Engineering Applications of Artificial Intelligence 132 (2024) 107914

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Resource allocation in ordinal classification problems: A prescriptive framework utilizing machine learning and mathematical programming

Lior Rabkin, Ilan Cohen, Gonen Singer *

Bar-Ilan University, Faculty of Engineering, Ramat-Gan, Israel

ARTICLE INFO

Keywords:
Ordinal classification
Ordinal neural network
Ordinal decision tree-based model
Mathematical programming
Cost minimization
Resource constraints
Resource allocation

ABSTRACT

Ordinal classification tasks that require the allocation of limited resources are prevalent in various real-world scenarios. Examples include assessing disease severity in the context of medical resource allocation and categorizing the quality of machines as good, medium, or bad to schedule maintenance treatment within capacity constraints. We propose a comprehensive analytic framework for scenarios that, in addition to including ordinal classification problems, also have constraints on the number of classified samples of classes due to resource limitations. The framework uses a probability matrix generated by a trained ordinal classifier as the input for an optimization model with a minimum misclassification cost objective and resource allocation constraints. We illustrated the equivalence between the formulation of the resource allocation problem into samples and the transportation problem, enabling the utilization of established transportation heuristics for our solution. To demonstrate the effectiveness and applicability of the framework, we applied it with various ordinal machine-learning models to both tabular data and image datasets. The proposed framework performs significantly better than the alternative common approach of using non-ordinal classifiers, achieving an average cost reduction of 1% with ordinal decision tree-based models and 4.4% with ordinal neural networks. Our results show that the proposed framework can provide an effective limited-resource allocation for ordinal classification problems. Our code is available at https://github.com/liorRabkin/hybrid-cost-sensitive-ml-optimization.

1. Introduction

Ordinal classification problems are commonly handled as multi-class classification scenarios, where the target class displays a specific ordinal order. These problems are typically associated with real-world applications such as categorizing disease severity, and classification of the emergency status of a patient (Silva et al., 2017; Nabi et al., 2019).

In classification problems, the presence of resource allocation issues due to scarcity can introduce real-world constraints, thereby affecting the distribution of classified samples across different classes (Li et al., 2018; Gartner et al., 2015; Abukasis et al., 2022). Applying the classification model directly to new data (test data) that stands apart from the training dataset will not ensure conformity to the resource constraint. Commonly, in classification problems, different decision-makers may look for different target values depending on their own purposes. Below, we explore various real-life scenarios that illustrate resource constraints in ordinal classification problems:

(i) Disease severity classification: In disease severity monitoring, accurately identifying the state of the disease is crucial for adopting an appropriate treatment method (Haba et al., 2023; Singer et al., 2021). The classes represent different levels of severity, such as mild, moderate, and severe. In this scenario, a medical practitioner might possess a restricted number of monitoring devices to identify disease progression (whether a patient is in the initial or advanced stage of the illness). Ideally, allocating these devices should prioritize the identification of a substantial portion of patients in the early disease stages. However, a medical center might prioritize recognizing patients in advanced disease stages, leveraging a recognized high recovery success rate, while optimizing the allocation of limited medical resources, including surgical rooms and teams.

(ii) Likelihood of churning prediction: In customer churn prediction (Wu et al., 2022), it is important to identify the likelihood of churning to enable proactive activities aimed at reducing customer churn and enhancing retention (Akan and Verma, 2022). The classes in this problem typically lie on an ordinal scale. In this scenario, a limited number of team leaders should be allocated to high-risk-level customers for retention calls, while a limited number of employees should be allocated to unstable customers who can potentially be retained with a simple phone call.

* Corresponding author.
E-mail addresses: Lior.rabkin@biu.ac.il (L. Rabkin), Ilan-reuven.cohen@biu.ac.il (I. Cohen), Gonen.Singer@biu.ac.il (G. Singer).

https://doi.org/10.1016/j.engappai.2024.107914
Received 9 November 2023; Received in revised form 1 January 2024; Accepted 13 January 2024
Available online 23 January 2024
0952-1976/© 2024 Elsevier Ltd. All rights reserved.

