# Optimizing Interval Data Classification Through Two-Stage Mixed Integer Programming: A Discriminant Analysis Perspective

Sanaz Hami Hassan Kiyadeh[1] , Javad Pourqasem[2*] , Appasamy Saraswathi[3]

[1] Department of Mathematics, The University of Alabama, 35487 Alabama, USA
[2] Department of Computer Engineering, Urmia University, 5756151818 Urmia, Iran
[3] Department of Mathematics, SRM Institute of Science and Technology, 603203 Kattankulathur, India

\* Correspondence: Javad Pourqasem (jpourmail@gmail.com)

**Abstract:** Efficient classification of interval data presents considerable challenges, particularly when group overlaps and data uncertainty are prevalent. This study introduces an innovative two-stage Mixed Integer Programming (MIP) framework for discriminant analysis (DA), which is designed to minimize misclassification of vertices while effectively addressing the problem of overlapping groups. By incorporating interval data structures, the proposed model captures both the shared characteristics within groups and the distinct separations between them. The first stage of the model focuses on the identification of group-specific boundaries, while the second stage refines classification by incorporating probabilistic estimates of group memberships. A Monte Carlo simulation is employed to evaluate the robustness of the model under conditions of imprecision and noise, and the results demonstrate its superior capability in handling overlapping data and classifying uncertain observations. Validation through numerical experiments illustrates the model's effectiveness in accurately resolving group overlaps, thereby improving classification performance. The approach offers significant advantages over traditional methods by probabilistically estimating group memberships, thus enhancing decision-making processes in uncertain environments. These findings suggest that the proposed MIP framework holds substantial promise for applications across a range of complex decision-making scenarios, such as those encountered in finance, healthcare, and engineering, where data imprecision is a critical concern.

**Keywords:** Interval data; Discriminant analysis (DA); Data envelopment analysis (DEA); Mixed integer programming; Monte Carlo method; Data classification

## 1 Introduction

Discriminant analysis (DA) is a statistical technique used to predict the group membership of a new observation. A set of known observations is employed to estimate the weights or parameters, as well as to compute the discriminant score, which serves to categorize the observations into two groups, $G_1$ and $G_2$. The method typically involves either minimizing the total deviations of misclassified observations or maximizing the total deviations of correctly classified observations. A new observation is classified into one of the groups by comparing its discriminant score to the evaluation score derived from the discriminant function.

Data envelopment analysis (DEA) evaluates the relative efficiency of a set of homogeneous decision-making units (DMUs) [1]. DEA is particularly effective when multiple inputs and outputs are involved, as traditional methods may struggle to assess efficiency in such complex scenarios [2]. It is especially valuable when conventional approaches, such as ratios (e.g., cost per unit), are insufficient due to the multifaceted nature of the inputs and outputs [3]. Despite their differences, there are notable similarities between DEA and DA. DA is used to differentiate two sets based on prior knowledge, whereas DEA categorizes two sets into efficient and inefficient classifications without the need for prior knowledge [4]. The goal programming method can be applied within each of these frameworks to improve their efficacy [5].

A novel non-parametric discriminant analysis approach has recently been proposed, which is capable of generating a set of weights for a linear discriminant function, thereby producing an assessment score to determine the group

membership of observations [6, 7]. This non-parametric approach is referred to as DEA/DA. However, the original DEA/DA method does not accommodate negative data. To address this limitation, an extended version of DEA/DA has been introduced, known as extended DEA/DA, which can handle negative data effectively [8].

The extended DEA/DA approach aims to minimize the overall discrepancies in misclassified observations. In contrast, the primary criterion for evaluating a DA model is the total number of observations that it classifies correctly. Sueyoshi [9] proposed a novel MIP technique to extend the DEA formulation [10, 11], which estimates weights by minimizing the total number of misclassified observations.

However, none of Sueyoshi's models address the classification of imprecise data. Jahanshahloo et al. [12] introduced a method that overcomes a limitation of Sueyoshi's models. Furthermore, Jahanshahloo et al. [13] proposed a method for classifying interval data using DA models by minimizing the sum of the external deviations of all misclassified vertices in the interval data. In recent years, numerous scholars have developed and implemented innovative models and applications of DEA, extending its use to a variety of fields, including energy efficiency, healthcare management, and sustainability assessment [14–21]. These advancements underscore the versatility of DEA in addressing complex decision-making and performance evaluation challenges across multiple industries.

This research presents an alternative formulation of MIP aimed at reducing the overall number of misclassified vertices in interval data, extending the work mentioned above.

The structure of the study is as follows: Section 2 presents the concepts of interval data and a method for classifying them into two groups. Sueyoshi's two-stage model is also reviewed at the end of this section. Section 3 develops a reformulation of Jahanshahloo's model using the MIP approach. Section 4 provides a numerical example to demonstrate the application of the model. Finally, Section 5 concludes with a summary of the findings and remarks on the analyzed model.

## 2 Background

### 2.1 Concepts on Interval Data and Discriminant Analysis

Suppose that there are $n$ observations, $(X_j, Y_j), (j = 1, \ldots, n)$, in which $X_j = (x_{1j}, \ldots, x_{mj})$ is the $j^{\text{th}}$ input vector $(j = 1, \ldots, n)$ and $Y_j = (y_{1j}, \ldots, y_{sj})$ is the $j^{\text{th}}$ output vector $(j = 1, \ldots, n)$. Consider each observation has $k$ independent factors, shown by $(x_{1j}, \ldots, x_{mj}, y_{1j}, \ldots, y_{sj})$, and $m + s = k$. Also, assume that all observations $(X_j, Y_j), (j = 1, \ldots, n)$ are interval data, i.e., $x_{ij} \in \left[x_{ij}^L, x_{ij}^U\right], (i = 1, \ldots, m), (j = 1, \ldots, n)$ and $y_{rj} \in \left[y_{rj}^L, y_{rj}^U\right], (r = 1, \ldots, s), (j = 1, \ldots, n)$ with positive constant lower and upper bounds of the interval. These observations can be classified into $G_1$ and $G_2$ groups, with $n_1$ and $n_2$ observations. In addition, we assume $n_1 + n_2 = n$ and $G_1 \cup G_2 = G$.

