
WORKING PAPER SERIES

2023-EQM-01

Convex and Nonconvex Nonparametric Frontier-based Classification Methods for Anomaly Detection

Qianying JIN

College of Economics and Management, Nanjing University of Aeronautics and Astronautics,
Nanjing, 211106, China. qianying.jin@nuaa.edu.cn.

Kristiaan KERSTENS

Univ. Lille, CNRS, IESEG School of Management, UMR 9221 - LEM - Lille Economie
Management, Lille, France, k.kerstens@ieseg.fr.

Ignace VAN DE WOESTYNE

KU Leuven, Research Centre for Operations Research and Statistics (ORSTAT), Brussels
Campus, War- moesberg 26, B-1000 Brussels, Belgium,
ignace.vandewoestyne@kuleuven.be.

IESEG School of Management
Lille Catholic University
3, rue de la Digue
F-59000 Lille
Tel: 33(0)3 20 54 58 92
www.ieseg.fr

Staff Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate. Any views expressed are solely those of the author(s) and so cannot be taken to represent those of IESEG School of Management or its partner institutions.

All rights reserved. Any reproduction, publication and reprint in the form of a different publication, whether printed or produced electronically, in whole or in part, is permitted only with the explicit written authorization of the author(s).

For all questions related to author rights and copyrights, please contact directly the author(s).

Convex and Nonconvex Nonparametric Frontier-based Classification Methods for Anomaly Detection

Qianying Jin* Kristiaan Kerstens[†] Ignace Van de Woestyne[‡]

This version: 16th February 2023

Abstract

Effective methods for determining the boundary of the normal class are very useful for detecting anomalies in commercial or security applications - a problem known as anomaly detection. This contribution proposes a nonparametric frontier-based classification (NPFC) method for anomaly detection. By relaxing the commonly used convexity assumption in the literature, a nonconvex NPFC method is constructed and the nonconvex nonparametric frontier turns out to provide a more conservative boundary enveloping the normal class. By reflecting on the monotonic relation between the characteristic variables and the membership, the proposed NPFC method is in a more general form since both input-type and output-type characteristic variables are incorporated. A biomedical data set is used to test the performance of the proposed NPFC methods. The results show that the proposed NPFC methods have competitive classification performance and have consistent advantages in detecting abnormal samples, especially the nonconvex NPFC method.

Keywords: Nonparametric Frontier; Convex; Nonconvex; Anomaly Detection

*College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106, China. qianying.jin@nuaa.edu.cn, Corresponding author.

[†]Univ. Lille, CNRS, IESEG School of Management, UMR 9221 - LEM - Lille Économie Management, Lille, France, k.kerstens@ieseg.fr.

[‡]KU Leuven, Research Centre for Operations Research and Statistics (ORSTAT), Brussels Campus, Warmoesberg 26, B-1000 Brussels, Belgium, ignace.vandewoestyne@kuleuven.be.

1 Introduction

Anomaly detection can be defined as the task of detecting anomalous data that differ in some aspects from the normal data which is known during training. The practical use and challenging nature of anomaly detection have gained considerable research attention and led to many methods being proposed. The applications of anomaly detection methods settle across sectors and disciplines, such as in medical diagnosis (Park, Huang, and Ding (2010); da Silva, Schmidt, da Costa, da Rosa Righi, and Eskofier (2021)), faults and failure detection in complex systems (Clifton, Clifton, Zhang, Watkinson, Tarassenko, and Yin (2014); Jiang, Li, Xie, and Du (2020)), and credit card or telecom fraud detection (Bhattacharyya, Jha, Tharakunnel, and Westland (2011); Ahmed, Mahmood, and Islam (2016); Al-Hashedi and Magalingam (2021)) to name but a few. In all of these commercial or security applications, detecting potential anomalies is of crucial importance to prevent from some catastrophic outcome. For comprehensive and structured investigations of anomaly detection methods and their application domains, one may refer to the surveys of, e.g., Markou and Singh (2003a,b); Ruff, Kauffmann, Vandermeulen, Montavon, Samek, Kloft, Dietterich, and Müller (2021); Pang, Shen, Cao, and Hengel (2021).

In anomaly detection, the typical dataset is unbalanced, with a very large number of normal data, but insufficient abnormal data to describe the anomaly or say novelty. Thus, classical binary classification methods may not be applicable for anomaly detection since they normally require the two classes to be more or less balanced. To address the anomaly detection problem from a discriminative perspective, one-class classification, occasionally also called single-class classification, is introduced (see Moya, Koch, and Hostetler (1993), El-Yaniv and Nisenson (2006), or Khan and Madden (2014)). The one-class classification anomaly detection methods are particularly based on the single class of normal data.

The fundamental idea of the one-class classification methods aims at finding a boundary around the normal class to describe the domain containing normal data only. If a new data point is located within the boundary, then it is regarded as normal; otherwise, it is an anomaly. Among the various one-class classification methods, one-class support vector classifiers (OC-SVCs) determine the boundary using only the normal data closest to it, i.e., the support vectors, not relying on any property of the distribution of the data. Over the past decades, OC-SVCs have been extensively studied and found suitable for anomaly detection in numerous applications (e.g., Alam, Sonbhadra, Agarwal, and Nagabhushan (2020)).

Two evolutionary methods of OC-SVCs are the support vector data description (SVDD)

method and the one-class support vector machine (OC-SVM) method. First, the SVDD method proposed by Tax and Duin (1999) defines a hypersphere with minimum radius that encloses the normal class. It gives a satisfactory performance but leads to a loose boundary for multivariate datasets (see Tax and Juszczak (2003)). Second, the OC-SVM method proposed by Schölkopf, Williamson, Smola, Shawe-Taylor, and Platt (1999) constructs a hyperplane to separate the normal class with the maximal margin from the origin in some feature space. In the OC-SVM method, all anomalies are assumed to be close to the origin, while the normal data points are far from the origin. In this respect, the OC-SVM method is not purely based on the normal class. Moreover, both the SVDD and OC-SVM methods have fixed their general shape of the boundary a priori: the former defines a hypersphere and the latter describes a hyperplane. For cases where the normal class cannot be directly described by a hypersphere or a hyperplane, the introduction of kernel functions can bring some flexibility Noble (2006).

In line with the fundamental idea of OC-SVCs, this contribution is also interested in determining the boundary of the normal class using only part of the normal data. Moreover, the boundary is expected to be around the data set without pre-determining the exact shape. To meet this, the Data Envelopment Analysis (DEA) method, which floats a piecewise linear boundary enveloping the observed data set, becomes of interest. It is a linear programming model proposed by Banker, Charnes, and Cooper (1984) and widely applied in production economics and finance (see the surveys and historical developments in Emrouznejad and Yang (2018); Emrouznejad, Banker, and Neralic (2019); Kaffash, Azizi, Huang, and Zhu (2020)). The piecewise linear boundary generated from the DEA model is commonly termed as a nonparametric frontier. The shape of the nonparametric frontier is determined by the originally observed data and the imposition of some weak maintained axioms, not necessarily in any feature space.

The main inspiration of introducing the nonparametric frontier to anomaly detection stems from the earliest work of Troutt, Rai, and Zhang (1996) and a modified version by Seiford and Zhu (1998). They propose to develop an acceptance frontier with DEA method for case based computer systems. Based on their work, theoretical extensions include characterizing data with non-discretionary characteristics (Leon and Palacios (2009)), incorporating importance measures of characteristics (Yan and Wei (2011)), etc. A series of empirical and experimental studies with the acceptance boundary were contemporaneously conducted by Pendharkar and coauthors in various domains, e.g., bankruptcy prediction (Pendharkar (2002)), mining breast cancer patterns (Pendharkar, Khosrowpour, and Rodger (2000)), etc..

In the above classification methods, a convex (C) nonparametric frontier is constructed

based on a certain group of training data. Then, this C nonparametric frontier is used to predict the group membership of test data. If a test data point is located within the C nonparametric frontier, then it has the same group membership as the training data which generate the C nonparametric frontier. Otherwise, it should be assigned to another group. We refer to these classification methods as C nonparametric frontier-based classification (C NPFC) methods.

