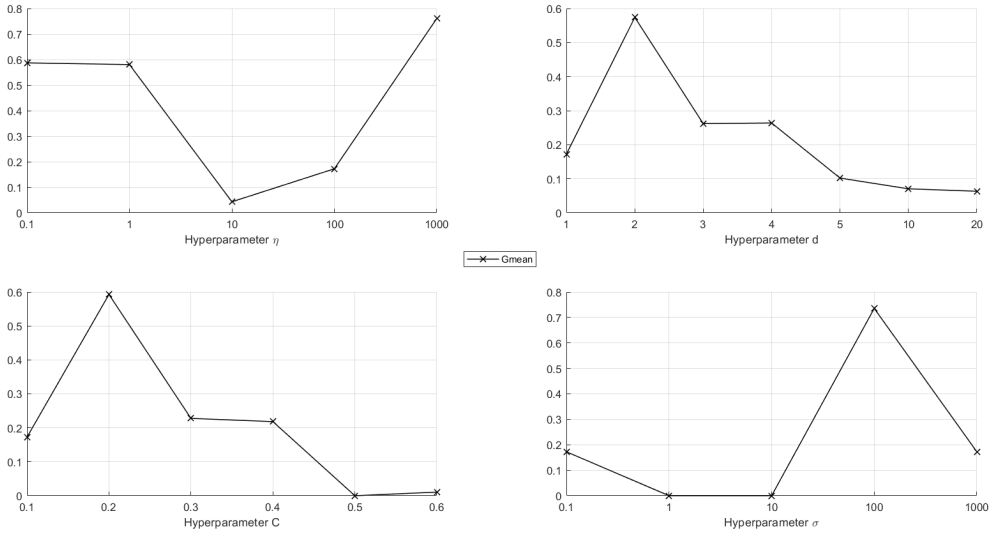

Figure 31: Hyperparameters sensitivity analysis for GESSVDD-GR-0-min on MNIST dataset (target class=0)

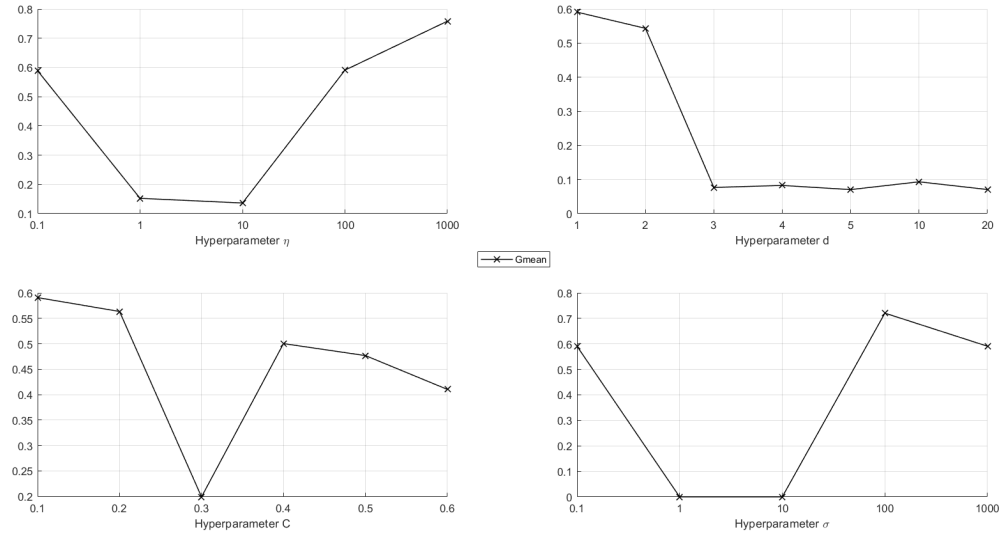

Figure 32: Hyperparameters sensitivity analysis for GESSVDD-GR-0-max on MNIST dataset (target class=0)