Let $J_1 = \{j \mid (X_j, Y_j) \in G_1\}$ and $J_2 = \{j \mid (X_j, Y_j) \in G_2\}$. All we need to do is to find a hyperplane $(\alpha, \beta)^t(X, Y) = d$, in which $(\alpha, \beta)^t(X, Y) \geq d$ for $(X, Y) \in G_1$ and $(\alpha, \beta)^t(X, Y) \leq d - \varepsilon$ for $(X, Y) \in G_2$. Note that $(\alpha, \beta)^t(X, Y)$ is a linear discriminant function, $d$ and $d - \varepsilon$ are the first and second groups' discriminant scores, respectively. A small positive number is utilized to prevent a trivial solution, specifically where all weights are equal to zero. Consequently, we have:

$$
\begin{aligned}
&\sum_{i=1}^{m} \alpha_i x_{ij} + \sum_{r=1}^{s} \beta_r y_{rj} \geq d, \quad j \in J_1 \\
&\sum_{i=1}^{m} \alpha_i x_{ij} + \sum_{r=1}^{s} \beta_r y_{rj} \leq d - \varepsilon, \quad j \in J_2 \\
&\sum_{i=1}^{m} \alpha_i + \sum_{r=1}^{s} \beta_r = 1 - 2u, u \in \{0, 1\} \\
&\alpha_i, \beta_r, d : \text{ unrestricted}, i = 1, \ldots, m, r = 1, \ldots, s
\end{aligned}
\tag{1}
$$

We assume the vector $(\alpha, \beta)$ is normalized. Note that all $(x_{ij}, y_{rj})$ $(i = 1, \ldots, m, r = 1, \ldots, s, j = 1, \ldots, n)$ in system (1) have certain values. Otherwise, they are formed as $x_{ij} \in \left[x_{ij}^{L_{ij}}, x_{ij}^{U_{ij}}\right], (i = 1, \ldots, m, j = 1, \ldots, n)$, and $y_{rj} \in \left[y_{rj}^{L_{rj}}, y_{rj}^{U_{rj}}\right], (r = 1, \ldots, s, j = 1, \ldots, n)$. Therefore, we have:

$$
\begin{aligned}
&\sum_{i=1}^{m} \alpha_i \left[x_{ij}^{L_{ij}}, x_{ij}^{U_{ij}}\right] + \sum_{r=1}^{s} \beta_r \left[y_{rj}^{L_{rj}}, y_{rj}^{U_{rj}}\right] \geq d, j \in J_1 \\
&\sum_{i=1}^{m} \alpha_i \left[x_{ij}^{L_{ij}}, x_{ij}^{U_{ij}}\right] + \sum_{r=1}^{s} \beta_r \left[y_{rj}^{L_{rj}}, y_{rj}^{U_{rj}}\right] \leq d - \varepsilon, \quad j \in J_2 \\
&\sum_{i=1}^{m} \alpha_i + \sum_{r=1}^{s} \beta_r = 1 - 2u, u \in \{0, 1\} \\
&\alpha_i, \beta_r, d : \text{ unrestricted}, i = 1, \ldots, m, r = 1, \ldots, s
\end{aligned}
\tag{2}
$$

The satisfaction of the preceding system is evident when the conditions of the foregoing system (3) are met. Let us first define

$$\Gamma_j^- = \{t_{ij} \mid t_{ij} \in \{L_{ij}, U_{ij}\}, i = 1, \ldots, m\}$$

$$\Gamma_j^+ = \{t_{rj} \mid t_{rj} \in \{L_{rj}, U_{rj}\}, r = 1, \ldots, s\}$$

Therefore, system (2) can be changed to

$$\sum_{i=1}^{m} \alpha_i x_{ij}^{t_{ij}} + \sum_{r=1}^{s} \beta_r y_{rj}^{t_{rj}} \geq d, \quad j \in J_1, t_{ij} \in \Gamma_j^-, t_{rj} \in \Gamma_j^+$$

$$\sum_{i=1}^{m} \alpha_i x_{ij}^{t_{ij}} + \sum_{r=1}^{s} \beta_r y_{rj}^{t_{rj}} \leq d - \varepsilon, \quad j \in J_2, t_{ij} \in \Gamma_j^-, t_{rj} \in \Gamma_j^+ \tag{3}$$

$$\sum_{i=1}^{m} \alpha_i + \sum_{r=1}^{s} \beta_r = 1 - 2u, u \in \{0, 1\}$$

$$\alpha_i, \beta_r, d : \text{ unrestricted}, i = 1, \ldots, m, r = 1, \ldots, s$$

## 2.2 Classification and Overlap Handling of Interval Data

The primary procedures for identifying overlaps in interval data are as follows:
**Stage 1.** Classifying and Identifying Overlap
The identification of overlap concerning system (3) is shown as follows:

$$\text{Min } \phi = \sum_{j \in G_1} \sum_{(t_{1j}; \ldots, t_{mj}) \in \Gamma^-} s_{1j}^{+t_{1j}, \ldots, t_{mj}} + \sum_{j \in G_2} \sum_{(t_{1j}; \ldots, t_{sj}) \in \Gamma^+} s_{1j}^{-t_{1j}, \ldots, t_{sj}}$$

$$\text{s.t} \quad \sum_{i=1}^{m} \alpha_i x_{ij}^{t_{ij}} + \sum_{r=1}^{s} \beta_r y_{rj}^{t_{rj}} + s_{1j}^{+t_{1j}, \ldots, t_{mj}} - s_{1j}^{-t_{1j}, \ldots, t_{sj}} = d, j \in J_1; (t_{1j}, \ldots, t_{mj}) \in \Gamma_j^-, (t_{1j}, \ldots, t_{sj}) \in \Gamma_j^+$$

$$\sum_{i=1}^{m} \alpha_i x_{ij}^{t_{ij}} + \sum_{r=1}^{s} \beta_r y_{rj}^{t_{rj}} + s_{2j}^{+t_{1j}, \ldots, t_{sj}} - s_{2j}^{-t_{1j}, \ldots t_{mj}} = d - \varepsilon, j \in J_2; (t_{1j}, \ldots, t_{mj}) \in \Gamma_j^-, (t_{1j}, \ldots, t_{sj}) \in \Gamma_j^+ \tag{4}$$

$$s_{1j}^{+1_j, \ldots, t_{mj}}, s_{1j}^{-t_j, \ldots, t_{sj}}, s_{2j}, s_{2j}^{-t_{1j}, t_{sj}} \geq 0, \quad j = 1, \ldots, n, (t_1, \ldots, t_m) \in \Gamma_j^-, (t_1, \ldots, t_s) \in \Gamma_j^+$$

$$\sum_{i=1}^{m} \alpha_i + \sum_{r=1}^{s} \beta_r = 1 - 2u, \quad u \in \{0, 1\}$$

$$\alpha_i, \beta_r, d : \text{ unrestricted}, i = 1, \ldots, m, r = 1, \ldots, s$$

In model (4), $s_{1j}^{+t_{1j}, \ldots, t_{mj}}$ and $s_{1j}^{-t_{1j}, \ldots f_{sj}}$ are respectively positive and negative deviations of the linear discriminant function $\sum_{i=1}^{m} \alpha_i x_{ij} + \sum_{r=1}^{s} \beta_r y_{rj}$ from the discriminant score $(d)$ in $G_1$. The positive deviation $(s_{1j}^{+t_{1j}, \ldots, t_{mj}} > 0, j \in G_1, (t_{1j}, \ldots, t_{mj}) \in \Gamma_j^-)$ is utilized to minimize the incorrect classification of the $j^{\text{th}}$ observation in $G_1$. Simultaneously, the negative deviation $\left(s_{1j}^{-t_{1j}, \ldots, t_{sj}} > 0, j \in G_1, (t_{1j}, \ldots, t_{sj}) \in \Gamma_j^+\right)$ indicated the correct classification of the $j^{\text{th}}$ observation in $G_1$. Furthermore, $s_{2j}^{+t_{1j}, \ldots f_{mj}}$ and $s_{2j}^{-t_{1j}, \ldots, t_{sj}}$ are respectively positive and negative deviations of the same linear discriminant function $\sum_{i=1}^{m} \alpha_i x_{ij} + \sum_{r=1}^{s} \beta_r y_{rj}$ from the discriminant score $(d - \varepsilon)$ in $G_2$. The positive number is employed to circumvent an evident solution (all weights equal to zero). In the present case, the negative deviation $\left(s_{2j}^{-t_{1j}, \ldots, t_{sj}}, j \in G_2, (t_{1j}, \ldots, t_{sj}) \in \Gamma_j^+\right)$ indicates an incorrect group classification, while the positive deviation $\left(s_{2j}^{+t_{1j}, \ldots, t_{mj}}, j \in G_1, (t_{1j}, \ldots, t_{mj}) \in \Gamma_j^-\right)$ shows a correct one.