(iii) Credit scores prediction: When a client requests a loan or credit from a lender, the lender needs to assess how risky it is to lend money to that client (Juraev and Rakhimberdiev, 2022). The scores of clients, which usually lie on an ordinal scale, reflect this risk (Dikkers and Rothkrantz, 2005; Wang et al., 2022). Borrowers often handle the problem of limited credit allocation, considering the prediction of the loaner's credit score (Zhu et al., 2022; Tran and Verhoeven, 2021).

(iv) Product's perceived value classification: In this task, the estimation of perceived value can be categorized into specific levels, such as low, medium, or high, representing a combination of various common emotional and functional dimensions, such as quality and value for money of the product (Sweeney and Soutar, 2001). The objective of this ordinal classification can be, for example, products allocation to limited online advertising spaces or integration into live streaming e-commerce with restricted capacity to enhance purchase intention (Zhu et al., 2023; Li et al., 2024).

In the experimental investigation section of our paper, we will use datasets representing the first two problems.

These sorts of applications typically involve misclassification costs that originate from the consequences of errors (Zhang et al., 2021; Frumosu et al., 2020). For instance, misclassifying patients with a high severity level of disease as those with a low severity level could have life-threatening consequences. However, in many cases precise misclassification cost values are unknown. Therefore, the gap between the actual and predicted classes, while considering the existing class order among classes, can be employed to reflect the misclassification errors as proposed in many research studies in the literature (Marudi et al., 2022; He, 2022). In some tasks, misclassification errors may display asymmetry. For instance, misclassifying ill patients as healthy in disease severity classification could pose a much higher risk, resulting in a different penalty than misclassifying healthy patients as ill. In these cases, when the misclassification costs are unknown, asymmetric and ordinal values reflecting the penalty of misclassification errors are proposed (Lima et al., 2020).

In this article, a decision-making framework is introduced to address the multi-class ordinal classification challenge with resource limitations, all with the goal of cost minimization. Our method merges an ordinal algorithm with an optimization model, collectively identifying the optimum resource allocation solution. Throughout the paper, various ordinal machine learning algorithms are showcased for the first phase of the framework. This approach effectively illustrates the universality of the framework, highlighting the application of various algorithms customized for tabular and image datasets, all working towards the same mission using a consistent methodology.

2. Background and literature review

This section presents the ordinal classification strategies proposed by researchers and ordinal task notation. Following that, we evaluate research studies that have utilized classification algorithms to address the challenge of constrained resource allocation.

2.1. Ordinal classification problems

In the machine learning literature, ordinal algorithms were developed to address problems in which the target value maintains an arbitrary scale, where only the relative ordering between the different values is significant, and the distances between them are irrelevant (Marudi et al., 2022; Lázaro and Figueiras-Vidal, 2023; Singer et al., 2020; Rosati et al., 2022). These ordinal algorithms were designed to tackle real-world applications, including tasks like categorizing disease severity, assessing product quality as good, medium, or bad in the industry, classifying the severity of traffic accident casualties in transportation, and predicting traffic intensities as high, moderate, or light in queuing systems (see, for example, Nabi et al. (2019), Wang et al. (2021), Yıldırım et al. (2019), Kim et al. (2023)). A comprehensive list of applications across various research areas and a proposed categorization of ordinal classification methods into distinct approaches, can be found at Gutierrez et al. (Gutiérrez et al., 2015).