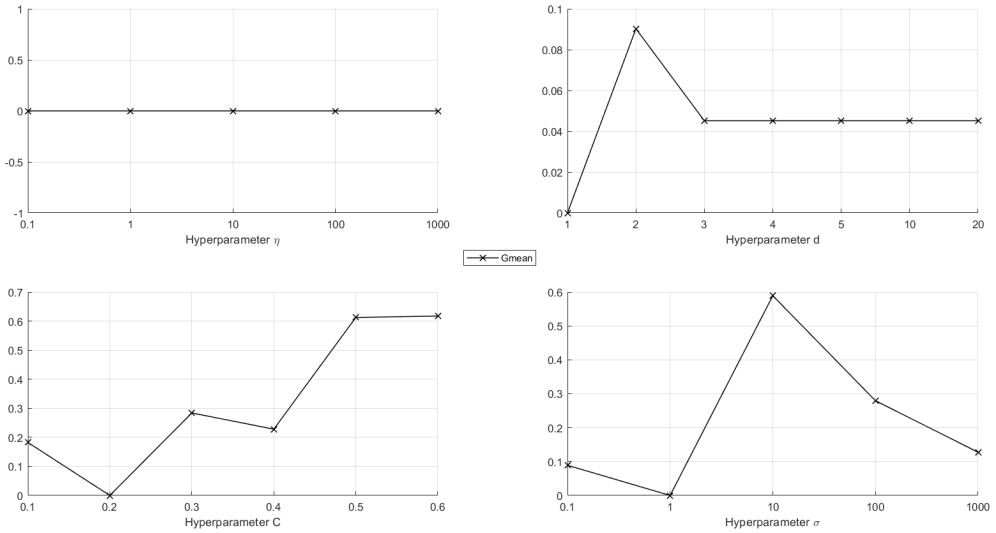

Figure 33: Hyperparameters sensitivity analysis for GESSVDD-$\mathcal{S}$-0-min on MNIST dataset (target class=0)

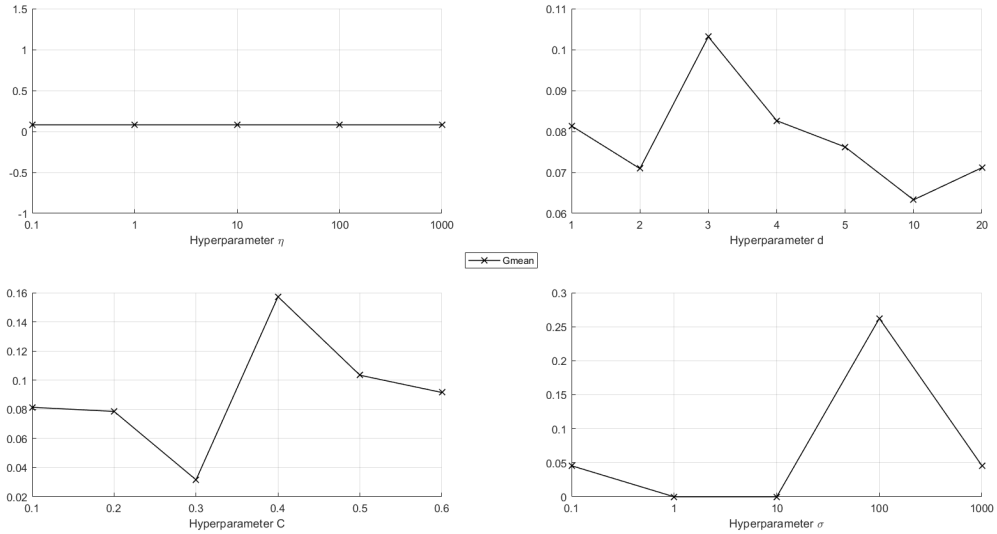

Figure 34: Hyperparameters sensitivity analysis for GESSVDD-$\mathcal{S}$-0-max on MNIST dataset (target class=0)

PUBLICATION V

Boosting Rare Benthic Macroinvertebrates Taxa Identification With One-Class Classification

F. Sohrab and J. Raitoharju

IEEE Symposium Series on Computational Intelligence 2020, 928–933
DOI:10.1109/SSCI47803.2020.9308359

Boosting Rare Benthic Macroinvertebrates Taxa Identification With One-Class Classification

Fahad Sohrab
Faculty of Information Technology and Communication Sciences
Tampere University
Tampere, Finland
fahad.sohrab@tuni.fi

Jenni Raitoharju
Programme for Environmental Information
Finnish Environment Institute
Jyväskylä, Finland
jenni.raitoharju@environment.fi

***Abstract*—Insect monitoring is crucial for understanding the consequences of rapid ecological changes, but taxa identification currently requires tedious manual expert work and cannot be scaled-up efficiently. Deep convolutional neural networks (CNNs) provide a promising way to increase the biomonitoring volumes significantly. However, taxa abundances are typically very imbalanced, and the amounts of training images for the rarest classes are too low for deep CNNs. As a result, the samples from the rare classes are often completely missed, while detecting them has biological importance. On the other hand, one-class classifiers are traditionally trained with much fewer samples to model a single class of interest. In this paper, we examine their capability to complement deep CNN based taxa identification by indicating samples potentially belonging to the rare classes of interest for human inspection. Our experiments confirm that the proposed approach may indeed support moving towards partial automation of the taxa identification task.**

***Index Terms*—biomonitoring, taxa identification, machine learning, one-class classification, support vector data description**