Here, let $\alpha^*, \beta^*$ and $d^*$ be the above model optimal solutions. Accordingly, a new observation,$(X_p, Y_p) \in \left(\left[X_p^{L_p}, X_p^{U_p}\right], \left[Y_p^{L_p}, Y_p^{U_p}\right]\right)$ can be classified as follows:

$$(1) \text{ If } \forall t_{ip} \in [L_{ip}, U_{ip}] \wedge \forall t_{rp} \in [L_{rp}, U_{rp}] \text{ and } \sum_{i=1}^{m} \alpha_i^* x_{ip}^{t_{ip}} + \sum_{r=1}^{s} \beta^* y_{rp}^{t_{rp}} \geq d^* \quad \text{then } p \in J_1$$

$$(2) \text{ If } \forall t_{ip} \in [L_{ip}, U_{ip}] \wedge \forall t_{rp} \in [L_{rp}, U_{rp}] \text{ and } \sum_{i=1}^{m} \alpha^*_i x_{ip}^{t_{ip}} + \sum_{r=1}^{s} \beta^*_r y_{rp}^{t_{rp}} \leq d^* - \varepsilon \quad \text{then } p \in J_2 \tag{5}$$

$$(3) \text{ If } \left(\exists t_{ip} \in [L_{ip}, U_{ip}] \vee \exists t_{rp} \in [L_{rp}, U_{rp}] \text{ and } \sum_{i=1}^{m} \alpha_i^* x_{ip}^{t_{ip}} + \sum_{r=1}^{s} \beta_r^* y_{rp}^{t_{rp}} < d^*\right)$$

$$\text{and } \left(\exists t'_{ip} \in [L_{ip}, U_{ip}] \vee \exists t'_{ip} \in [L_{rp}, U_{rp}] \text{ and } \sum_{i=1}^{m} \alpha_i^* x_{ip}^{t'_{ip}} + \sum_{r=1}^{s} \beta_r^* y_{rp}^{t'_{ip}} > d^*\right)$$

then part of the new observation belongs to $G_1$ and its rest belongs to $G_2$.

**Theorem 1.** The model (4) always has bounded optimal solution.

Refer to the study [4] for the proof.

**Theorem 2.** Assume that $C(G_1)$ and $C(G_2)$ be the convex hulls of $G_1$ and $G_2$. Accordingly, $C(G_1) \cap C(G_2) = \phi$, if $\phi^* = 0$.

Refer to the study [4] for the proof.

**Stage 2.** Identifying Overlap Using the Monte Carlo Approach

Now, we describe a technique for computing overlap with the use of the Hit and Miss Monte Carlo algorithm [3]. This method quantifies the area within the overlapping zone for any interval data. Consider $P^{\text{th}}$ observation, $(X_p, Y_p) \in \left( \left[ X_p^{L_p}, X_p^{U_p} \right], \left[ Y_p^{L_p}, Y_p^{U_p} \right] \right)$, as an interval data satisfying part (3) of (5). We consider $x_{ip} \in \left[ x_{ip}^{L_{ip}}, x_{ip}^{U_{ip}} \right], (i = 1, \ldots, m)$ and $y_{rp} \in \left[ y_{rp}^{L_{rp}}, y_{rp}^{U_{rp}} \right], (r = 1, \ldots, s)$. Accordingly, the whole region measure concerning the $P^{\text{th}}$ observation is $\prod_{i=1}^{m} \left( x_{ip}^{U_{ip}} - x_{ip}^{L_{ip}} \right) \prod_{r=1}^{s} \left( y_{rp}^{U_{rp}} - y_{rp}^{L_{rp}} \right)$. A vector is generated by generating random numbers as $(\bar{X}_p, \bar{Y}_p) = (\bar{x}_{1p}, \ldots, \bar{x}_{mp}, \bar{y}_{1p}, \ldots, \bar{y}_{sp})$, where for each $i(i = 1, \ldots, m), \bar{x}_{ip} \sim U\left( x_{ip}^{L_{ip}}, x_{ip}^{U_{ip}} \right)$, indicates that $\bar{x}_{ip}$ is distributed uniformly on $\left( x_{ip}^{L_{ip}}, x_{ip}^{U_{ip}} \right)$, that is, $\bar{x}_{ip} = x_{ip}^{L_{ip}} + \left( x_{ip}^{L_{ip}} - x_{ip}^{U_{ip}} \right)$ where $u_i \sim U(0,1)$. Similarly, for each $r, (r = 1, \ldots, s), \bar{y}_{rp} \sim U\left( y_{rp}^{L_{rp}}, y_{rp}^{U_{rp}} \right)$, indicates that $\bar{y}_{rp}$ is distributed uniformly on $\left( y_{rp}^{L_{rp}}, y_{rp}^{U_{rp}} \right)$, that is, $\bar{y}_{rp} = y_{rp}^{L_{rp}} + \left( y_{rp}^{L_{rp}}, y_{rp}^{U_{rp}} \right)$ where, $u_r \sim U(0,1)$. Here, consider the $P^{\text{th}}$ observation $(\bar{X}_p, \bar{Y}_p) = (\bar{x}_{1p}, \ldots, \bar{x}_{mp}, \bar{y}_{1p}, \ldots, \bar{y}_{sp})$.

(a) If $\sum_{i=1}^{m} \alpha_i^* \bar{x}_{ip} + \sum_{r=1}^{s} \beta_r^* \bar{y}_{rp} \leq d^*$, then the $P^{\text{th}}$ observation $(\bar{X}_p, \bar{Y}_p)$ is in the overlapped region of $G_1$ (6)

(b) If $\sum_{i=1}^{m} \alpha_i^* x_{ip} + \sum_{r=1}^{s} \beta_r^* y_{rp} > d^*$, then the $P^{\text{th}}$ observation $(\bar{X}_p, \bar{Y}_p)$ is in the overlapped region of $G_2$ (7)

Using inequalities (6) and (7) and counting the number of Hits, we can estimate the measure of the overlapped region of the $P^{\text{th}}$ observation with the relation below:

$$V_P = \frac{N_H}{N} \tag{8}$$

where, $N_H$ indicates the vector number $(\bar{X}_p, \bar{Y}_p)$ satisfying conditions (6) or (7), and $N$ stands for vector number.