These C NPFC methods are all constructed from a single group of data. In this respect, these methods should have a unique advantage in anomaly detection problems, where only the group of normal data is sufficiently available. However, the existing C NPFC methods originating from Troutt, Rai, and Zhang (1996) are mainly designed to solve a classical binary classification problem. Therefore, this unique advantage of relying on only a single group of data to achieve correct classification has not received any attention in the literature related to C NPFC methods.

To the best of our knowledge, the C NPFC methods have not been applied to solve the anomaly detection problem. Moreover, the C NPFC methods are essentially frontier analysis methods, and anomaly detection is one of the main tasks in supervised machine learning. Machine learning and frontier analysis are two relatively disconnected fields. In the literature, a research trend of applying well-known machine learning techniques to frontier analysis seems to emerge (e.g., Aparicio, Esteve, Rodriguez-Sala, and Zofio (2021); Valero-Carreras, Aparicio, and Guerrero (2021); Zhu, Zhu, and Emrouznejad (2021); Esteve, Aparicio, Rodriguez-Sala, and Zhu (2023)). However, research applying frontier analysis methods to machine learning seems not to be developing in parallel. Therefore, the adaptation of C NPFC methods to the anomaly detection problem can help bringing frontier analysis methods into the field of machine learning, thus creating some connection between two otherwise more or less unrelated fields.

In addition, there are in our opinion two key shortcomings of the existing C NPFC methods which may constrain their classification capacity. The goal of this contribution is to remedy both.

First, the existing C NPFC methods are limited to construct a C nonparametric frontier. If the boundary of the class happens to be C , then a C nonparametric frontier offers a reasonable estimate. But, with our ignorance as to the real shape of the boundary, the convexity assumption can be overly optimistic. Pendharkar, Rodger, and Yaverbaum (1999, p. 231) mention this as a potential harm to the capacity of the DEA frontier-based classification method while comparing it to the neural networks which are not constrained by convexity.

Second, the existing C NPFC methods are limited to situations in which all characteristic variables have the property called conditional monotonicity. That is, acceptability of a case to a class increases with the increase or decrease of all characteristic variables. Thus, a radial DEA model without outputs or without inputs are adopted in the literature (Lovell and Pastor (1999)). In conjunction, these two restrictions are severe - a generic data set need not be separable by a C boundary and it can simultaneously possess monotonically increasing and decreasing characteristics.

In this contribution, a general NPFC method is proposed to solve the anomaly detection problems and it eliminates the above two shortcomings. First, the convexity assumption is interpreted as reflecting a substitution relation between the characteristic variables. This relation does not always hold in practice. Therefore, we propose to relax the convexity assumption and construct a nonconvex (NC) NPFC method. This NC NPFC method is based on the Free Disposal Hull (FDH) model, initially proposed by Deprins, Simar, and Tulkens (1984). Solving the FDH model results in a monotonous and staircase shaped nonparametric frontier enveloping the observed data. This NC nonparametric frontier is more conservative than the C nonparametric frontier. Second, the assumption of free disposability is interpreted as reflecting the monotonic relation between the characteristic variables and the membership. Therefore, both monotonically increasing and decreasing characteristic variables can be incorporated into the model simultaneously. With these two modifications, a generalized NPFC method is constructed: it can portray both monotonically increasing and monotonically decreasing characteristic variables, and it can generate C or NC nonparametric frontiers.

To meet the above objectives, this contribution is structured as follows. Section 2 introduces the models and procedures used to construct the NPFC method. In Section 3, an experimental analysis is used to show the classification performance of our C and NC NPFC methods relative to that of the OC-SVM and SVDD methods. Finally, Section 4 is concluded with a summary of the contributions and a discussion of potential future research topics.

2 Nonparametric Frontier-based Classification Methods

2.1 Problem Description

In anomaly detection problems, there is usually only a sufficient number of normal observations. The number of anomalous observations is very limited and therefore being insufficient for training a classifier. Thus, the training set consists of normal observations only. Let $G = \{Z_1, \dots, Z_n\}$ be the set of training observations.

Each training observation $Z_j \in G$ is characterized by a number of characteristic variables. These characteristic variables can be exclusively differentiated into two monotonic types, namely the monotonically decreasing characteristic variables denoted by $X = \{x_1, \dots, x_m\}$ and the monotonically increasing characteristic variables denoted by $Y = \{y_1, \dots, y_m\}$. The former is also termed as input-like characteristic variables, and the latter is termed as output-like characteristic variables. Generally, the observation is represented by $Z_j = (X_j, Y_j) \in \mathbb{R}^m \times \mathbb{R}^s$.

The monotonicity constraints are prior-knowledge of the classification problem about the relationship between the characteristic variables and the group membership. Consider the example of credit card default. All other characteristics being the same, cardholders with higher annual income are less likely to default compared to cardholders with lower income. That is, the probability of default should not decrease in the presence of better characteristics while the rest remains the same. Specifically, a characteristic variable is defined as being output-like if the probability of being normal increases (decreases) with the increase (decrease) of its value, e.g., the annual income in the example of credit card default. A characteristic variable is defined as being input-like if the probability of being normal increases (decreases) with the decrease (increase) of its value.

Given the training set G , an acceptance possibility set (APS) is constructed from the training observations and the imposition of some weak maintained axioms. It is a data-based description of the normal group. Any data point within this APS is perceived as normal and anomalous otherwise. Then, the boundary of the APS, termed as a nonparametric frontier, is used for anomaly detection. It consists part of the normal training observations. A test data point that has the same characteristic variables as the training observations is classified as normal if it lies within the nonparametric frontier and anomalous otherwise.

2.2 Convex and Nonconvex Acceptance Possibility Set

In this subsection, the normal observations from the training set G are used to describe the domain containing all possible normal data points. It describes all possible combinations of characteristic values for which the corresponding evaluated data point can be classified as normal.

In production analysis, a production possibility set (PPS) is used to describe the attainable set in production. For all the combinations of the inputs and the output within the PPS, these are attainable (producible) under a certain given technology. Instead of discussing the producibility under the PPS, the attainable set in classification describes the attainability in accepting an observation as normal. Hence, we define an APS to describe the attainable set of the normal group based on the training set G .

If a data point has the same characteristic values as a normal observation from the training set G , then it is in the APS. Based on the monotonicity constraint, any data point with less X and more Y than an observation $Z_j \in G$ is perceived as having better characteristics and should be in the APS of the normal group. A free disposal set denoted by T_j is introduced to describe the situation under the monotonicity constraint. For every observation $Z_j \in G$, $T_j = \{(X, Y) \in \mathbb{R}^m \times \mathbb{R}^s \mid X \leq X_j \text{ and } Y \geq Y_j\}$.