I. Introduction

To understand the consequences of climate change and other anthropogenic changes in different aquatic ecosystems, it is crucial to widely monitor different animal groups. Also international environmental legislation, such as the EU Water Framework Directive (WFD) [1], acknowledges the task of monitoring aquatic ecosystems. Since changes in the abundances of benthic macroinvertebrate taxa can provide an early warning sign of environmental problems in aquatic ecosystems, they are widely used as indicating factors in WFD-compliant ecological status assessment and environmental decision making [2], [3]. Simultaneously, they have also been identified as one of the most challenging groups to be monitored [4]. The task currently requires tedious manual expert work making it expensive, time-consuming, and error-prone. The recent advances in machine learning, especially deep convolutional neural networks (CNNs), provide a promising way to scale-up monitoring and provide faster information for environmental decision making. In the future, the samples of benthic macroinvertebrates may be imaged with an automated imaging device and then identified using a deep learning model trained with a sample dataset.

The overall accuracy obtained by automatic identification of benthic macroinvertebrates is approaching human expert-level [5] and, already in the near future, it may be possible to use machines to handle the majority of the samples, while human experts manually identify only the difficult and interesting cases, such as specimens potentially belonging to rare species. A significant challenge that needs to be addressed is induced by the very imbalanced taxa abundances. For some rare species, the number of training images is simply too low for a deep CNN and, as a result, the identification often fails. This problem is largely overlooked in the recent works [5], [6] that consider only the overall identification accuracy. The low number of misclassified specimens from rare species hardly affects the overall accuracy, while they are important for monitoring biodiversity. In this paper, we propose a mechanism that can indicate a reasonably-sized subset of specimens as potential samples of rare species for human expert inspection. To this end, we propose complementing the deep CNN based identification with one-class classifiers. One-class classifiers are traditionally trained with much fewer samples than deep networks, and our experimental results support the assumption that they can help detect samples from the rare species.

The rest of this paper is organized as follows. In Section II, we discuss the related work on machine learning in biomonitoring and one-class classification. The proposed system is described in detail in Section III. The details of experimental setup along with the results are provided in detail in Section IV. Finally, the conclusion and future work are presented in Section V.

II. Related work

A. Machine learning in biomonitoring

Machine learning is rapidly gaining recognition as a promising tool for many biomonitoring applications, such as identifying fish species [7], forest surveillance [8], or monitoring Arctic flowering seasons [9]. In this paper, we concentrate on benthic macroinvertebrate identification. Nevertheless, primary challenges are similar for most biomonitoring applications, and the solutions may be easily applied to other applications. For example, the identification task is very fine-grained. Fine-grained classification aims at distinguishing subordinate categories of a common superior category [10]. Domain experts

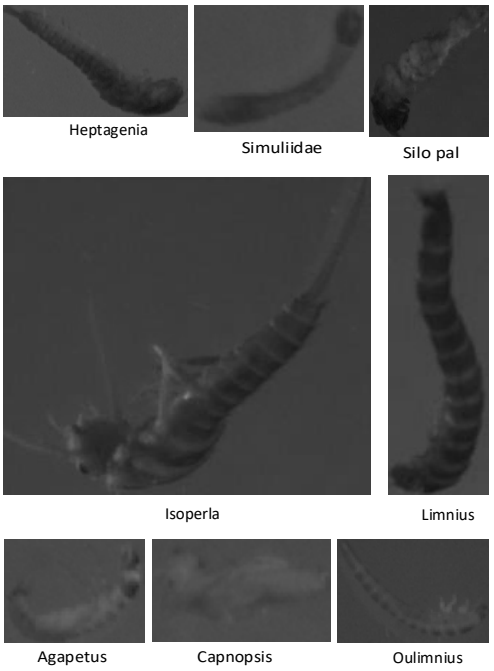

Fig. 1. Example images of different benthic macroinvertebrate taxa from FIN-Benthic 2 dataset [5]

usually define those subordinate categories with complicated rules, which typically focus on subtle differences in particular regions. For a non-expert, it may be hard to see any difference between similar species as illustrated by example images in Fig. 1. At the same time, the intra-class variance may be large due to different development stages [11]. Taxa distributions in the nature, and thus also the available reference datasets, are very imbalanced [5]. Furthermore, some taxonomists continue to object the shift toward automated methods due to different doubts and fears [12]. The last problem may be eased by providing better mechanisms for dividing the identification task between machines and human experts so that the machine first handles only the most routine-like cases [6].