## 2.3 MIP Method of Extended DEA/DA

Sueyoshi [9] introduced a novel two-stage MIP formulation as an extension of the extended DEA/DA model aimed at decreasing the overall count of misclassified observations:

**Stage 1:** Classifying and Identifying Overlap

The initial phase of extended DEA/DA is redefined as follows:

$$
\begin{aligned}
&\text{Min } s \\
&\text{s.t. } \sum_{i=1}^{k} \left( \lambda_i^+ - \lambda_i^- \right) z_{ij} - d + s \geq 0, \quad j \in J_1 \\
&\sum_{i=1}^{k} \left( \lambda_i^+ - \lambda_i^- \right) z_{ij} - d - s \leq 0, \quad j \in J_2 \\
&\sum_{i=1}^{k} \left( \lambda_i^+ + \lambda_i^- \right) = 1, \\
&\text{NLC}: \xi_i^+ \geq \lambda_i^+ \geq \varepsilon \xi_i^+ \quad \& \quad \xi_i^- \geq \lambda_i^- \geq \varepsilon \xi_i^-, \quad i = 1, \ldots, k \\
&\xi_i^+ + \xi_i^- \leq 1, \quad i = 1, \ldots, k \\
&\text{NZC}: \sum_{i=1}^{k} \left( \xi_i^+ + \xi_i^- \right) = k, \quad \xi_i^+, \xi_i^- \in \{0, 1\} \\
&d, s: \text{ unrestricted, all other variables } \geq 0
\end{aligned}
\tag{9}
$$

where, $Z = (X, Y)$.

Now, let's consider that $\lambda_i^* \left( = \lambda_i^{+*} - \lambda_i^{-*} \right), d^*$ and $s^*$ to be model (9) optimal solution.

(a) If $s^* \leq 0$ (no overlap), the MIP computation is terminated since the whole observation set is classified obviously into $G_1$ and $G_2$ by $\sum_{i=1}^{k} \lambda_i^* Z_i = d^*$.

(b) If $s^* > 0$ (an overlap), the subsequent phase can commence.

**Stage 2:** Handling Overlap:

Assume that:

$$C_1 = \left\{ j \in J_1 \mid \sum_{i=1}^{k} \lambda_i^* z_{ij} > d^* + s^* \right\}, C_2 = \left\{ j \in J_2 \mid \sum_{i=1}^{k} \lambda_i^* z_{ij} < d^* - s^* \right\} \qquad (10)$$

and

$$D_1 = J_1 - C_1 \qquad (11)$$

$$D_2 = J_2 - C_2 \qquad (12)$$

Since the observations' group membership in the overlap $\{(X_j, Y_j) \mid j \in D_1 \cup D_2\}$ is unknown, we need to reformulate stage two as follows:

$$
\begin{aligned}
&\text{Min } \sum_{j \in D_1} y_j + \sum_{j \in D_2} y_j \\
&\text{s.t. } \sum_{i=1}^{k} \left( \lambda_i^+ - \lambda_i^- \right) z_{ij} - c + M y_j \geq 0, j \in D_1 \\
&\sum_{i=1}^{k} \left( \lambda_i^+ - \lambda_i^- \right) z_{ij} - c - M y_j \leq 0 - \varepsilon, j \in D_2 \\
&\sum_{i=1}^{k} \left( \lambda_i^+ + \lambda_i^- \right) = 1 \\
&\text{NLC} : \xi_i^+ \geq \lambda_i^+ \geq \varepsilon \xi_i^+ \quad \& \quad \xi_i^- \geq \lambda_i^- \geq \varepsilon \xi_i^-, i = 1, \ldots, k \\
&\xi_i^+ + \xi_i^- \leq 1, i = 1, \ldots, k \\
&\text{NZC} : \sum_{i=1}^{k} \left( \xi_i^+ + \xi_i^- \right) = k, \xi_i^+, \xi_i^- \in \{0, 1\} \\
&c : \text{unrestrictes, all other variables } \geq 0
\end{aligned} \qquad (13)
$$

where, $M$ represents a specified large number. A key characteristic of stage two is that its goal function seeks to minimize the quantity of misclassified observations. The binary variable $y_j$ signifies the presence of such misclassifications. The classification of a recently sampled observation $Z_r = (z_{1r}, \ldots, z_{kr})$, is as follows:

First, we verify the below states.

(a-1) If $\sum_{i=1}^{k} \lambda_i^* z_{ir} > d^* + s^*$, then the observation belongs to $G_1$.

(b-1) If $\sum_{i=1}^{k} \lambda_i^* z_{ir} < d^* - s^*$, then the observation belongs to $G_2$.

(c-1) If $d^* - s^* \leq \sum_{i=1}^{k} \lambda_i^* z_{ir} \leq d^* + s^*$, the observation group membership is unknown because it belongs to the overlapped region, then we verify two cases below:

(a-2) If $\sum_{i=1}^{k} \lambda_i^* z_{ir} > c^*$, then the observation belongs to $G_1$.

(b-2) If $\sum_{i=1}^{k} \lambda_i^* z_{ir} \leq c^* - \varepsilon$, then the observation belongs to $G_2$.

## 3   Modified MIP Approach

With regard to presented models (9) and (13) above, it can be seen that the sets of constraints $\xi_i^+ + \xi_i^- \leq 1, (i = 1, \ldots, k)$ and $\sum_{i=1}^{k} \left( \xi_i^+ + \xi_i^- \right) = k$ are equivalent to the constraints $\xi_i^+ + \xi_i^- = 1, i = 1, \ldots, k$.

Therefore, two models (9) and (13) reformulated as follows, respectively:

$$
\begin{aligned}
&\text{Min } s \\
&\text{s.t. } \sum_{i=1}^{k} \left( \lambda_i^+ - \lambda_i^- \right) z_{ij} - d + s \geq 0, \quad j \in J_1 \\
&\sum_{i=1}^{k} \left( \lambda_i^+ - \lambda_i^- \right) z_{ij} - d - s \geq 0, \quad j \in J_2 \\
&\sum_{i=1}^{k} \left( \lambda_i^+ + \lambda_i^- \right) = 1 - 2u, \\
&\text{NLC} : \xi_i^+ \geq \lambda_i^+ \geq \varepsilon \xi_i^+ \quad \& \quad \xi_i^- \geq \lambda_i^- \geq \varepsilon \xi_i^-, i = 1, \ldots, k \\
&\xi_i^+ + \xi_i^- = 1, \quad i = 1, \ldots, k, \quad \xi_i^+, \xi_i^-, u \in \{0, 1\} \\
&d, s : \text{unrestricted, all other variables} \geq 0
\end{aligned}
$$

and

$$
\begin{aligned}
&\text{Min } \sum_{j \in D_1} y_j + \sum_{j \in D_2} y_j \\
&\text{s.t. } \sum_{i=1}^{k} \left( \lambda_i^+ - \lambda_i^- \right) z_{ij} - c + M y_j \geq 0, \quad j \in D_1 \\
&\sum_{i=1}^{k} \left( \lambda_i^+ - \lambda_i^- \right) z_{ij} - c - M y_j \leq 0 - \varepsilon, \quad j \in D_2 \\
&\sum_{i=1}^{k} \left( \lambda_i^+ + \lambda_i^- \right) = 1 - 2u, \\
&\text{NLC} : \xi_i^+ \geq \lambda_i^+ \geq \varepsilon \xi_i^+ \quad \& \quad \xi_i^- \geq \lambda_i^- \geq \varepsilon \xi_i^-, \quad i = 1, \ldots, k \\
&\xi_i^+ + \xi_i^- = 1, \quad i = 1, \ldots, k, \quad \xi_i^+, \xi_i^-, u \in \{0, 1\} \\
&c : \text{unrestrictes, all other variables} \geq 0
\end{aligned}
$$