Over the years, many ordinal classification methods have been proposed to address classification tasks on ordinal data. In these studies, as discussed in Section 1, misclassification costs are incurred and influenced by the distance between the actual and predicted classes. The greater the distance, the higher the misclassification cost. In the disease severity classification problem, for example, a larger distance between the predicted severity level of a patient and the actual severity level may result in less appropriate treatment for the patient, leading to a higher risk. The primary limitation of these methods arises when the cost matrix is not known and predefined, as multiple options for different cost matrices may reflect the same ordinal scale of the problem (Abukasis et al., 2022). In this problem, usually the order information is introduced into the learning process within the classification algorithms. For instance, because classes are typically arranged in ascending or descending order of quality, Marudi et al. (2022), Singer et al. (2020) suggested assigning reward or cost values to each class to capture the ordinal nature of the class variable. These values are subsequently incorporated into the learning process of ordinal classification models.

2.2. Ordinal classification notation

Now, we introduce definitions used throughout the paper. We denote a fully labeled training dataset by $D = (X, Y)$. $X \in \mathbb{R}^{|S|\times|F|}$ is a set of samples S, with a set of features F. The labels are stored in the vector $Y \in \mathbb{R}^{|S|\times|1|}$, i.e., the $Y_{s,1}$ entry represents the class a sample $x_s \in X$ belongs to, briefly denoted by y_s. We assume a set of different classes, denoted as C, such that $\forall i \in S, y_i \in C$.

Assume an ordinal cost matrix $O \in \mathbb{R}^{|C|\times|C|}$, where $O_{i,j}$ denote the penalty cost between the predicted class i and the real class j. Given that, in many classification problems, the misclassification costs are often unknown or not predefined, as discussed in Section 2.1, we adopt the approach proposed by Singer et al. (2020) and Marudi et al. (2022). They introduced a function $\upsilon(\cdot)$ that assigns distinct values to classes on the ordinal scale, considering the potential magnitude of classification errors between predicted and actual classes. Specifically, for every $c_i, c_j \in C$ where $i < j$, it holds that $\upsilon(c_i) < \upsilon(c_j)$. Thus, $O_{i,j} = O_{j,i} = |\upsilon(c_i) - \upsilon(c_j)|$. According to the objective of the misclassification cost problem, a sample $x_s \in X$ should be classified into the class i that minimizes the cost. $\mathcal{M}$ is a classification model that returns probability matrix $\hat{Y} \in (0,1)^{|S|\times|C|}$, i.e., a vector of estimated probabilities for each sample, for each possible class. The $\hat{Y}_{s,i}$ entry represents the probability that a sample $x_s \in X$ belongs to class i, $\hat{Y}_{s,i} = P(i|x_s)$. Each sample is assigned to the class that has the highest probability (that is, the maximum likelihood), i.e., $\hat{y}_s = \arg\max_i P(i|x_s)$.

2.3. Machine learning and optimization models for resource-constrained classification tasks

A number of research investigations have suggested the integration of machine learning techniques and optimization models to address classification challenges that involve resource constraints. Typically, during the machine learning classification stage, individual probabilities are produced for each sample within the dataset, indicating the likelihood of successful classification for each class. The optimization model operates to optimize an objective function while also ensuring compliance with resource constraints specified by decision-makers.

The study in Li et al. (2018) introduced a resource allocation strategy designed for categorizing workload tasks that share comparable resource needs. The primary goal was to optimize the allocation of existing resources, concurrently diminishing energy consumption by minimizing the count of operational physical machines.

Another study Pessach et al. (2020) presented a methodology that merges interpretable machine learning models with mathematical programming. The primary goal of this strategy was to assign well-matched candidates to a limited number of open positions, thereby reducing the risk of turnover. The machine learning model was used to produce success probabilities for employee-to-position matching, while the mathematical modeling considered the constraint of a limited number of available positions and the workforce's diversity across various departments.

In Zhang et al. (2020), an approach that combines classification models with an optimization model is presented to address the challenge of allocating limited resources to target customer groups with diverse characteristics in the context of precision marketing for a new product.

The work presented in Eshghali et al. (2023) suggests a fusion of a machine learning algorithm and an optimization model to enhance the productivity of hospital operating rooms, while taking into account both ellective and emergency patients. Due to the inherent randomness in emergency patient arrival, a random forest machine learning model is used to obtain the emergency patient surgery duration and arrival time, serving as input for an optimization model designed to determine the most optimal scheduling. The research outlined in Gartner et al. (2015) introduced an approach that incorporates diagnosis-related group classifications using machine learning model into a resource allocation model based on mixed-integer programming. This approach was aimed at optimizing the utilization of resources like operating rooms and beds.