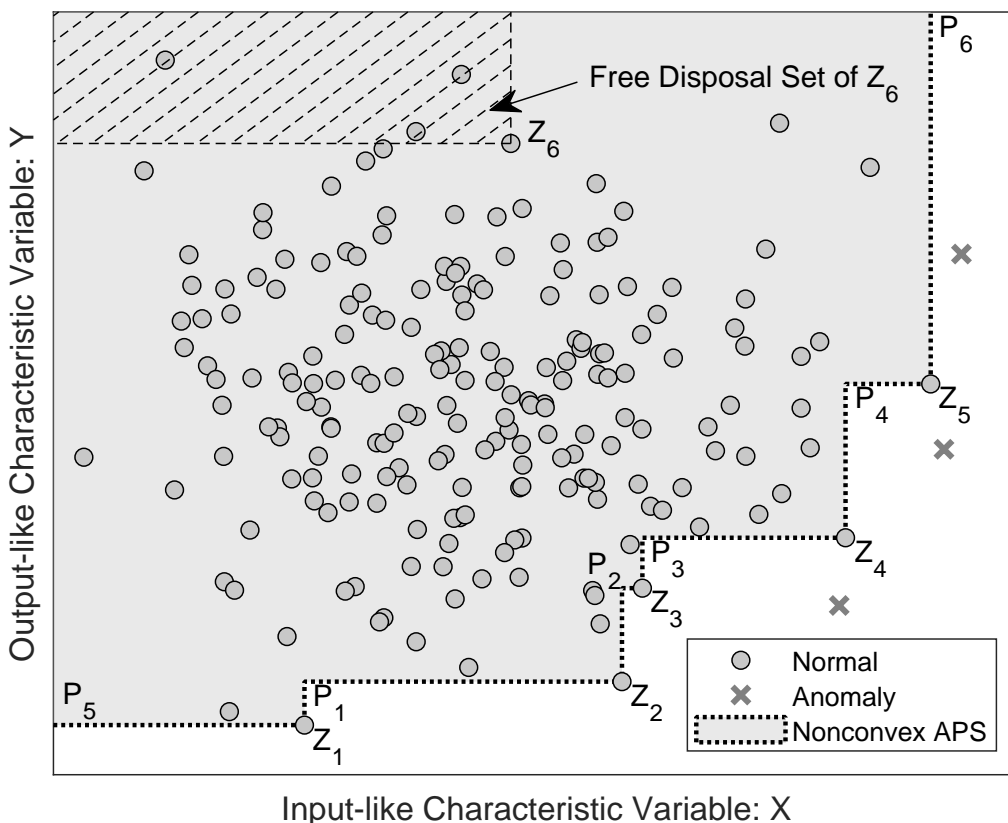
The union of all the free disposal sets of the observations from G constitutes a NC APS denoted by T_{NC} . Specifically, T_{NC} depicts the normal group based on the n training observations as follows:

$$\begin{aligned}
 T_{NC} &= \bigcup_{j=1}^n T_j \\
 &= \left\{ (X, Y) \in \mathbb{R}^m \times \mathbb{R}^s \mid \sum_{j=1}^n \lambda_j X_j \geq X, \sum_{j=1}^n \lambda_j Y_j \leq Y, \sum_{j=1}^n \lambda_j = 1, \lambda_j \in \{0, 1\} \right\}. \tag{1}
 \end{aligned}$$

Figure 1 sketches a typical figure displaying a NC APS. All grey circles are observations known from the normal group and constitute the training set G . These observations are characterized by two characteristic variables, namely $X \in \mathbb{R}^1$ and $Y \in \mathbb{R}^1$. For the monotonically decreasing characteristic variable X , the smaller its value the higher is the probability of belonging to the normal group. While for the monotonically increasing characteristic variable Y , the larger its value the higher is the probability of belonging to the normal group.

In Figure 1, the observations from G are known to be normal. Thus, they are apparently

Figure 1: Nonconvex APS of the normal group G



in the APS. Then, we take the observation Z_6 as an example to explain the free disposal set. The free disposal set is built based on the monotonic relation of the characteristic variables. If a data point has either a smaller X or a larger Y than Z_6 , then it is more likely belonging to the normal group than Z_6 does. Since Z_6 belongs to the normal group, thus a data point which has either a smaller X or a larger Y than Z_6 can be accepted as normal. This draws the dashed hatched area, which is located above and to the left of Z_6 . This dashed hatched area represents the free disposal set of Z_6 , namely T_6 . If a new data point is located within this dashed hatched area, then it is regarded as normal just like the observation Z_6 . For other observations from the training set G , their free disposal sets are derived in the same way. The union of all these free disposal sets constitutes the NC APS of the normal group. In Figure 1, this is the shaded area restricted to the second quadrant located above and to the right of the dotted polyline $P_5Z_1P_1Z_2P_2Z_3P_3Z_4P_4Z_5P_6$.

In addition to the monotonicity assumption, the convexity assumption is commonly adopted in the literature. Mathematically, the axiom on convexity implies that for any two

observations from one set, the linear combination of these two observations belong to the same set. In classification, this convexity axiom could explain a substitution relation between two characteristic variables. For example, both Z_2 and Z_4 in Figure 1 are normal training observations. With the convexity assumption, their linear combinations, which locates on the line between Z_2 and Z_4 , are also regarded as belonging to normal.

The C APS, denoted by T_C , is the convex hull of the NC APS. It depicts the normal group based on the n training observations as follows:

$$T_C = \left\{ (X, Y) \in \mathbb{R}^m \times \mathbb{R}^s \mid \sum_{j=1}^n \lambda_j X_j \geq X, \sum_{j=1}^n \lambda_j Y_j \leq Y, \sum_{j=1}^n \lambda_j = 1, \lambda_j \geq 0 \right\}. \quad (2)$$

Figure 2: Convex APS of the normal group G

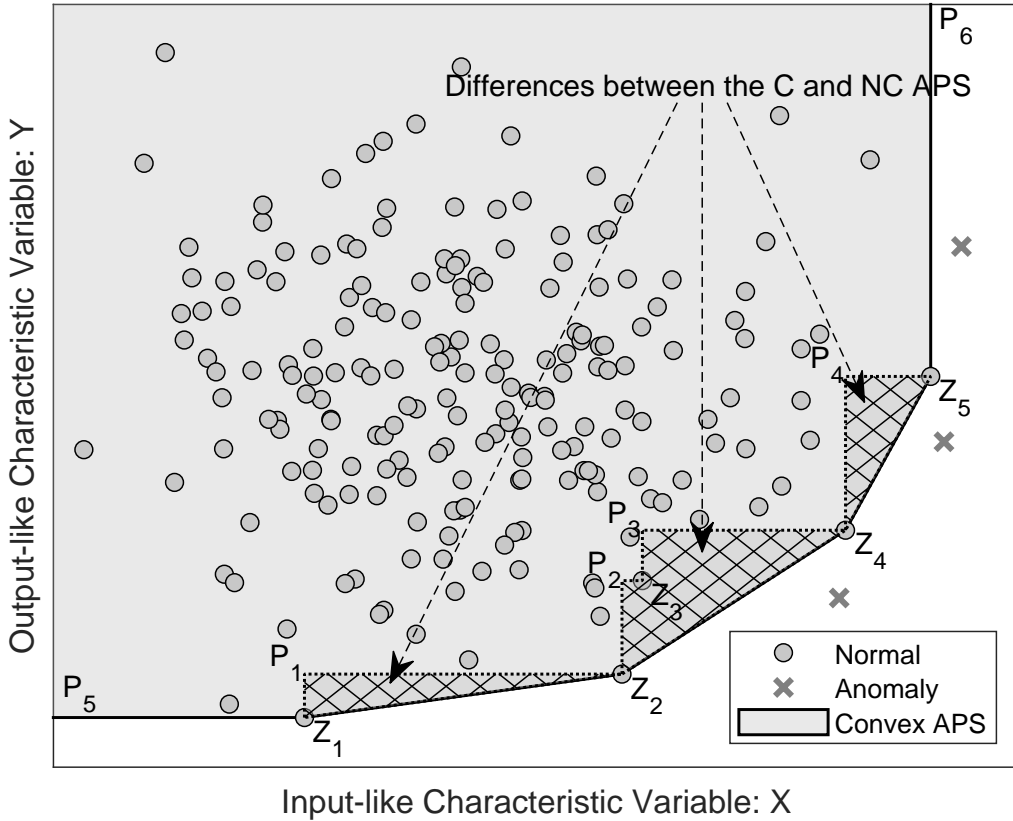


Figure 2 shows a figure of a C APS. The same training observations as in Figure 1 are used to illustrate the construction of a C APS. This C APS is built based on the NC APS with an additional assumption on convexity. Under the convexity assumption, the data points derived from the linear combination of observations must also be regarded as

belonging to the normal group. For example, the data points on the line Z_2Z_4 should be classified into the normal group due to the convexity assumption. Then, the data points in the polygon $Z_2P_2P_3Z_4Z_2$ marked by grid lines can also be classified into the normal group with the monotonicity assumption. Overall, three polygons marked by grid lines are added, namely $Z_1P_1Z_2Z_1$, $Z_2P_2P_3Z_4Z_2$ and $Z_4P_4Z_5Z_4$. Thus, the C APS of the normal group is the shaded area restricted to the second quadrant located above and to the right of the solid polyline $P_5Z_1Z_2Z_4Z_5P_6$.