The machine learning techniques for image-based taxa identification have developed from using handcrafted features with shallow networks [13], [14] towards using deep neural networks, which operate on images as inputs [5], [6]. A significant challenge with deep neural networks is the need of huge amounts of training data. This has lead to efforts to create imaging devices capable of providing high-quality images with minimal manual effort [11], [15]. Nevertheless, the existing datasets, such as FIN-Benthic2 [5] used in this paper, typically have very imbalanced classes. The smallest taxa simply do not provide enough information for training deep neural networks. However, such rare taxa and changes in their abundances may be biologically and environmentally interesting. The performance of the deep neural networks for the very rare species may be enhanced, e.g., by data-augmentation [16] or special loss functions [17], but also these approaches tend to overfit to the few training samples and do not generalize well for unseen samples. In this paper, we suggest combining one-class classifiers with the trained deep neural network to provide an additional mechanism for detecting samples potentially belonging to the rare classes for human inspection.

B. One-class classification

The main idea in one-class classification is to create a representative model of a class of interest, typically called target or positive class, using data from this class only. During inference, the model is used to predict whether unseen samples belong to the target class or are outliers. Traditional one-class classifiers can produce successful class models with only tens and hundreds of samples [18]–[20]. At the same time, computational complexity of the training phase makes training on very large or high-dimensional datasets infeasible. However, testing the trained one-class classifiers is fast, and very large datasets can be evaluated with a low computational cost. One-class classifiers have been applied, for example, on highly imbalanced phage-bacteria datasets for fast identification of potential phage candidates for a given bacteria [21]. In [19], one-class classification was applied to the facial image analysis problem. River target detection in remote sensing images was studied in [22], where the proposed one-class classification based system reduced the time complexity in target detection. A document classification system based on principal component analysis (PCA) and one-class classification was proposed in [23], where PCA helped achieve dimensionality reduction for one-class classification.

We denote the target data as $\mathrm{X} = [\mathbf{x}_1, ..., \mathbf{x}_n]$, where n is the number of target items and $\mathbf{x}_i$ are D-dimensional vectors. One-Class Support Vector Machine (OC-SVM) [24] separates all the data points from the origin and maximizes the distance from this hyperplane to the origin:

$$\begin{aligned} \min_{\mathbf{w},\xi_i,\rho} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{Cn}\sum_{i=1}^{n}\xi_i - \rho \\ \text{s.t.} \quad & \mathbf{w} * \mathbf{x}_i \geq \xi_i - \rho, \ \forall i \in \{1, \ldots, n\} \\ & \xi_i \geq 0, \ \forall i \in \{1, \ldots, n\}, \end{aligned} \tag{1}$$

where $\mathbf{w}$ is a weight vector, slack variables ξ_i allow some data points to lie within the margin, and hyper-parameter C sets an upper bound on the fraction of training samples allowed within the margin and a lower bound on the number of training samples used as Support Vector.

Another classical one-class classification method is Support Vector Data Description (SVDD) [25]. An SVDD model is trained by forming the smallest hypersphere, which includes

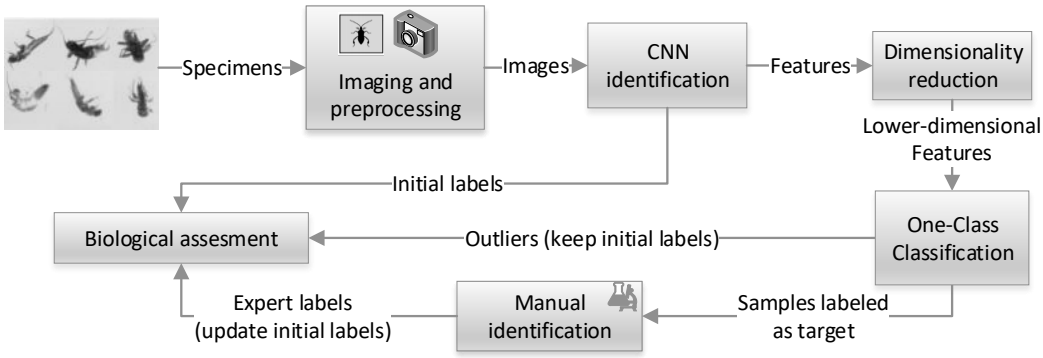

Fig. 2. The proposed taxa identification pipeline

all the target data. SVDD minimized the following function:

$$\begin{aligned} \min \quad & F(R,\mathbf{a}) = R^2 + C\sum_{i=1}^{n}\xi_i \\ \text{s.t.} \quad & \|\mathbf{x}_i - \mathbf{a}\|_2^2 \le R^2 + \xi_i, \ \forall i \in \{1,\dots,n\}, \\ & \xi_i \ge 0, \ \forall i \in \{1,\dots,n\}, \end{aligned} \tag{2}$$

where R is the radius, $\mathbf{a}$ is the center of hypersphere, ξ_i are slack variables allowing some training samples to be left outside the hypersphere, and hyper-parameter C controls the number of allowed outliers. Both OC-SVM and SVDD can be solved in one step using Lagrange multipliers.