While these research studies have primarily concentrated on integrating machine learning techniques and optimization models to tackle classification challenges related to resource allocation, none have considered ordinal classification problems, where the class variable exhibits a specific order. Furthermore, these studies have typically provided solutions tailored to specific applications using tabular datasets. In this paper, we introduce a comprehensive analytical framework for addressing ordinal classification problems with constraints on the number of samples classified per class due to resource limitations. We demonstrated that the formulation of the resource allocation problem into samples is equivalent to the transportation problem, allowing the use of well-known transportation heuristics to solve our problem. Additionally, we showed that the characteristics of the problem formulation enable the finding of a solution with integer variables. We illustrate the effectiveness of this framework using both tabular and image datasets, employing ordinal decision tree-based models and ordinal neural networks, respectively.

The remainder of the paper is structured as follows: In Section 3, we introduce the proposed framework designed to address ordinal classification problems with resource constraints. Section 4 is dedicated to presenting the outcomes of our numerical experiments, which assess and benchmark the proposed framework against commonly used approaches. In this section, we present the outcomes obtained by utilizing ordinal decision tree-based models on tabular data and employing ordinal neural networks for image data analysis. Finally, Section 5 offers concluding remarks to summarize and wrap up the paper.

3. Methods

3.1. Hybrid framework for a ordinal classification problem with limited resources constraints

This section introduces a Constraint-based Ordinal Classification Framework (COCF) for minimizing the costs of the resource-constrained ordinal classification problem, by utilizing both the ordinal machine learning method and optimization model. A key property of our framework is the generation of the probability matrix $\hat{Y}$ by an ordinal classifier that can be used as input for our optimization model, aiming to achieve minimum costs while satisfying the resource constraints. Fig. 1 presents the proposed framework consisting of two phases. An ordinal machine learning algorithm is trained using a training dataset in the first phase. The output of this phase, the probability matrix, is used in the second phase as input for the optimization model. Then, using the validation data, we select the best hyper-parameters for the ordinal classifier based on the costs achieved after the optimization model in the second phase. The optimization model is solved to find the classifications that meet the constraints and create the lowest cost. The test dataset goes through the machine learning phase, resulting in a probability matrix that is used as input to the optimization phase, which produces a classification for each instance as presented in Fig. 2.

In our research, in the experimental study, we employ two types of ordinal methods to demonstrate the effectiveness of our proposed framework on both tabular and image datasets, as explained in Section 3.2. A detailed description of the optimization model is described in Section 3.3.

3.2. Ordinal machine learning methods

In the experimental study, we will evaluate the effectiveness of the proposed COCF framework on both tabular and image datasets. Specifically, we will explore its performance on: (i) A tabular dataset representing the likelihood of churning prediction problem, which corresponds to the second real-life scenario in the introduction. (ii) An image dataset representing the problem of disease severity classification, corresponding to the first real-life scenario in the introduction. For such problem domains, interpretability of the ordinal classification models is vital to support human decision-making and improve understanding of resource allocation reasons. Additionally, different decision-makers, each responsible for distinct types of resources, may seek different class values. To address these challenges, we choose to utilize ordinal decision tree-based methods for the tabular dataset (Marudi et al., 2022; Singer et al., 2020, 2021; Singer and Marudi, 2020; Singer and Cohen, 2020). These methods enable the construction of interpretable tree models based on objective defined by decision-makers, as explained later in Eq. (2) (Marudi et al., 2022; Singer et al., 2020). In the case of the image dataset, we use an ordinal neural network method (Chen et al., 2019; Liu et al., 2022). This method is tailored to handle the unique characteristics of image-based ordinal classification tasks.