## 4 The Proposed Algorithm

Consider a decisional case with two Groups ($G_1$ and $G_2$) of interval observations which alternatively have $n_1$ and $n_2$ observations. The number of both groups' observations reaches $n$ and every observation is characterized by $m$ independent input factors and $s$ independent output factors, denoted by $(x_{ij}, y_{rj})$ for $i = 1, \ldots, m, r = 1, \ldots, s$ and $j = 1, \ldots, n$.

Mathematically, the MIP approach for the interval data has the following two-stage computational process.

1: Classifying and Identifying Overlap.

2: Handling Overlap.

### 4.1 Classifying and Identifying Overlap

To classify the observations and identify the presence or absence of the overlapped region in between them, we solve the following Mathematic Programming (MP):

$$\text{Min } s$$

$$\text{s.t.} \sum_{i=1}^{m} \alpha_i x_{ij}^{t_{ij}} + \sum_{r=1}^{s} \beta_r y_{rj}^{t_{rj}} + s \geq d, \quad j \in J_1, t_{ij} \in \Gamma_j^-, t_{rj} \in \Gamma_j^+$$

$$\sum_{i=1}^{m} \alpha_i x_{ij}^{t_{ij}} + \sum_{r=1}^{s} \beta_r y_{rj}^{t_{rj}} - s \leq d, \quad j \in J_2, t_{ij} \in \Gamma_j^-, t_{rj} \in \Gamma_j^+ \quad (14)$$

$$\sum_{i=1}^{m} \alpha_i + \sum_{r=1}^{s} \beta_r = 1 - 2u, \quad u \in \{0, 1\}$$

$$\alpha_1, \beta_r, d, s : \text{urestricted}, \quad i = 1, \ldots, m, r = 1, \ldots, s$$

Now, let $\alpha^*, \beta^*, d^*$ and $s^*$ be the optimal solution of the previous MP.

(a) If $s^* \leq 0$ (no overlap), the MIP computation is stopped since the observations are classified thoroughly into $G_1$ or $G_2$ by $\alpha^* X + \beta^* Y = d^*$.

(b) If $s^* > 0$ (an overlap), the following step starts.

### 4.2 Handling Overlap

At this stage, it is obvious that all or some vertices of some interval observations belong to the overlapped region. To handle the classification of such observations, we first define the following sets, $V_1$ and $V_2$, to be the sets of the vertices of the observations belonging to $G_1$ and $G_2$, respectively.

$$V_1 = \left\{ \left( x_{ij}^{t_{ij}}, y_{rj}^{t_{rj}} \right), \mid j \in J_1, t_{ij} \in \Gamma_j^-, t_{rj} \in \Gamma_j^+ \right\}$$

$$V_2 = \left\{ \left( x_{ij}^{t_{ij}}, y_{rj}^{t_{rj}} \right), \mid j \in J_2, t_{ij} \in \Gamma_j^-, t_{rj} \in \Gamma_j^+ \right\} \quad (15)$$

Also, suppose that

$$K_1 = \left\{ \left( x_{ij}^{t_{ij}}, y_{rj}^{t_{rj}} \right) \in V_1 \mid \sum_{i=1}^{M} \alpha_i^* x_{ij}^{t_{ij}} + \sum_{r=1}^{s} \beta_r^* y_{rj}^{t_{rj}} > d^* + s^*, \quad j \in J_1 \right\}$$

$$K_2 = \left\{ \left( x_{ij}^{t_{ij}}, y_{rj}^{t_{rj}} \right) \in V_2 \mid \sum_{i=1}^{m} \alpha_i^* x_{ij}^{t_{ij}} + \sum_{r=1}^{s} \beta_r^* y_{rj}^{t_{rj}} < d^* - s^*, \quad j \in J_2 \right\} \quad (16)$$

to be the sets of vertices that are correctly classified into $G_1$ or $G_2$. Now, assume that:

$$Q_1 = \{j \mid (X_j, Y_j) \in V_1 - K_1\}, Q_2 = \{j \mid (X_j, Y_j) \in V_2 - K_2\} \quad (17)$$

Thus, the HO stage is formulated as follows:

$$\text{Min } \sum_{j \in Q_1} y_j + \sum_{j \in Q_2} y_j$$

$$\text{s.t} \sum_{i=1}^{m} \alpha_i x_{ij}^{t_{ij}} + \sum_{r=1}^{s} \beta_r y_{rj}^{t_{rj}} \geq c - M y_j, \quad j \in Q_1, t_{ij} \in \Gamma_j^-, t_{rj} \in \Gamma_j^+$$

$$\sum_{i=1}^{m} \alpha_i x_{ij}^{t_{ij}} + \sum_{r=1}^{s} \beta_r y_{rj}^{t_{rj}} \leq (c - \varepsilon) + M y_j, \quad j \in Q_2, t_{ij} \in \Gamma_j^-, t_{rj} \in \Gamma_j^+ \quad (18)$$

$$\sum_{i=1}^{m} \alpha_i + \sum_{r=1}^{s} \beta_r = 1 - 2u, \quad u \in \{0, 1\}$$

$$\alpha_i, \beta_r, c, d : \text{unrestricted}, i = 1, \ldots, m, r = 1, y_j \in \{0, 1\}, j \in Q_1, j \in Q_2, \text{ all other variables } \geq 0$$

A significant characteristic of the model (18) is its ability to minimize the quantity of misclassified vertices. The binary variable $y_j$ signifies the misclassified vertices presence.

Subsequently, consider the newly Sampled $P^{\text{th}}$ observation $(X_p, Y_p)$ with $X_p \in \left[X_p^{L_p}, X_p^{U_p}\right]$ and $Y_p \in \left[Y_p^{L_p}, Y_p^{U_p}\right]$. Going through stage one, we assume to have $\alpha^*, \beta^*, d^*$ and $s^*$, as the optimal solution.

$$
\begin{aligned}
&(1) \text{ If } \forall\, (t_{ip}, t_{rp}) \in \left(\Gamma_p^-, \Gamma_p^+\right), \text{ and } \sum_{i=1}^{m} \alpha_i^* x_{ij}^{t_{ip}} + \sum_{r=1}^{s} \beta_r^* y_{rp}^{t_{rp}} > c^* \quad \text{then } (X_p, Y_p) \in G_1 \\
&(2) \text{ If } \forall\, (t_{ip}, t_{rp}) \in \left(\Gamma_p^-, \Gamma_p^+\right), \text{ and } \sum_{i=1}^{m} \alpha_i^* x_{ij}^{t_{ip}} + \sum_{r=1}^{s} \beta_r^* y_{rp}^{t_{rp}} < c^* \quad \text{then } (X_p, Y_p) \in G_2
\end{aligned}
\tag{19}
$$

Otherwise, all or parts of the $P^{\text{th}}$ observation pertain to the overlapped region. Therefore, stage two starts to decide whether this observation would pertain to $G_1$ or $G_2$.