A recent extension of SVDD, Subspace Support Vector Data Description (S-SVDD) [26] maps the data to an optimised d-dimensional subspace suitable for one-class classification as $\mathrm{Q}\mathbf{x}_i$. S-SVDD is solved iteratively, alternating the steps of solving SVDD in the current subspace and improving the subspace projection Q. The second step computes the gradient of Lagrangian of Eq. (2), $\Delta\mathcal{L}$, and updates Q as

$$\mathrm{Q} = \mathrm{Q} - \eta(\Delta\mathcal{L} + \beta\Delta\Psi), \tag{3}$$

where Ψ is an additional regularization term enforcing more variance, β is a weight for controlling the importance of Ψ, and η is a learning rate. Ψ is calculated as

$$\Psi = \mathrm{Tr}(\mathrm{Q}\mathrm{X}\lambda\lambda^{\intercal}\mathrm{X}^{\intercal}\mathrm{Q}^{\intercal}), \tag{4}$$

where $\mathrm{Tr}(\cdot)$ is a trace operator and different values for λ result in different versions of S-SVDD. In this paper, we use unregularized version (i.e. $\lambda_i = 0$) denoted as S-SVDD and two regularized versions S-SVDDr1 with $\lambda_i = 1$ and SVDDr2, where λ is used to select only the support vectors.

III. Proposed system

Our work aims at helping the move from fully manual taxa identification of benthic macroinvertebrates to a semi-automated approach, where a trained machine learning model can handle most of the specimens, while the human experts can concentrate on difficult and potentially most interesting cases. As our starting point, we assume the typical scenario where we have a trained deep neural network model that gives a satisfactory overall accuracy, while it fails to correctly identify specimens from rare species, which have biological/environmental importance. We propose using one-class classifiers as an additional mechanism that can be used together with a deep CNN to pinpoint specimens that potentially belong to the rare species for human expert inspection. The proposed approach does not require changing the existing CNN in any way and it can be used to complement any kind of CNN. Here, we only assume that the features extracted by the deep CNN provide sufficient information to discriminate the classes of interest even though the CNN itself fails to identify the samples belonging to these classes correctly.

The proposed overall framework is shown in Fig. 2. In the first phase, collected macroinvertebrate samples are imaged, and the images are preprocessed as needed (e.g., normalization, resizing). The images are fed to a trained deep CNN for initial identification. Features extracted from the second last layer of the network are mapped to a lower dimensionality to make the one-class models smaller and more focused on the key information. The mapping may be done using the traditional dimensionality reduction techniques, such as PCA or Linear Discriminant Analysis (LDA), or learned specifically for the one-class classification task, such as in S-SVDD. The lower-dimensional features are then fed to one or more one-class classifiers, each trained to model a single rare class of interest. Finally, the specimens which are classified to the target class are manually identified by a human expert, while otherwise the initial CNN identification is used in the subsequent biological assessment of the results. Note that the experts use the actual specimens with microscopic analysis, while the machine learning components rely on images and

features extracted from the images.

As one-class classifiers use only target class data for training, they may not be able to accurately distinguish unseen target samples from outliers, which have a high similarity with the target class. However, this may be even a benefit in our application. Trying to separate target samples from very similar outliers is naturally error-prone. Therefore, it is better to direct also these unclear cases for expert identification instead of trying to build a model as accurate as possible. In general, our goal is to detect as many samples from the target class as possible with the minimum amount of samples that require manual identification. However, it is not straightforward how to evaluate the performance of different one-class classifiers on the given task. Depending on biomonitoring goals and importance of the target class, it may vary how much human effort is acceptable to maximize the number of detected target class specimens.

IV. Experimental setup and results

A. Dataset

We used FIN-Benthic2 dataset [5] in our experiments. The dataset is publicly available[1] and consists of 460004 images of 9631 benthic macroinvertebrate specimens belonging to 39 different taxa. The number of images per taxon varies from 490 to 44240 making the dataset very imbalanced. The images are of varying sizes and in PNG-format. FIN-Benthic2 provides 10 different data splits for training, validation, and testing. Each split has been formed so that the images of a single specimen (max 50) are in the same set (train/validation/test). In this paper, we consider only image-based identification, and we leave it for future work to investigate how to exploit the fact that we actually have several images corresponding to the same specimen. We used Split 1 as our data splitting.

For one-class classification, we selected three different taxa, *Capnopsis schilleri*, *Nemoura cinerea*, and *Leuctra nigra*, as our target classes. Each of these taxa is rare, and VGG16 has poor performance on them. The target classes were selected as a proof-of-concept, not based on their environmental importance. The image numbers for the selected classes are shown in Table I.