For the NC case, the APS consists of data points which are located within the free disposal area of certain training observation from G . While for the C case, except for the above situation, if a data point is located within the free disposal area of a convex combination of two training observations from G , it also constitutes the APS and therefore belongs to the normal group. Obviously, $T_{NC} \subseteq T_C$: a NC monotonic hull is a subset of a C monotonic hull. Put differently, the NC APS provides a tighter envelopment of the training observations than the C APS does.

To simplify the expressions, we use the following notation to stand for the APS of the normal group under the NC and C cases:

$$T_\Lambda = \left\{ (X, Y) \in \mathbb{R}^m \times \mathbb{R}^s \mid \sum_{j=1}^n \lambda_j X_j \geq X, \sum_{j=1}^n \lambda_j Y_j \leq Y, \sum_{j=1}^n \lambda_j = 1, \lambda_j \in \Lambda \right\}. \quad (3)$$

where

$$(i) \Lambda \equiv \Lambda^C = \{\lambda_j \geq 0\}, \quad \text{or} \quad (ii) \Lambda \equiv \Lambda^{NC} = \{\lambda_j \in \{0, 1\}\}.$$

2.3 Convex and Nonconvex Nonparametric Frontiers

Instead of using all the training observations, the APS of the normal group can be simply described by a number of training observations located on the nonparametric frontier. These training observations have the least preferable characteristic values and are located on the worst-practice frontier. Any data point with better characteristics than these training observations is assigned to be normal. On the contrary, any data point with worse characteristics than these training observations is assigned to be abnormal. In the following, the nonparametric frontiers under the C and NC settings are introduced correspondingly.

Before constructing the nonparametric frontiers, we introduce the directional distance function (DDF) measure to gauge the relative distance of a data point Z to the frontier. Fol-

lowing Chambers, Chung, and Färe (1998), T_Λ is represented by the DDF measure ($D_{\Lambda,g}(Z)$) defined below:

$$D_{\Lambda,g}(Z) = \sup\{\delta \in \mathbb{R} \mid Z + \delta g \in T_\Lambda\}. \quad (4)$$

where $g = (g_X, g_Y) \in \mathbb{R}^m \times \mathbb{R}^s$ represents the projection direction. To be meaningful, $g_{x_i} > 0$ for all $i \in \{1, \dots, m\}$ and $g_{y_r} < 0$ for all $r \in \{1, \dots, s\}$. In this way, the characteristic variables X are non-decreasing and the characteristic variables Y are non-increasing while increasing the value of δ , which is the favorable behavior. Note that δ is a free decision variable that can take positive, zero or negative values.

All the C nonparametric frontier-based classification methods in the literature adopt either an input-oriented or an output-oriented radial efficiency measure. However, the adoption of these radial efficiency measures may lead to infeasibilities for the observations located outside the APS, when there are both input-like and output-like characteristic variables. While the DDF measure in expression (4) is well-defined for all possible data points, and for different monotonic types of characteristic variables.

The value of $D_{\Lambda,g}(Z)$ serves as an indicator that positions an observation relative to the frontier of the APS (T_Λ). A non-negative $D_{\Lambda,g}(Z)$ means that Z belongs to T_Λ . Specifically, an observation with $D_{\Lambda,g}(Z)$ equal to 0 means this observation is located on the frontier. If an observation Z is located outside T_Λ , then $D_{\Lambda,g}(Z)$ becomes negative and this observation is projected onto the frontier in the direction opposite to g .

Note that different choices of the direction vectors g lead to various distance values denoted by $D_{\Lambda,g}(Z)$. However, this choice does not change the sign of $D_{\Lambda,g}(Z)$. In the following, the direction vector is applied with $g = (|X_0|, -|Y_0|)$ for the observation $Z = (X_0, Y_0)$. This invests the DDF measure with a proportional interpretation (see Briec (1997); Kerstens and Van de Woestyne (2011)). Such a percentage interpretation is not indispensable to assign a membership, but it remains convenient.

Based on the APS of the normal group defined by expression (3), the proportional DDF measure is then computed accordingly. With respect to T_Λ , the DDF of a data point $Z_0 =$

(X_0, Y_0) is obtained by solving the following program:

$$\begin{aligned}
& \max_{\lambda_j, \bar{\delta}_\Lambda} \quad \bar{\delta}_\Lambda \\
& \text{s.t.} \quad \sum_{j=1}^n \lambda_j x_{i,j} \geq x_{i,0} + \bar{\delta}_\Lambda |x_{i,0}| \quad \forall i \in \{1, \dots, m\} \\
& \quad \quad \sum_{j=1}^n \lambda_j y_{r,j} \leq y_{r,0} - \bar{\delta}_\Lambda |y_{r,0}| \quad \forall r \in \{1, \dots, s\} \\
& \quad \quad \sum_{j=1}^n \lambda_j = 1 \\
& \quad \quad \lambda_j \in \Lambda \quad \quad \quad \forall j \in \{1, \dots, n\}
\end{aligned} \tag{5}$$

where

$$(i) \Lambda \equiv \Lambda^C = \{\lambda_j \geq 0\}, \quad \text{or} \quad (ii) \Lambda \equiv \Lambda^{\text{NC}} = \{\lambda_j \in \{0, 1\}\}.$$

In the C case, model (5) is a linear programming (LP) problem, while it involves solving a binary mixed integer program (BMIP) for the NC case. To remedy the computational issue in the NC case, a fast implicit enumeration-based method is proposed by Cherchye, Kuosmanen, and Post (2001) requiring only to compute minima and maxima of lists of ratios. Instead of solving a BMIP model, the following exact solution is obtained for model (5) under the NC case:

$$\bar{\delta}_{\Lambda^{\text{NC}}}^* = \max_{j=1, \dots, n} \left(\min_{i=1, \dots, m} \left(\frac{x_{i,0} - x_{i,j}}{|x_{i,0}|} \right), \min_{r=1, \dots, s} \left(\frac{y_{r,j} - y_{r,0}}{|y_{r,0}|} \right) \right). \tag{6}$$

The assumption on convexity differentiates the NC APS ($T_{\Lambda^{\text{NC}}}$) from the C APS (T_{Λ^C}). However, this does not change the definition of the DDF measure, only the value of the DDF measure may change. Thus, $\bar{\delta}_{\Lambda^{\text{NC}}} \leq \bar{\delta}_{\Lambda^C}$, since $T_{\Lambda^{\text{NC}}} \subseteq T_{\Lambda^C}$.

By solving model (5) for all observations from G , a frontier set denoted by F_Λ is generated. F_Λ consists of the observations from G that have $\bar{\delta}_\Lambda^* = 0$. Normally, the set F_Λ under the NC case is different from that under the C case. All frontier observations in F_{Λ^C} can also be found in $F_{\Lambda^{\text{NC}}}$. However, not all frontier observations in $F_{\Lambda^{\text{NC}}}$ belong to F_{Λ^C} , since some frontier observations under the NC case might be dominated by convex combinations of certain observations. Therefore, $F_{\Lambda^C} \subseteq F_{\Lambda^{\text{NC}}}$.

We consider Figures 1 and 2 in Section 2.2 as an example to illustrate the results of

model (5). Under the NC setting, model (5) is executed for all training observations. Only observations Z_1, Z_2, Z_3, Z_4 and Z_5 have $\bar{\delta}_{\Lambda^{\text{NC}}}^* = 0$, while the other observations have $\bar{\delta}_{\Lambda^{\text{NC}}}^* > 0$. Thus, the NC frontier is represented by the frontier set $F_{\Lambda^{\text{NC}}} = \{Z_1, Z_2, Z_3, Z_4, Z_5\}$. The NC nonparametric frontier is the dotted polyline $P_5Z_1P_1Z_2P_2Z_3P_3Z_4P_4Z_5P_6$. Similarly, model (5) is executed for all observations under the C setting. Observations Z_1, Z_2, Z_4 and Z_5 still have $\bar{\delta}_{\Lambda^{\text{C}}}^* = 0$, but observation Z_3 has $\bar{\delta}_{\Lambda^{\text{C}}}^* > 0$ as all the other observations do. Thus, $F_{\Lambda^{\text{C}}} = \{Z_1, Z_2, Z_4, Z_5\}$. The C nonparametric frontier is the solid polyline $P_5Z_1Z_2Z_4Z_5P_6$.