In this case,

$$
\begin{aligned}
&(1) \text{ If } \forall\, (t_{ip}, t_{rp}) \in \left(\Gamma_p^-, \Gamma_p^+\right), \text{ and } \sum_{i=1}^{m} \alpha_i^* x_{ij}^{t_{ip}} + \sum_{r=1}^{s} \beta_r^* y_{rp}^{t_{rp}} > c^* \quad \text{then } (X_p, Y_p) \in G_1 \\
&(2) \text{ If } \forall\, (t_{ip}, t_{rp}) \in \left(\Gamma_p^-, \Gamma_p^+\right), \text{ and } \sum_{i=1}^{m} \alpha_i^* x_{ij}^{t_{ip}} + \sum_{r=1}^{s} \beta_r^* y_{rp}^{t_{rp}} < c^* \quad \text{then } (X_p, Y_p) \in G_2
\end{aligned}
\tag{20}
$$

Here, since $\varepsilon$ is a very small number, we withdraw it.

(3) Otherwise, some vertices of the arbitrary interval data $(X_p, Y_p)$ pertain to $G_1$ and the rest pertain to $G_2$.

To overcome this, we assume $A_1$ to be part of $(X_p, Y_p)$ that pertains to $G_1$ and $A_2$ part of $(X_p, Y_p)$ that belongs to $G_2$.

Now, we use the Monte Carlo method to measure how much of the observation $(X_p, Y_p)$ pertains to $G_1$ and $G_2$ by counting the number of Hits $(N_{H_1}$ or $N_{H_2})$ of all the arbitrary vectors $(N)$ that pertain to one of our Groups $(G_1$ or $G_2)$.

Let $V_{P_1} = \frac{N_{H_1}}{N}$ be the overlapped region that belong to $G_1$.

$$
\begin{aligned}
&(3.\ a) \text{ If } V_{P_1} \geq \frac{1}{2}, \text{ then } (X_p, Y_p) \in G_1 \text{ and } \Delta = 1 - V_{P_1} \\
&(3.\ b) \text{ If } V_{P_1} < \frac{1}{2}, \text{ then } (X_p, Y_p) \in G_2 \text{ and } \Delta = V_{P_1}
\end{aligned}
\tag{21}
$$

where, $\Delta$ indicates the error by which $(X_p, Y_p)$ can be assumed to belong to one of the Groups. In other words, we intend to measure the probability under which the interval observations $(X_p, Y_p)$, belong to either group. Also, since $N$ is a given large number, it is not practically possible for $N_{H_1}$ to reach to half of $N$, exactly. Therefore, we can withdraw the probability of $V_{p_1} < \frac{1}{2}$, and we do not consider it.

## 5  Examples

### 5.1  Example 1

Suppose $G_1$ and $G_2$ are two interval data groups with one input and output, as listed in Table 1.

**Table 1.** Input and output sets of Example 1

| Groups | $DMU_j$ | $X_j^{L_j}$ | $X_j^{U_j}$ | $Y_j^{L_j}$ | $Y_j^{U_j}$ |
|---|---|---|---|---|---|
| Group1 $(G_1)$ | $P_1$ | 4 | 5 | 4 | 5 |
| | $P_2$ | 2 | 3 | 5 | 6 |
| Group2 $(G_2)$ | $P_3$ | 1 | 2 | 3 | 4 |
| | $P_4$ | 4 | 5 | 1 | 3 |

Using model (15) for the classification of the observations $P_1, P_2, P_3$ and $P_4$, we obtain the following:

$$
\alpha^* = 0.25,\ \beta^* = 0.75,\ d^* = 3.75,\ s^* = -0.25.
$$

Since $s^*(= -0.25)$ is negative, it is obvious that the two groups are separated completely with no overlap between them. The strongly separating hyperplane is: $0.25X + 0.75Y = 3.75$, as shown in Figure 1.
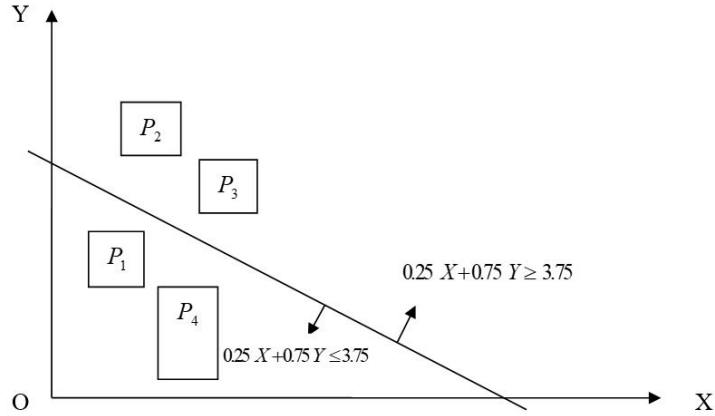
**Figure 1.** Without the presence of overlap

**Table 2.** The coordinates of the new observation of Example 1

| New Observation | $X^L$ | $X^U$ | $Y^L$ | $Y^U$ |
|---|---|---|---|---|
| $P_5$ | 2 | 3 | 1 | 2 |

Now consider the new observation $P_5$ in Table 2.

So, we have:

$$\alpha^* X^L + \beta^* Y^L = (0.25)(2) + (0.75)(1) = 1.25 < 3.75 = d^*$$
$$\alpha^* X^L + \beta^* Y^U = (0.25)(2) + (0.75)(2) = 2 < 3.75 = d^*$$
$$\alpha^* X^U + \beta^* Y^L = (0.25)(3) + (0.75)(1) = 1.5 < 3.75 = d^*$$
$$\alpha^* X^U + \beta^* Y^U = (0.25)(3) + (0.75)(2) = 2.25 < 3.75 = d^*$$

As can be seen, all the vertices belonging to $P_5$ are located below the line $0.25X + 0.75Y = 3.75$ and, therefore, belong to the group $G_2$.

### 5.2 Example 2

Consider the new observation $P_6$ that has some of its vertices in $G_1$ and the others in $G_2$, as shown in Table 3.

**Table 3.** The coordinates of the new observation of Example 2

| New Observation | $X^L$ | $X^U$ | $Y^L$ | $Y^U$ |
|---|---|---|---|---|
| $P_6$ | 6 | 7 | 2 | 4 |

Regarding the model optimal solutions and the lower and upper bounds of the independent input and output factors, we will obtain:

$$\alpha^* X^L + \beta^* Y^L = (0.25)(6) + (0.75)(2) = 3 < 3.75 = d^*$$
$$\alpha^* X^L + \beta^* Y^U = (0.25)(6) + (0.75)(4) = 4.5 > 3.75 = d^*$$
$$\alpha^* X^U + \beta^* Y^L = (0.25)(7) + (0.75)(2) = 3.25 < 3.75 = d^*$$
$$\alpha^* X^U + \beta^* Y^U = (0.25)(7) + (0.75)(4) = 4.75 > 3.75 = d^*$$

It is clear that two of the vertices belong to $G_1$ and the rest belong to $G_2$.