Ordinal decision tree-based models. In the experiments with the tabular dataset, we employed the ordinal decision tree and ordinal random forest algorithms as proposed in Marudi et al. (2022), Singer et al. (2020) in the first phase of the framework (see Fig. 1). The ordinal entropy, referred to as objective-based entropy (OBE), is used in the tree construction process, which is formulated as

$$H(\tau) = -\sum_{i=1}^{|C|} \varepsilon_i P(c_i) \log_2 P(c_i), \tag{1}$$

where

$$\varepsilon_i(c_i, \tau) = \frac{|v(c_i) - \tau(s)|^\alpha}{\sum_{j=1}^{|C|} |v(c_j) - \tau(s)|^\alpha}. \tag{2}$$

$P(c_i)$ is defined as the probability that an instance belongs to class c_i, and ε_i represents the normalized distance between the value of class c_i, $v(c_i)$, and the objective $\tau(s)$. The ordinal decision tree-based algorithms do not use an ordinal cost matrix, however considering the costs of errors from predefined objective $\tau(s)$, where s represents a statistical function calculated over the data. In Singer et al. (2020) for example, the costs reflect the deviation of the values of the classes relative to the value of the most likely class, i.e., $\tau(s) = v(c^{mode})$. $\alpha \geq 0$ is a normalization parameter that biases the distances of the classes compared to the objective $\tau(s)$. The probability matrix $\hat{Y}$ obtained by the ordinal decision tree-based methods, used as input for the optimization model. The hyper-parameters α and s are chosen by grid search over the validation dataset.

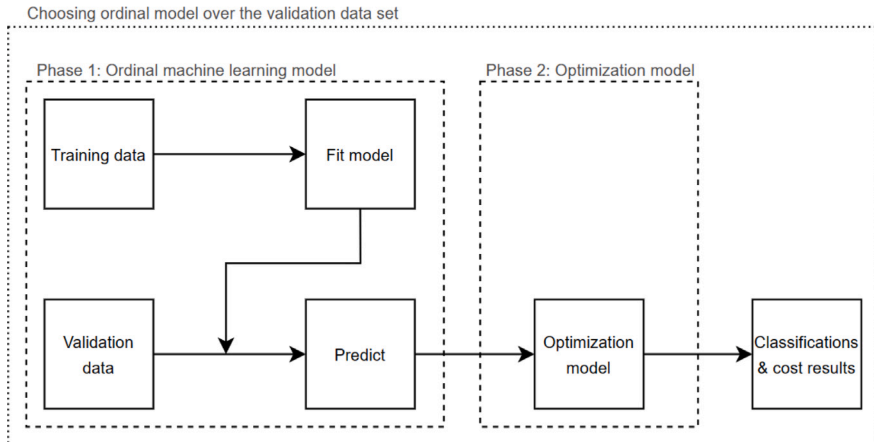


Fig. 1. Schematic illustration of Constraint-based Ordinal Classification Framework (COCF).

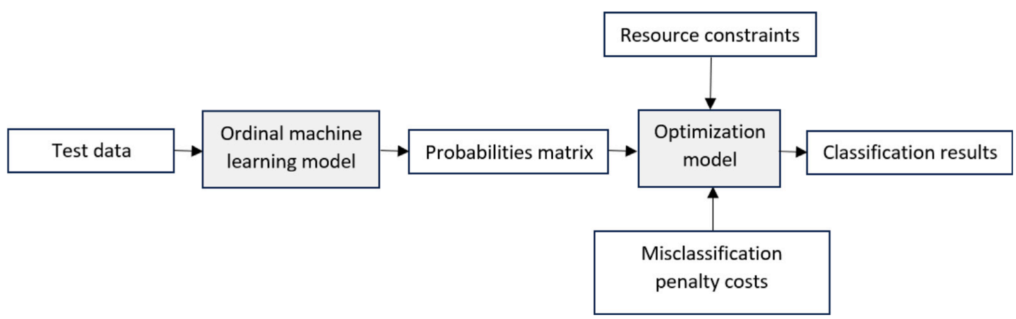


Fig. 2. Schematic process of applying the chosen ordinal machine learning model to test data.