2.4 Nonparametric Frontier-based Classification Rules

The nonparametric frontier represented by the observations in the frontier set F_{Λ} is then used to determine the membership of a new data point. Specifically, the following model is used to calculate the distance of the new data point $Z_0 = (X_0, Y_0)$ relative to the nonparametric frontier:

$$\begin{aligned}
& \max_{\lambda_j, \delta_{\Lambda}} \quad \delta_{\Lambda} \\
& \text{s.t.} \quad \sum_{j \in F_{\Lambda}} \lambda_j x_{i,j} \geq x_{i,0} + \delta_{\Lambda} |x_{i,0}| \quad \forall i \in \{1, \dots, m\} \\
& \quad \quad \sum_{j \in F_{\Lambda}} \lambda_j y_{r,j} \leq y_{r,0} - \delta_{\Lambda} |y_{r,0}| \quad \forall r \in \{1, \dots, s\} \\
& \quad \quad \sum_{j \in F_{\Lambda}} \lambda_j = 1 \\
& \quad \quad \lambda_j \in \Lambda \quad \quad \quad \forall j \in F_{\Lambda}
\end{aligned} \tag{7}$$

where

$$(i) \Lambda \equiv \Lambda^{\text{C}} = \{\lambda_j \geq 0\}, \quad \text{or} \quad (ii) \Lambda \equiv \Lambda^{\text{NC}} = \{\lambda_j \in \{0, 1\}\}.$$

For the new data point $Z_0 = (X_0, Y_0)$, the optimal distance measure calculated from model (7) (δ_{Λ}^*) is the same as that calculated from model (5) ($\bar{\delta}_{\Lambda}^*$). The difference is that only the observations in the frontier set F_{Λ} are used in the left-hand side of the inequalities in model (7). Although $\delta_{\Lambda}^* = \bar{\delta}_{\Lambda}^*$ always holds, the decrease in sample size can save some computational time.

The decision variable δ_{Λ} in model (7) is a free variable. If $\delta_{\Lambda}^* \geq 0$, then it indicates there exists a projection point that dominates the new data point $Z_0 = (X_0, Y_0)$. This projection point is generated from the left-hand side of the inequality constraints in model (7) and is represented by $Z_b = (\sum_{j \in F_{\Lambda}} \lambda_j^* X_j, \sum_{j \in F_{\Lambda}} \lambda_j^* Y_j)$. It is either an observation from the NC

frontier set $F_{\Lambda^{NC}}$ or a convex combination of the observations from F_{Λ^C} . In the case where $\delta_{\Lambda}^* \geq 0$, the following inequalities hold: $\sum_{j \in F_{\Lambda}} \lambda_j^* X_j \geq X_0$ and $\sum_{j \in F_{\Lambda}} \lambda_j^* Y_j \leq Y_0$. Comparing to the projection point Z_b which is normal, the new data point $Z_0 = (X_0, Y_0)$ has less X and more Y . Therefore, it should be assigned to the normal group.

By contrast, if $\delta_{\Lambda}^* < 0$, then the new data point Z_0 dominates the projection point Z_b . That is, $\sum_{j \in S_{\Lambda}} \lambda_j^* X_j < X_0$ and $\sum_{j \in S_{\Lambda}} \lambda_j^* Y_j > Y_0$. The projection point Z_b is on the boundary of the APS. Comparing to the projection point Z_b , the new data point Z_0 has more X and less Y , and therefore it is situated outside the APS. Hence, the new data point Z_0 is assigned as an anomaly if there is no further information.

To sum up, the membership of the new data point Z_0 is determined by the sign of the optimal DDF δ_{Λ}^* calculated from model (7). The classification rules are summarized as follows:

$$\begin{aligned} & \text{If } \delta_{\Lambda}^* \geq 0, \text{ then } Z_0 \text{ belongs to the normal group;} \\ & \text{Otherwise, } Z_0 \text{ belongs to the group of anomalies.} \end{aligned} \tag{8}$$

Algorithm 1 NPFC Methods

Given:

Training data set: G , Input-like characteristics variables: X

Output-like characteristics variables: Y , Assumption on Convexity

Training:

- 1: Let $F_{\Lambda} = \emptyset$
- 2: **for** $j = 1, 2, \dots, n$
- 3: Calculate the corresponding $\bar{\delta}_{\Lambda}^*$ using model (5)
- 4: **if** $\bar{\delta}_{\Lambda}^* = 0$, **then**
- 5: Z_j is added to F_{Λ}
- 6: **end if**
- 7: **end for**
- 8: Export the trained F_{Λ}

Testing:

- 1: Calculate δ_{Λ}^* of a new data point Z_0 using model (7) and the trained F_{Λ}
 - 2: Classify Z_0 using Rule (8)
-

3 Experimental Analysis

3.1 Experimental Setup

The proposed C and NC NPFC methods are applied to a real-life data set (see Cox, Johnson, and Kafadar (1982)).¹ This data set arose in a study that aims at identifying carriers of a rare genetic disorder. Because the disease is rare, the number of carriers whose data are available is relatively smaller comparing to the number of normal samples. Specifically, this biomedical data set contains 194 observations after excluding 15 observations which have missing values. Among them, 127 observations are normal samples and 67 observations are disease carriers which are deemed as anomalies. Each observation is characterized by five characteristic variables, namely, age and 4 blood measurements.

A k -fold cross-validation is used in this analysis. Specifically, k takes the values of 2, 5 and 10. In the detailed setting, we stick to the general situation of anomaly detection where only one group of data is available. That is, only the normal observations are used for the training process, while all the abnormal observations are excluded from the training process and are waited to be classified in the test process. For example, when k takes the value of 2, then the normal samples are divided into 2 disjoint subsets: each subset contains around 50% observations of the normal sample. In the first iteration, a classifier is trained using 63 normal observations. Then, the trained classifier is examined with a test data set which contains the remaining 64 normal observations and the 67 abnormal observations. In the second iteration, the classifier is trained using the 64 normal observations. Then, the trained classifier is examined with a test data which contains the remaining 63 normal observations and the same 67 abnormal observations. The classification performance reported by the 2-fold cross-validation is thus the average of the corresponding measures computed in each iteration. Similar settings are applied for the 5-fold and 10-fold cross-validation.

The classification performance of the classifiers is characterized by 6 measures, namely, accuracy, precision, recall, specificity, F_1 score and G-mean. These performance measures are listed in equations (9) to (14):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

¹The data set is collected from the Statlib data archive at: <http://lib.stat.cmu.edu/datasets/>.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (11)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (12)$$

$$F_1 \text{ score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (13)$$

$$\text{G-mean} = \sqrt{\text{Precision} \cdot \text{Recall}} \quad (14)$$

whereby TN , TP , FN , and FP represent true negative, true positive, false negative and false positive cases, respectively. An anomaly is depicted as positive, while the normal case is depicted as negative.

Accuracy captures the percentage of correctly predicted samples out of all test samples. Precision depicts the percentage of true positive (abnormal) samples out of all predicted positive samples. Recall represents the percentage of correctly predicted positive (abnormal) samples out of all true positive samples. Specificity indicates the percentage of correctly predicted negative (normal) samples out of all true negative samples. To reflect and penalize the inequalities between the precision and the recall, both harmonic and geometric means are used, namely the F_1 score and the G-mean.