To handle the case, the second stage of the MIP approach to interval data starts to classify the incorrectly classified vertices.

By solving model (18) for the overlap handling of the misclassified vertices, we obtain:

$$\alpha^* = 0.25, \quad \beta^* = 0.75, \quad c^* = -3.49$$

Therefore:

$$(0.25)(6) + (0.75)(2) = 3.00 > -3.49 = c^*$$
$$(0.25)(6) + (0.75)(4) = 4.5 > -3.49 = c^*$$
$$(0.25)(7) + (0.75)(2) = 3.25 > -3.49 = c^*$$
$$(0.25)(7) + (0.75)(4) = 4.75 > -3.49 = c^*$$

So we can conclude that $P_6$ belongs to $G_1$.

### 5.3 Example 3

Consider the two-stage MIP method for classification of 10 observations where $P_1, P_2, P_3, P_4$ and $P_5$ belong to group $G_1$ and $P_6, P_7, P_8, P_9$ and $P_{10}$ belong to group $G_2$. The observations are interval data, each with two independent input factors and two independent output factors. The input and output values of the observations are listed in Table 4.

**Table 4.** Input and output sets of Example 3

| **Groups** | $DMU_j$ | $X_j^{L_j}$ | $X_j^{U_j}$ | $Y_j^{L_j}$ | $Y_j^{U_j}$ |
|---|---|---|---|---|---|
| | $P_1$ | 4 | 6 | 7 | 9 |
| | $P_2$ | 9 | 10 | 9 | 10 |
| Group1 ($G_1$) | $P_3$ | 4 | 6 | 13 | 14 |
| | $P_4$ | 8 | 11 | 12 | 13 |
| | $P_5$ | 12 | 13 | 9 | 10 |
| | $P_6$ | 2 | 3 | 4 | 5 |
| | $P_7$ | 2 | 3 | 6 | 8 |
| Group2 ($G_2$) | $P_8$ | 1 | 3 | 1 | 2 |
| | $P_9$ | 7 | 9 | 6 | 8 |
| | $P_{10}$ | 4 | 5 | 2 | 4 |

The ideal solution for classifying these observations applying model (14) is derived as follows: $\alpha^* = 0, \beta^* = 1, d^* = 7.5, s^* = 0.5$.

Since $s^*(= 0.5)$ is positive, we can conclude that an overlap exists between groups. So, at the COI stage, we will have the subsets $K_1$ consisting of all the correctly classified vertices of $G_1$, $K_2$ consisting of all the correctly classified vertices of $G_2$, $\bar{Q}_1$ consisting of all the incorrectly classified vertices of $G_1$, $\bar{Q}_2$ consisting of all the incorrectly classified vertices of $G_2$.

It is obvious that the overlap area includes all the vertices belonging to $\bar{Q}_1 \cup \bar{Q}_2$ and is restricted between the lines $Y = 7$ and $Y = 8$, as shown in Figure 2.
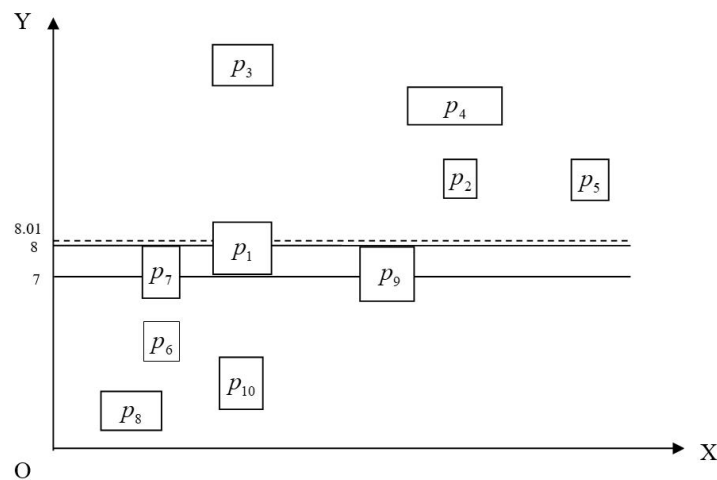


**Figure 2.** The presence of overlap

As it is visually depicted in Figure 2, $\bar{Q}_1$ and $\bar{Q}_2$ are recognized as:

$$\bar{Q}_1 = \left\{ \left( X_{p_1}^{L_1}, Y_{p_1}^{L_1} \right), \left( X_{p_1}^{L_1}, Y_{p_1}^{U_1} \right) \right\}$$

$$\bar{Q}_2 = \left\{ \left( X_{p_7}^{L_7}, Y_{p_9}^{U_7} \right), \left( X_{p_7}^{U_7}, Y_{p_7}^{U_7} \right), \left( X_{p_9}^{L_9}, Y_{p_9}^{U_9} \right), \left( X_{p_9}^{U_9}, Y_{p_9}^{U_9} \right) \right\}$$

$$K_1 = V_1 - \bar{Q}_1$$

$$K_2 = V_2 - \bar{Q}_2$$

Here, $V_1$ and $V_2$ are the sets of the whole vertices belonging to $G_1$ and $G_2$ respectively.

It is also observed from Figure 2 that the observations $P_1, P_2, P_3$ and $P_4$ thoroughly belong to $G_1$ and the observations $P_6, P_8$, and $P_{10}$ thoroughly belong to $G_2$. The observations $P_1, P_7$ and $P_9$ have parts of them in the overlap area.

Because the overlap exists, the next stage starts to handle observations $P_1, P_7$ and $P_9$. Let $y_1$ be the binary variable for enumerating the misclassified vertices of $P_1$ in $\bar{Q}_1$, $y_2$ be the binary variable for enumerating the misclassified vertices of $P_7$ in $\bar{Q}_2$ and $y_3$ be the binary variable for enumerating the misclassified vertices of $P_9$ in $\bar{Q}_2$. Going through the model (18), we obtain the following optimal solution for the HO stage:

$$\alpha^* = 0, \beta^* = 1, c^* = 8.01, y_1^* = 1, y_2^* = 0, y_3^* = 0$$

Therefore, the discriminant function defined at the next stage is formulated as $Y = 8.01$ which makes the misclassified vertices of $P_7$ and $P_9$ to belong to $V_2$ (The set of vertices belonging to $G_2$). Also, since the observation $P_1$ still has parts of it in $G_1$ and the rest in $G_2$, we need to use the Monte Carlo method to handle it.

Let $R_1$ be the part of $P_1$ that belongs to $G_1$, $N$ be the total number of the arbitrary vectors produced by the Monte Carlo method and $N_{H_1}$ be the number of such produced vectors that belong to $G_1$. Going through the Monte Carlo method with the assumption $N = 500,000$, we obtain the following:

$$V_{R_1} = \frac{N_{H_1}}{N} = 0.495 < 0.5$$

So it is concluded that $P_1$ belongs to $G_2$, because the possibility under which $P_1$ belongs to $G_2 (= 0.505)$ is more than the possibility under which $P_1$ belongs to $G_1 (= 0.495)$. Also, the error considered for $P_1$ to belong to $G_2$ is equal to 0.495 .