TABLE I
Image numbers in Split 1 of FIN-Benthic2 dataset

	Train	Validation	Test
Capnopsis schilleri	600	100	350
Nemoura cinerea	650	100	50
Leuctra nigra	1100	50	200
Whole dataset	321407	45912	92685

B. Classifiers and their parameters

As our base-model, we fine-tuned a VGG16 network [27] pre-trained on ImageNet using FIN-Benthic2 dataset. To make VGG16 suitable for our task, we added two dense layers on top of the VGG16 convolutional output. The first added layer is composed of 4069 neurons with ReLU activation. The second added layer is the output layer composed of 39 neurons using soft-max activation. We also added two dropout layers on top of the mentioned dense layers to avoid overfitting. The dropout rate was set to 40 percent. We fine-tuned the whole network for 50 epochs using Stochastic Gradient Descent with a learning rate of 0.007 and selected the final network based on the validation set accuracy. As the original images are of varying size, we first scaled them to 64x64. The overall accuracy of the network on the test set was 0.872. This is similar to earlier published results [5], while we did not concentrate on optimizing this step in this work.

We extracted the output of the second last VGG16 layer (i.e., 4096-D) for further analysis and first applied PCA for dimensionality reduction. We used only the target class training samples to obtain the PCA mapping and then applied this mapping for all the remaining data. We kept the first 100 principal components as our final feature vectors used for training and testing the one-class classifiers. Finally, we trained different one-class classifiers (separate models for each target species) using feature vectors of the training images of the target species. The hyper-parameters were optimized using the validation set. In the end, we tested the models with the full test set, where all the images not belonging to the target class were considered as outliers.

Note that both the CNN architecture (VGG16) and the dimensionality reduction techniques (PCA) are well-known and widely used techniques and they were selected for the experiments just as a proof-of-concept. The proposed approach can be used with any CNN model and more advanced approaches for learning the optimal dimensionality reduction will be a topic for future work.

The one-class classifiers considered were OC-SVM, SVDD, S-SVDD, S-SVDDr1, and S-SVDDr2 (See Section II-B). We used both linear and non-linear (kernel) versions. For the kernel version, we used the RBF kernel, i.e.,

$$\mathbf{K}_{ij} = \exp\left(\frac{-||\mathbf{x}_i - \mathbf{x}_j||^2}{2\sigma^2}\right), \tag{5}$$

where σ is an additional hyper-parameter. The hyper-parameters C, d, η, β and σ were selected from the following values:

- $C \in \{0.01, 0.05, 0.1, 0.2, 0.3\}$,
- $d \in \{1, 2, 3, 4, 5, 10, 20, 50, 100\}$,
- $\eta \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$,
- $\beta \in \{0.01, 0.1, 1, 10, 100\}$,
- $\sigma \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$.

For comparison purposes, we used the PCA output as the input also for S-SVDD even though it could be also used to learn the projection from the original features. For both linear and non-linear case, we also report the outcome of an ensemble of all the considered one-class classifiers (Ensemble-OCC). In Ensemble-OCC, a test instance is classified to the

[1]FIN-Benthic2 dataset is available at http://urn.fi/urn:nbn:fi:att:dc6440ad-43bd-4349-8fb9-0e0d1971a7e8

TABLE II
ONE-CLASS CLASSIFIER RESULTS FOR DIFFERENT TARGET SPECIES

	Capnopsis schilleri				*Nemoura cinerea*				*Leuctra nigra*			
	TPR	GM	TP	TP+FP	TPR	GM	TP	TP+FP	TPR	GM	TP	TP+FP
CNN classification												
VGG16	0.046	0.214	16	101	0.020	0.141	1	39	0.170	0.412	34	174
Linear one-class classification												
OC-SVM	0.906	0.613	317	54367	0.660	0.357	33	74739	0.625	0.437	125	64304
SVDD	0.346	0.586	121	701	0.280	0.525	14	1422	0.730	0.832	146	4860
S-SVDD	0.557	0.740	195	1893	0.480	0.676	24	4385	0.805	0.838	161	11910
S-SVDDr1	0.609	0.773	213	1977	0.340	0.567	17	5209	0.805	0.837	161	12103
S-SVDDr2	0.706	**0.825**	247	3573	0.560	**0.702**	28	11178	0.855	**0.876**	171	9625
Ensemble-OCC	0.620	0.779	217	2077	0.380	0.601	19	4487	0.820	0.848	164	11497
Non-linear one-class classification												
OC-SVM	0.034	0.185	12	87	0.000	0.000	0	51	0.220	0.469	44	102
SVDD	0.331	0.574	116	658	0.300	0.543	15	1441	0.730	0.832	146	4904
S-SVDD	0.503	0.705	176	1169	0.440	0.649	22	3890	0.815	0.853	163	10085
S-SVDDr1	0.540	0.730	189	1404	0.400	0.622	20	3138	0.780	0.854	156	6221
S-SVDDr2	1.000	0.000	350	92685	0.220	0.465	11	1762	0.995	0.003	199	92683
Ensemble-OCC	0.503	0.705	176	1160	0.300	0.542	15	1883	0.775	0.851	155	6245

target class if most of the one-class classifiers predict the instance as positive.