Note that when the data set is unbalanced, then the measure of accuracy can not appropriately reflect the overall classification performance (Kumar, Goel, Sinha, and Bhardwaj (2022)). Thus, the F_1 score and the G-mean are used for characterizing the overall classification performance in this study. However, the results of accuracy are still reported.

The performance of the proposed NPFC methods are compared with the OC-SVM and SVDD methods. It is important to remark that this comparison is intended to show that the proposed NPFC methods can be a good candidate for the anomaly detection problem, rather than always beat the best OC-SVM and SVDD methods. Thus, the OC-SVM method with a Gaussian kernel and the SVDD method with a polynomial kernel are chosen to be compared with. The OC-SVM method with a Gaussian kernel is implemented using the MATLAB function “`fitsvm`”. The SVDD method with a polynomial kernel is implemented using the MATLAB code available from Qiu (2022).

In implementing the NPFC method, age is taken as an input-like characteristic variable while the four blood measures are taken as output-like characteristic variables. This is supported by the expert knowledge released in Cox, Johnson, and Kafadar (1982) which claims that young people tend to have higher measurements. As for the assumption on convexity, it is decided by the potential substitution relation between the characteristic variables. How-

ever, in this biomedical data set, there is no prior information on this substitution relation. Thus, both the C and NC NPFC methods are adopted in this analysis.

3.2 Experimental Results

The classification performance results under the 2-fold cross-validation are presented in Table 1. The four columns correspond to the performance results under the OC-SVM, SVDD, C and NC NPFC methods, respectively. Horizontally, each row displays the average classification performance of the various methods under the corresponding measures (9)-(14). In each row, the best result among the four methods is highlighted in bold.

Table 1: Performance Results of 4 Classifiers under the 2-fold Cross-Validation

	$k=2$			
	OC-SVM	SVDD	C NPFC	NC NPFC
Accuracy	0.8161	0.7315	0.8389	0.7587
Precision	0.8983	0.9444	0.8670	0.7164
Recall	0.7239	0.5224	0.8209	0.8806
Specificity	0.9134	0.9531	0.8575	0.6303
F_1 Score	0.8016	0.6439	0.8415	0.7897
G-mean	0.8063	0.6855	0.8427	0.7941

Three observations can be made from Table 1. First, regarding the overall classification performance which is reflected by the three overall measures, i.e., accuracy, F_1 score and G-mean, it is observed that the C NPFC method achieves the best overall performance with respect to all three overall measures. The OC-SVM method also performs well with respect to the three overall measures, second only to the C NPFC method. Second, the SVDD method achieves the best performance in identifying the normal samples. It is shown in Table 1 that 95.31% of the normal samples have been correctly identified with the SVDD method. Its overall performance is relatively poor because it fails to detect 47.76% of the abnormal samples. Third, in contrast to the SVDD method, the NC NPFC method correctly detects 88.06% of the abnormal samples. However, its overall performance is also relatively poor, because it only identifies 63.03% of the normal samples.

A further comparative analysis is done for the four listed methods under the 5-fold and 10-fold cross-validation. These performance results are reported in Table 2. Horizontally, the first block of the columns contains the results under the 5-fold cross-validation while the second block is under the 10-fold cross-validation. Within each of these two horizontal

blocks, the results of the performance measures (9)-(14) are arranged in the same way as in Table 1. In each row, the best result within each block is highlighted in bold.

Table 2: Performance Results of 4 Classifiers under the 5-fold and 10-fold Cross-Validation

	$k=5$				$k=10$			
	OC-SVM	SVDD	C NPFC	NC NPFC	OC-SVM	SVDD	C NPFC	NC NPFC
Accuracy	0.7684	0.6536	0.7921	0.8224	0.7440	0.5118	0.7729	0.8369
Precision	0.9659	0.9857	0.9451	0.9057	0.9881	0.9944	0.9813	0.9584
Recall	0.7075	0.5313	0.7642	0.8537	0.7045	0.4224	0.7448	0.8433
Specificity	0.9292	0.9760	0.8646	0.7382	0.9526	0.9846	0.9231	0.8051
F_1 Score	0.8156	0.6789	0.8423	0.8762	0.8221	0.5915	0.8462	0.8968
G-mean	0.8261	0.7166	0.8484	0.8779	0.8341	0.6472	0.8546	0.8988

Similarly, three observations with respect to the overall performance, the performance of identifying normal cases, and the performance of detecting anomalies are obtained from Table 2. First, the NC NPFC method is observed to have the best overall performance under both cross-validation settings. The C NPFC method now ranks second in overall performance, slightly behind the NC NPFC method. Second, the SVDD method always performs the best in correctly identifying normal samples. When k takes the value of 10, it can correctly identify 99.44% of normal samples. However, in detecting anomalies, the SVDD method performs rather worse with a recall of 42.24%. Third, the NC NPFC method is consistently superior in correctly detecting abnormal samples. Its recall rates are 85.37% and 84.33% under 5-fold and 10-fold cross-validation, respectively. Meanwhile, the NC NPFC method shows a significant improvement in the performance of identifying normal samples. Thus, it ranks first in overall performance under both cross-validation settings.

The above comparative analysis is derived from a specific k -fold cross-validation. In Figure 3, the performance results with various values of k are displayed and compared. Figures 3(b) and 3(c) show the classification performance under the two overall performance measures, F_1 score and the *G-mean*. Figures 3(d) to 3(f) show the classification performance for the normal or abnormal groups accordingly. In every sub-figure, the blue dashed line marked with crosses represents the results of the OC-SVM method; the green dash-dotted line marked with asterisks represents the results of the SVDD method; the red solid line marked with circles represents the results of the NC NPFC method; the black dashed line marked with squares represents the results of the C NPFC method.

The dynamic process from 2-fold cross-validation to 10-fold cross-validation is accompanied by an increasing number of normal samples for the training process. With increased normal training samples, the classification methods are expected to be better trained. Note

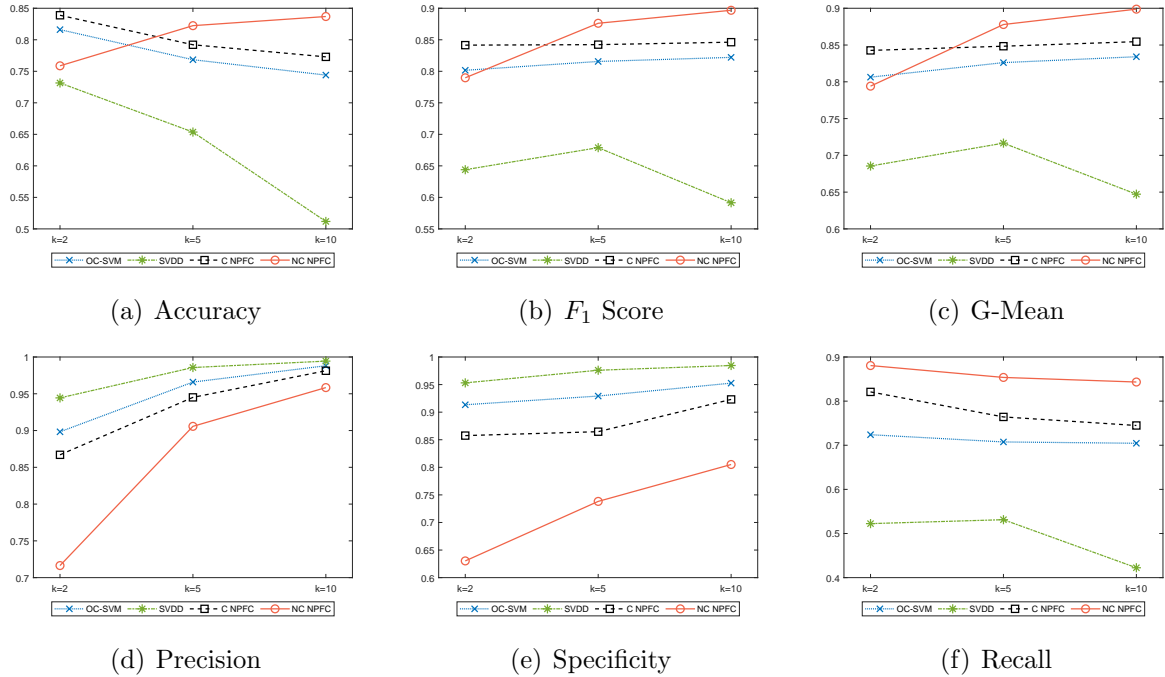


Figure 3: Performance Results of 4 Classifiers for Various k -fold Cross-Validation

that the data sets become unbalanced under the 5-fold and 10-fold cross-validation, so accuracy can not appropriately reflect the overall classification performance. Thus, the results on accuracy are displayed in Figure 3(a) but are not analyzed. From Figure 3, the increase of normal samples during training has a similar impact on the classification methods, but to different degrees for various methods.