## 6 Conclusions

This study introduced an innovative two-stage MIP framework for discriminant analysis of interval data, addressing key challenges such as misclassification and group overlap. The proposed methodology effectively minimized the total number of misclassified vertices and identifies overlap zones, providing probabilistic insights into group memberships. Through comprehensive examples, the model demonstrated its ability to classify interval data, manage overlaps, and identified structural similarities and differences within datasets. Despite its strengths, the performance of the method is influenced by the choice of parameters and M, which can lead to variability in classification outcomes. Future research should focus on standardizing parameter selection and extending the framework to accommodate data in fuzzy or other non-traditional formats. The proposed approach represents a significant advancement in interval data classification, offering a robust and adaptable tool for complex decision-making scenarios.

**Data Availability**

The data supporting our research results are included within the article.

**Conflicts of Interest**

The authors declare no conflict of interest.

**References**

[1] H. Saleh, M. Shafiee, and M. Sanji, "Modifying the interconnecting activities through an adjusted dynamic DEA model: A slacks-based measure approach," *J. Appl. Res. Ind. Eng.*, vol. 7, no. 3, pp. 287–300, 2020. https://doi.org/10.22105/jarie.2020.229872.1163

[2] B. Ebrahimi, S. Fischer, and M. Milovancevic, "An improved mixed-integer DEA approach to determine the most efficient unit," *Optimality*, vol. 1, no. 2, pp. 224–231, 2024. https://doi.org/10.22105/opt.v1i2.59

[3] K. Kumar Mohanta and D. S. Sharanappa, "Neutrosophic data envelopment analysis: A comprehensive review and current trends," *Optimality*, vol. 1, no. 1, pp. 10–22, 2024. https://doi.org/10.22105/opt.v1i1.19

[4] M. Yazdani and H. Kazemipor, "Evaluating the efficiency and effectiveness of the organization by combining the BSC and DEA evaluation systems," *Financial Banking Strategic Stud.*, vol. 1, no. 4, pp. 272–280, 2024. https://doi.org/10.22105/fbs.2024.457582.1085

[5] R. S. Keyser, "A fuzzy goal programming approach for solving multi-objective minimum cost flow problems with possibilistic coefficients," *Uncertainty Discourse Appl.*, vol. 1, no. 1, pp. 29–40, 2024.

[6] T. Sueyoshi, "Dea-discriminant analysis in the view of goal programming," *Eur. J. Oper. Res.*, vol. 115, no. 3, pp. 564–582, 1999. https://doi.org/10.1016/s0377-2217(98)00014-9

[7] T. Sueyoshi and Y. Kirihara, "Efficiency measurement and strategic classification of Japanese banking institutions," *Int. J. Syst. Sci.*, vol. 29, no. 11, pp. 1249–1263, 1998. https://doi.org/10.1080/00207729808929613

[8] T. Sueyoshi, "Extended DEA-discriminant analysis," *Eur. J. Oper. Res.*, vol. 131, no. 2, pp. 324–351, 2001. https://doi.org/10.1016/s0377-2217(00)00054-0

[9] T. Sueyoshi, "Mixed integer programming approach of extended DEA-discriminant analysis," *Eur. J. Oper. Res.*, vol. 152, no. 1, pp. 45–55, 2004. https://doi.org/10.1016/S0377-2217(02)00657-4

[10] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *Eur. J. Oper. Res.*, vol. 2, no. 6, pp. 429–444, 1978. https://doi.org/10.1016/0377-2217(78)90138-8

[11] W. W. Cooper, L. M. Seiford, and J. Zhu, "Data envelopment analysis: History, models, and interpretations," in *Handbook on Data Envelopment Analysis*. Springer, Boston, 2011, pp. 1–39. https://doi.org/10.1007/978-1-4419-6151-8_1

[12] G. R. Jahanshahloo, F. H. Lotfi, F. R. Balf, and H. Z. Rezai, "Discriminant analysis of interval data using Monte Carlo method in assessment of overlap," *Appl. Math. Comput.*, vol. 191, no. 2, pp. 521–532, 2007. https://doi.org/10.1016/j.amc.2007.02.113

[13] G. R. Jahanshahloo, F. H. Lotfi, H. Z. Rezai, and F. R. Balf, "Using Monte Carlo method for ranking efficient DMUs," *Appl. Math. Comput.*, vol. 162, no. 1, pp. 371–379, 2005. https://doi.org/10.1016/j.amc.2003.12.139

[14] M. T. Y. Shikhzahedi, A. Amirteimoori, S. Kordrostami, and S. A. Edalatpanah, "Inverse data envelopment analysis model to improve efficiency by increasing outputs," *Decis. Making Appl. Manage. Eng.*, vol. 7, no. 2, pp. 1–14, 2024.

[15] A. P. Yekta, M. J. S. Noveiri, M. Maghbouli, and S. A. Edalatpanah, "Dynamic DEA with common weights: A case study of Iranian airlines," *J. Uncert. Sys.*, vol. 17, no. 2, 2024. https://doi.org/10.1142/s1752890924400014

[16] S. Edalatpanah, F. H. Lotfi, K. Kerstens, and P. Wanke, *Analytical Decision Making and Data Envelopment Analysis*. Springer, Singapore, 2024.

[17] H. Shi, X. Liu, and S. Chen, "Decision-making conflict measurement of old neighborhoods renovation based on mixed integer programming DEA-discriminant analysis (MIP DEA–DA) models," *Buildings*, vol. 14, no. 3, p. 785, 2024. https://doi.org/10.3390/buildings14030785

[18] Y. Pan, C. Zhang, C. Lee, and S. Lv, "Environmental performance evaluation of electric enterprises during a power crisis: Evidence from DEA methods and AI prediction algorithms," *Energy Econ.*, vol. 130, p. 107285, 2024. https://doi.org/10.1016/j.eneco.2023.107285

[19] M. Tavassoli and M. Ghandehari, "Classification and forecasting of sustainable-resilience suppliers via developing a novel fuzzy MIP model and DEA in the presence of zero data," *Oper. Manage. Res.*, pp. 1–26, 2023. https://doi.org/10.1007/s12063-023-00401-z

[20] I. Surjandari, N. R. Maulina, and C. Bahri, "Efficiency analysis of halal certification bodies in Indonesia: A hybrid data envelopment analysis and machine learning approach," *Qual. Quant.*, pp. 1–15, 2024. https://doi.org/10.1007/s11135-024-02006-5

[21] P. Nguyen, T. Nguyen, C. Wang, M. Vu, L. A. T. Nguyen, H. Pham, M. Thi Pham, and H. Q. Le, "Linking investment decisions-based on firm performance and open innovation practices in Vietnam's wire and cable market using data envelopment analysis models," *J. Open Innov. Technol. Marke Complexity*, vol. 9, no. 2, p. 100080, 2023. https://doi.org/10.1016/j.joitmc.2023.100080