Ordinal neural network. In the experiments with the images dataset, we employed the neural network with ordinal loss from Chen et al. (2019) as an ordinal machine learning model in the first phase of the COCF framework (see Fig. 1). An ordinal loss is proposed using the ordinal cost matrix representing the costs predicting a sample belongs to class i when the observed class is y_s and the estimated probabilities of the samples for each one of the possible classes $\hat{Y}$ as follows,

$$OL = \sum_{\forall x_s} \sum_{\forall i \in C} O_{i,y_s} \hat{Y}_{s,i}. \tag{3}$$

The probability matrix $\hat{Y}$ obtained by the ordinal Neural Network, used as input for the optimization model.

3.3. Resource-constrained optimization model

We now describe the optimization model used in phase two of the proposed COCF framework shown in Fig. 1. The model receives the probability matrix $\hat{Y}$ generated from the ordinal model trained in the first phase of the proposed framework and a misclassification cost matrix $O \in \mathbb{R}^{|C|\times|C|}$.

Our assumption is that the available resources for each class, denoted as $\mathcal{B}_i$ for each class $i \in C$, impose limitations on the number of samples in the considered dataset S' (test dataset) that can be classified as belonging to this class. The learning process takes place on the training dataset, and the model's performance is validated on a separate validation dataset, which often contains different number of samples compared to the test dataset. Thus, we project the resource constraints of the test dataset onto the training and validation datasets. This is achieved by calculating the available resources for each class i as a percentage of all classifications within the considered dataset, expressed as $n_i = \frac{\mathcal{B}_i}{|S'|}$. In this way, we use the resource constraint for each class i for the training and validation processes as a product of the dataset size and n_i. We define a constraint vector $N \in [0,1]^{|C|}$ as input to the optimization model, where these n_i's values relate to all classes.

Table 1
Notation in the resource-constrained optimization model.

Notation	Size	Meaning
S	$\lvert S\rvert$	Set of samples
C	$\lvert C\rvert$	Set of classes
O	$\lvert C\rvert \times \lvert C\rvert$	Ordinal cost matrix
N	$\lvert C\rvert \times \lvert 1\rvert$	Constraint vector
$\hat{Y}$	$\lvert S\rvert \times \lvert C\rvert$	Probability matrix
R	$\lvert S\rvert \times \lvert C\rvert$	Result vector

The output of the optimization model is represented as $R \in \{0,1\}^{|S|\times|C|}$, which is a binary matrix. In this matrix, $R_{s,i} = 1$ indicates that a sample s classified as belonging to class i. The model ensures that each sample is classified into exactly one class. Table 1 presents the notation used in describing the problem.

We now can model the resource allocation problem as follows:

$$\min \quad \sum_{s=1}^{|S|} \sum_{i=1}^{|C|} R_{s,i} \sum_{j=1}^{|C|} (O_{i,j} \cdot \hat{Y}_{s,j}) \tag{4}$$

$$\text{s.t.} \quad \sum_{s=1}^{|S|} R_{s,i} \leq n_i \cdot |S|, \quad \forall i \in C \tag{5}$$

$$\sum_{i=1}^{|C|} R_{s,i} = 1, \quad \forall s \in S. \tag{6}$$

Our goal is to minimize a cost function. We define two constraints within the optimization model. As defined in Eq. (5), the first constraint guarantees that the number of samples classified as label i remains below the resource constraint, expressed as a percentage of the total number of classifications. The second constraint, presented in Eq. (6), ensures that each sample receives a single label assignment. Note that the equality notation of Eq. (6) can be replaced by the inequality ≥ 1, while the same optimal solution of a single label assignment is obtained. An optimal solution for the matrix R is achieved by minimizing the objective function in Eq. (4), while satisfying these constraints.