First, as more normal samples are used for training, more normal samples can be correctly identified in the test process. This is validated in Figure 3(e) by the progressively increase in specificity and holds for all four classification methods. The increase in specificity means that less normal samples are misclassified as being abnormal. Correspondingly, the precision in Figure 3(d) also shows a consistent increase for all four classification methods. Among the four classification methods, the improvement in specificity and precision of the NC NPFC method is the most significant.

Second, while adding normal training samples increases specificity, it has a negative effect on recall (Figure 3(f)). From the 2-fold cross-validation to the 10-fold cross-validation, the recall of all classification methods decreases. For the SVDD method, this decrease in recall is large with a value of 10.00%, but for the NC NPFC method, the decrease in recall is smaller with a value of only 3.73%.

Third, from Figures 3(b) and 3(c), the increase in normal training samples has an in-

consistent impact on the overall performance for different classification methods. This is because the increase in normal training samples has an opposite impact on correctly identifying normal and abnormal samples. For the SVDD method, when k goes from 5 to 10, its overall classification performance deteriorates because the decrease in recall (10.90%) greatly outweighs the increase in specificity (0.85%). As for the OC-SVM, C and NC NPFC methods, the positive impact outweighs the negative impact. Thus, their overall classification performance is progressively improving.

In general, the proposed NPFC methods show a competitive classification performance, and even outperform the listed OC-SVM and SVDD methods in some of the performance measures. Moreover, they show unique advantages in correctly detecting abnormal samples, especially the NC NPFC method. As the number of normal training samples increases, the NC NPFC method can greatly improve its performance in identifying the normal cases while maintaining a relatively better performance in detecting the anomalies. All these support that the proposed NPFC methods, especially the NC NPFC method, can be well applied to the anomaly detection problem.

4 Conclusions

Although anomaly detection is a popular research problem, no consensus has been reached on the best classification method. This contribution proposes for the first time that the NPFC method can be used for anomaly detection. In the NPFC method, the nonparametric frontier is generated from the group of normal training data and consists of only a few training observations. Moreover, the shape of the nonparametric frontier is determined by the training observations and the imposition of some weak maintained axioms, rather than being predetermined as a hyperplane or a hypersphere. A test data point only has to be compared with this nonparametric frontier for deriving its membership.

In addition to bringing the NPFC method to anomaly detection, this contribution also makes two innovations at the methodological level. First, the convexity assumption is explained as a substitution relation between the characteristic variables, thus it can be reasonably relaxed. Having the convexity assumption relaxed, a NC NPFC method is constructed for anomaly detection and it ends up with a better classification performance comparing to the C NPFC method. Second, the other assumption of free disposability is explained to reflect the monotonic relation between the characteristic variables and the membership. Therefore, characteristic variables with both the monotonically increasing and the monoton-

ically decreasing relations can be included in the model without data transformation. To sum up, assigning reasonable interpretations to the assumptions justifies the NPFC methods and also contributes to the construction of a more general NPFC method.

The experiment analysis on a biomedical data set reveals that the proposed NPFC methods have competitive overall performance and have consistent advantages in detecting abnormal samples. This advantage in correctly detecting abnormal samples is consistent with the goal of anomaly detection. Moreover, the NC NPFC method performs better than the C NPFC method, especially in detecting abnormal samples. Thus, if there is no prior information on the substitution relation among characteristic variables, then the NC NPFC method should be favored.

We end with developing some perspectives for potential future research. First, it is an open question to which extent the proposed NPFC methods could be further enhanced for better classification performance by a further relaxation of some of the axioms inherited from production theory. Just as relaxing convexity yields a monotonous frontier instead of a C piecewise linear frontier, one may wonder whether it is possible to weaken the currently maintained axiom of disposability. A recent theoretical attempt to do so is developed in Briec, Kerstens, and Van de Woestyne (2016) and empirically implemented in Briec, Kerstens, and Van de Woestyne (2018). Second, one may equally wonder to which extent the same ideas can be transposed in the limited literature employing double separating frontiers in a classification setting (e.g, see Sueyoshi (2006), Chang and Kuo (2008) and Wu, An, and Liang (2011)). Third, while we have in this contribution compared the NPFC methods to the OC-SVM and SVDD methods, it could be interesting to compare the best of the NPFC methods to some of the best performing state of the art classification methods in anomaly detection to check their relative classification and prediction accuracies.

References

- AHMED, M., A. N. MAHMOOD, AND M. R. ISLAM (2016): “A Survey of Anomaly Detection Techniques in Financial Domain,” *Future Generation Computer Systems*, 55, 278–288.
- AL-HASHEDI, K. G., AND P. MAGALINGAM (2021): “Financial Fraud Detection Applying Data Mining Techniques: A Comprehensive Review from 2009 to 2019,” *Computer Science Review*, 40, 100402.
- ALAM, S., S. K. SONBHADRA, S. AGARWAL, AND P. NAGABHUSHAN (2020): “One-Class Support Vector Classifiers: A Survey,” *Knowledge-Based Systems*, 196, 105754.

- APARICIO, J., M. ESTEVE, J. J. RODRIGUEZ-SALA, AND J. L. ZOFIO (2021): *The Estimation of Productive Efficiency Through Machine Learning Techniques: Efficiency Analysis Trees* vol. 312, pp. 51–92. Springer International Publishing, Cham.
- BANKER, R. D., A. CHARNES, AND W. W. COOPER (1984): “Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis,” *Management Science*, 30(9), 1078–1092.
- BHATTACHARYYA, S., S. JHA, K. THARAKUNNEL, AND J. C. WESTLAND (2011): “Data Mining for Credit Card Fraud: A Comparative Study,” *Decision Support Systems*, 50(3), 602–613.
- BRIEC, W. (1997): “A Graph-Type Extension of Farrell Technical Efficiency Measure,” *Journal of Productivity Analysis*, 8(1), 95–110.
- BRIEC, W., K. KERSTENS, AND I. VAN DE WOESTYNE (2016): “Congestion in Production Correspondences,” *Journal of Economics*, 119(1), 65–90.
- (2018): “Hypercongestion in Production Correspondences: An Empirical Exploration,” *Applied Economics*, 50(27), 2938–2956.
- CHAMBERS, R., Y. CHUNG, AND R. FÄRE (1998): “Profit, Directional Distance Functions, and Nerlovian Efficiency,” *Journal of Optimization Theory and Applications*, 98(2), 351–364.
- CHANG, D., AND Y. KUO (2008): “An Approach for the Two-group Discriminant Analysis: An Application of DEA,” *Mathematical and Computer Modelling*, 47(9-10), 970–981.
- CHERCHYE, L., T. KUOSMANEN, AND T. POST (2001): “FDH Directional Distance Functions with An Application to European Commercial Banks,” *Journal of Productivity Analysis*, 15(3), 201–215.
- CLIFTON, L., D. A. CLIFTON, Y. ZHANG, P. WATKINSON, L. TARASSENKO, AND H. YIN (2014): “Probabilistic Novelty Detection with Support Vector Machines,” *IEEE Transactions on Reliability*, 63(2), 455–467.
- COX, L. H., M. M. JOHNSON, AND K. KAFADAR (1982): “Exposition of Statistical Graphics Technology,” in *Proceedings of the Statistical Computation Section*, pp. 55–56. American Statistical Association, Washington, D.C.