C. Performance metrics

In one-class classification, sensitivity and specificity are popular performance metrics used to evaluate the performance of the trained models. Sensitivity, also known as recall or True Positive Rate (TPR), is the fraction of correctly classified target class samples correctly:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}}, \tag{6}$$

where TP is the total number of correctly predicted positive samples, and P is the total number of positive samples in the dataset. Specificity, also called the True Negative Rate (TNR), is the proportion of outlier samples that are correctly identified:

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{N}}, \tag{7}$$

where TN is an acronym for True Negatives, i.e., the total number of correctly predicted outliers, and N represents the total number of outliers in the dataset. Geometric Mean (GM) takes into account both TPR and TNR, and it is defined as the square root of the product of TPR and TNR:

$$\mathrm{GM} = \sqrt{\mathrm{TPR} \times \mathrm{TNR}}. \tag{8}$$

GM reflects both the ability of the model to detect target class samples and its ability to keep the overall amount of samples to be manually identified low. Therefore, we use it as our primary performance measure and use it for optimizing the hyper-parameters also. Furthermore, we report the total number of correctly identified target samples, i.e., True Positives (TP), and the total number of samples needing manual identification, i.e., True Positives and False Positives (TP+FP).

D. Experimental results

We give the experimental results in Table II. We see that one-class classifiers, using the same features as VGG16, can indeed detect samples from rare species much better than the deep network with a reasonable overhead (TP+FP). Here, it should be remembered that up to 50 images can represent the same specimen and, therefore, the actual number of specimens needing manual inspection may be significantly smaller than the reported number of images.

Comparing the different one-class classifiers, we see that the results for the linear versions are more robust, while the kernel versions in some cases produce models that classify (almost) all the samples as target class or outliers. The best one-class classifier for all the considered taxa in terms of GM is the linear S-SVDDr2 model. We also notice that ensembling the one-class classifiers did not further improve the results.

V. CONCLUSION AND FUTURE WORK

We proposed a taxa identification framework, where specimens potentially representing rare species are directed for human expert inspection. We showed that one-class classifiers can complement a deep neural network with high overall classification accuracy in a way that allows dividing the tasks between machine and human experts. This supports moving from fully manual to semi-automated taxa identification in biomonitoring. The best one-class classification model in terms of Geometric Mean was regularized linear Subspace Support Vector Data Description.

In this paper, we considered images separately, while we actually have multiple images of a single specimen. In our future work, we will consider how to exploit this information. For example, we may require a certain fraction of images to be

classified as target class to assign the specimen for human inspection, or we may use multi-modal one-class classifiers, e.g., [28], by considering each image as a separate modality. We will experiment with more advanced dimensionality reduction techniques and consider how to use classification confidences of both the CNN and one-class classifiers to further reduce the number of samples requiring human inspection. We will also experiment with different classifier types, such as class-specific classifiers, in our general identification framework.

ACKNOWLEDGMENT

Thanks to the Academy of Finland project 324475 for funding the work of Jenni Raitoharju.

REFERENCES

[1] EU Water Framework Directive (WDF), Directive 2000/60/EC, *Journal of the European Communities*, vol. L327/1, 2000.

[2] D. Buchner, A. Beermann, A. Laini, P. Rolauffs, S. Vitecek, D. Hering, and F. Leese, "Analysis of 13,312 benthic invertebrate samples from german streams reveals minor deviations in ecological status class between abundance and presence/absence data," *PloS one*, vol. 14, no. 12, 2019.

[3] Y. Sun, Y. Takemon, and Y. Yamashiki, "Freshwater spring indicator taxa of benthic invertebrates," *Ecohydrology & Hydrobiology*, 2019.

[4] S. Poikane, R. Johnson, L. Sandin, A. Schartau *et al.*, "Benthic macroinvertebrates in lake ecological assessment: a review of methods, intercalibration and practical recommendations," *Science of the total environment*, vol. 543, pp. 123–134, 2016.

[5] J. Ärje, J. Raitoharju, A. Iosifidis, V. Tirronen, K. Meissner, M. Gabbouj, S. Kiranyaz, and S. Kärkkäinen, "Human experts vs. machines in taxa recognition," *Signal Processing: Image Communication*, p. 115917, 2020.

[6] J. Raitoharju and K. Meissner, "On confidences and their use in (semi-) automatic multi-image taxa identification," in *IEEE Symposium Series on Computational Intelligence*, 2019.