- DA SILVA, D. B., D. SCHMIDT, C. A. DA COSTA, R. DA ROSA RIGHI, AND B. ESKOFIER (2021): “Deepsigns: A Predictive Model Based on Deep Learning for The Early Detection of Patient Health Deterioration,” *Expert Systems with Applications*, 165, 113905.
- DEPRINS, D., L. SIMAR, AND H. TULKENS (1984): “Measuring Labor Efficiency in Post Offices,” in *The Performance of Public Enterprises: Concepts and Measurements*, ed. by M. Marchand, P. Pestieau, and H. Tulkens, pp. 243–268. North Holland, Amsterdam.
- EL-YANIV, R., AND M. NISENSEN (2006): “Optimal Single-Class Classification Strategies,” *Advances in Neural Information Processing Systems*, 19.
- EMROUZNEJAD, A., R. D. BANKER, AND L. NERALIC (2019): “Advances in Data Envelopment Analysis: Celebrating the 40th Anniversary of DEA and the 100th Anniversary of Professor Abraham Charnes’ Birthday,” *European Journal of Operational Research*, 278(2), 365–367.
- EMROUZNEJAD, A., AND G.-L. YANG (2018): “A Survey and Analysis of the First 40 Years of Scholarly Literature in DEA: 1978-2016,” *Socio-Economic Planning Sciences*, 61, 4–8.
- ESTEVE, M., J. APARICIO, J. J. RODRIGUEZ-SALA, AND J. ZHU (2023): “Random Forests and the Measurement of Super-Efficiency in the Context of Free Disposal Hull,” *European Journal of Operational Research*, 304(2), 729–744.
- JIANG, T., Y. LI, W. XIE, AND Q. DU (2020): “Discriminative Reconstruction Constrained Generative Adversarial Network for Hyperspectral Anomaly Detection,” *IEEE Transactions on Geoscience and Remote Sensing*, 58(7), 4666–4679.
- KAFFASH, S., R. AZIZI, Y. HUANG, AND J. ZHU (2020): “A Survey of Data Envelopment Analysis Applications in the Insurance Industry 1993–2018,” *European Journal of Operational Research*, 284(3), 801–813.
- KERSTENS, K., AND I. VAN DE WOESTYNE (2011): “Negative Data in DEA: A Simple Proportional Distance Function Approach,” *Journal of the Operational Research Society*, 62(7), 1413–1419.
- KHAN, S. S., AND M. G. MADDEN (2014): “One-Class Classification: Taxonomy of Study and Review of Techniques,” *Knowledge Engineering Review*, 29(3), 345–374.
- KUMAR, A., S. GOEL, N. SINHA, AND A. BHARDWAJ (2022): “A Review on Unbalanced Data Classification,” in *Proceedings of International Joint Conference on Advances in*

- Computational Intelligence*, ed. by M. S. Uddin, P. K. Jamwal, and J. C. Bansal, pp. 197–208, Singapore. Springer Nature Singapore.
- LEON, C. F., AND F. PALACIOS (2009): “Evaluation of Rejected Cases in an Acceptance System with Data Envelopment Analysis and Goal Programming,” *Journal of the Operational Research Society*, 60(10), 1411–1420.
- LOVELL, K., AND J. T. PASTOR (1999): “Radial DEA Models without Inputs or without Outputs,” *European Journal of Operational Research*, 118(1), 46–51.
- MARKOU, M., AND S. SINGH (2003a): “Novelty Detection: A Review-Part 1: Statistical Approaches,” *Signal Processing*, 83(12), 2481–2497.
- (2003b): “Novelty Detection: A Review-Part 2: Neural Network Based Approaches,” *Signal Processing*, 83(12), 2499–2521.
- MOYA, M. M., M. W. KOCH, AND L. D. HOSTETLER (1993): “One-Class Classifier Networks for Target Recognition Applications,” in *World Congress on Neural Networks*, vol. 3, pp. 797–801. Lawrence Erlbaum Associates, Portland.
- NOBLE, W. S. (2006): “What is a Support Vector Machine?,” *Nature Biotechnology*, 24(12), 1565–1567.
- PANG, G., C. SHEN, L. CAO, AND A. V. D. HENGEL (2021): “Deep Learning for Anomaly Detection: A Review,” *ACM Computing Surveys (CSUR)*, 54(2), 1–38.
- PARK, C., J. Z. HUANG, AND Y. DING (2010): “A Computable Plug-In Estimator of Minimum Volume Sets for Novelty Detection,” *Operations Research*, 58(5), 1469–1480.
- PENDHARKAR, P., M. KHOSROWPOUR, AND J. RODGER (2000): “Application of Bayesian Network Classifiers and Data Envelopment Analysis for Mining Breast Cancer Patterns,” *Journal of Computer Information Systems*, 40(4), 127–132.
- PENDHARKAR, P., J. RODGER, AND G. YAUERBAUM (1999): “Association, Statistical, Mathematical and Neural Approaches for Mining Breast Cancer Patterns,” *Expert Systems with Applications*, 17(3), 223–232.
- PENDHARKAR, P. C. (2002): “A Potential Use of Data Envelopment Analysis for the Inverse Classification Problem,” *Omega*, 30(3), 243–248.
- QIU, K. (2022): “Support Vector Data Description (SVDD Version 2.2),” pp. Available from: <https://github.com/iqiukp/SVDD-MATLAB>.

- RUFF, L., J. R. KAUFFMANN, R. A. VANDERMEULEN, G. MONTAVON, W. SAMEK, M. KLOFT, T. G. DIETTERICH, AND K.-R. MÜLLER (2021): “A Unifying Review of Deep and Shallow Anomaly Detection,” *Proceedings of the IEEE*, 109(5), 756–795.
- SCHÖLKOPF, B., R. C. WILLIAMSON, A. SMOLA, J. SHAWE-TAYLOR, AND J. PLATT (1999): “Support Vector Method for Novelty Detection,” in *Advances in Neural Information Processing Systems*, ed. by S. A. Solla, K.-R. Müller, and T. K. Leen, vol. 12, pp. 582–588. MIT Press, Cambridge, Massachusetts.
- SEIFORD, L., AND J. ZHU (1998): “An Acceptance System Decision Rule with Data Envelopment Analysis,” *Computers & Operations Research*, 25(4), 329–332.
- SUEYOSHI, T. (2006): “DEA-Discriminant Analysis: Methodological Comparison Among Eight Discriminant Analysis Approaches,” *European Journal of Operational Research*, 169(1), 247–272.
- TAX, D. M. J., AND R. P. W. DUIN (1999): “Support Vector Domain Description,” *Pattern Recognition Letters*, 20(11-13), 1191–1199.
- TAX, D. M. J., AND P. JUSZCZAK (2003): “Kernel Whitening for One-Class Classification,” *International Journal of Pattern Recognition and Artificial Intelligence*, 17(3), 333–347.
- TROUTT, M., A. RAI, AND A. ZHANG (1996): “The Potential Use of DEA for Credit Applicant Acceptance Systems,” *Computers & Operations Research*, 23(4), 405–408.
- VALERO-CARRERAS, D., J. APARICIO, AND N. M. GUERRERO (2021): “Support Vector Frontiers: A New Approach for Estimating Production Functions Through Support Vector Machines,” *Omega*, 104, 102490.
- WU, J., Q. AN, AND L. LIANG (2011): “A Modified Super-efficiency DEA Approach for Solving Multi-groups Classification Problems,” *International Journal of Computational Intelligence Systems*, 4(4), 606–618.
- YAN, H., AND Q. WEI (2011): “Data Envelopment Analysis Classification Machine,” *Information Sciences*, 181(22), 5029–5041.
- ZHU, N., C. ZHU, AND A. EMROUZNEJAD (2021): “A Combined Machine Learning Algorithms and DEA Method for Measuring and Predicting the Efficiency of Chinese Manufacturing Listed Companies,” *Journal of Management Science and Engineering*, 6(4), 435–448.