[7] A. A. dos Santos and W. N. Gonçalves, "Improving pantanal fish species recognition through taxonomic ranks in convolutional neural networks," *Ecological Informatics*, vol. 53, 2019.

[8] N. Wyniawskyj, M. Napiorkowska, D. Petit, P. Podder, and P. Marti, "Forest monitoring in guatemala using satellite imagery and deep learning," in *IEEE International Geoscience and Remote Sensing Symposium*, 2019.

[9] J. Ärje, D. Milioris, D. Tran, A. Iosifidis, J. Raitoharju, M. Gabbouj, J. Jepsen, and T. H. ye, "Automatic flower detection and classification system using a light-weight convolutional neural network," in *EUSIPCO workshop on "Signal Processing, Computer Vision and Deep Learning for Autonomous Systems"*, 2019.

[10] Z. Yang, T. Luo, D. Wang, Z. Hu, J. Gao, and L. Wang, "Learning to navigate for fine-grained classification," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 420–435.

[11] J. Raitoharju, E. Riabchenko, I. Ahmad, A. Iosifidis, M. Gabbouj, S. Kiranyaz, V. Tirronen, J. Ärje, S. Kärkkäinen, and K. Meissner, "Benchmark database for fine-grained image classification of benthic macroinvertebrates," *Image and Vision Computing*, vol. 78, 2018.

[12] M. Kelly, S. Schneider, and L. King, "Customs, habits, and traditions: the role of nonscientific factors in the development of ecological assessment methods," *Wiley Interdisciplinary Reviews: Water*, vol. 2, no. 3, 2015.

[13] D. Lytle, G. Martinez-Munoz, W. Zhang, N. Larios, L. Shapiro, R. Paasch, A. Moldenke, E. Mortensen, S. Todorovic, and T. Dietterich, "Automated processing and identification of benthic invertebrate samples," *Journal of the North American Benthological Society*, vol. 29, no. 3, 2010.

[14] S. Kiranyaz, T. Ince, J. Pulkkinen, M. Gabbouj, J. Ärje, S. Kärkkäinen, V. Tirronen, M. Juhola, T. Turpeinen, and K. Meissner, "Classification and retrieval on macroinvertabrate image databases," *Computers in Biology and Medicine*, vol. 41, no. 7, 2011.

[15] J. Ärje, C. Melvad, M. R. Jeppesen, S. A. Madsen, J. Raitoharju, M. S. Rasmussen, A. Iosifidis, V. Tirronen, M. Gabbouj, K. Meissner *et al.*, "Automatic image-based identification and biomass estimation of invertebrates," *Methods in Ecology and Evolution*, 2020.

[16] J. Raitoharju, E. Riabchenko, K. Meissner, I. Ahmad, A. Iosifidis, M. Gabbouj, and S. Kiranyaz, "Data enrichment in fine-grained classification of aquatic macroinvertebrates," in *ICPR Workshop on Computer Vision for Analysis of Underwater Imagery*, 2016.

[17] C. Huang, Y. Li, C. Change Loy, and X. Tang, "Learning deep representation for imbalanced classification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[18] S. S. Khan and M. G. Madden, "One-class classification: taxonomy of study and review of techniques," *The Knowledge Engineering Review*, vol. 29, no. 3, 2014.

[19] V. Mygdalis, A. Iosifidis, A. Tefas, and I. Pitas, "One class classification applied in facial image analysis," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016.

[20] F. Sohrab, J. Raitoharju, A. Iosifidis, and M. Gabbouj, "Ellipsoidal subspace support vector data description," *IEEE Access*, vol. 8, pp. 122 013–122 025, 2020.

[21] J. F. López, J. A. L. Sotelo, D. Leite, and C. Peña-Reyes, "Applying one-class learning algorithms to predict phage-bacteria interactions," in *2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*. IEEE, 2019.

[22] S. Bo and Y. Jing, "One-class classification based river detection in remote sensing image," in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 2017.

[23] B. S. Kumar and V. Ravi, "Text document classification with pca and one-class svm," in *Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications*. Springer, 2017.

[24] B. Schölkopf, R. Williamson, A. Smola, and J. Shawe-Taylor, "Sv estimation of a distribution's support," 1999.

[25] D. Tax and R. Duin, "Support vector data description," *Machine learning*, vol. 54, no. 1, 2004.

[26] F. Sohrab, J. Raitoharju, M. Gabbouj, and A. Iosifidis, "Subspace support vector data description," in *International Conference on Pattern Recognition*, 2018.

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015.

[28] F. Sohrab, J. Raitoharju, A. Iosifidis, and M. Gabbouj, "Multimodal subspace support vector data description," *Pattern Recognition*, p. 107648, 2020.