The resource allocation problem is equivalent to the formulation of the transportation problem, in which the supply exceeds total demand (Ford and Fulkerson, 1956; Babu et al., 2020; Winston, 2004). In this analogy, the set of classes C can represent the warehouses, with $n_i \cdot |S|$ denoting the quantity of goods available at warehouse c_i. The set of samples S can represent the stores, where 1 is the quantity of goods demanded at store x_s. The expression $\sum_{j=1}^{|C|}(O_{i,j} \cdot \hat{Y}_{s,j})$ denotes the transportation cost per unit for any warehouse c_i to any store x_s. The problem involves determining an optimal transportation scheme between the warehouses and the stores. Given the simplicity of finding a basic feasible solution when the total supply equals total demand (referred to as a balanced transportation problem), we will create a dummy sample to solve our problem, with demand equal to the amount of excess supply, and assign a zero cost to shipments between the dummy sample and classes (Winston, 2004). As the constraint matrix of our optimization problem is completely unimodular and the resource constraints and demand values are integers, the values of the variables in each basic solution will be integers (Heller, 1957; Shapiro, 1979). To determine the optimal solution for this problem, the process involves establishing an initial feasible solution in the first step and then determining the optimal solution using this initial solution (Korukoğlu and Ballı, 2011). To find an initial solution, well-known transportation methods such as Vogel's Approximation Method (VAM) and the northwest corner method can be utilized (Korukoğlu and Ballı, 2011; Tularam and Bhayo, 2014). While some heuristics quickly offer an initial feasible solution that may not effectively minimize total costs, others may yield highly effective solutions for minimizing total costs but require significantly more time (Tularam and Bhayo, 2014). VAM is a commonly used heuristic that usually provides a better starting solution than other methods with a competitive computational time complexity of $\mathcal{O}(mn)$, where m and n in our problem are the number of classes $|C|$ and the number of instances $|S|$, respectively (Chaudhuri et al., 2013). Although VAM does not guarantee an optimal solution, it yields an optimal or close-to-optimal starting point for small-sized transportation problems. In the case of large-sized problems, in terms of the number of samples and classes, an improved version of VAM, such as IVAM (Korukoğlu and Ballı, 2011), can be employed to obtain more efficient initial solutions.

4. Experimental study

In the following section, we present the results of our experimental investigation of the proposed COCF and compare them with the results obtained using the conventional approach of combining a non-ordinal classifier with an optimization model (referred to as CNOCF, which stand for Constraint-based Non-Ordinal Classification Framework). The methodology will be applied to both tabular and image datasets. For the tabular dataset, our COCF will utilize an ordinal decision tree and an ordinal random forest in the first phase, while their corresponding non-ordinal algorithms will be applied for the purpose of comparison in the CNOCF. When working with image dataset, we apply an ordinal neural network in the first phase of the COCF and the corresponding non-ordinal neural network in the CNCOF. Details about the ordinal methods can be found in Section 3.2. In all our experiments involving the ordinal decision-tree based algorithms, we employed a function $v(c_i) = i, \forall i$. For the ordinal loss function of the neural network and optimization model, we utilized an ordinal cost matrix O as proposed in Chen et al. (2019), where $\forall i,j : O_{i,j} = 2 \cdot |v(c_j) - v(c_i)| + 1$. This matrix is designed so that a larger deviation between the actual and predicted classes reflects errors with more significant consequences. It is worth noting that subtracting 1 from all values in the ordinal cost matrix, in order to assign a cost of 0 for correctly classified samples, does not impact the classification results but only reduces the overall cost of the solution. As the framework aims to minimize ordinal misclassification errors, it is important to note that it may not necessarily optimize other performance measures (Abukasis et al., 2022). The misclassification errors will be presented as calculated by,

$$\sum_{s=1}^{|S|} \sum_{i=1}^{|C|} R_{s,i} O_{i,y_s}. \tag{7}$$

Although the proposed framework is designed to yield favorable ordinal error results, we will also evaluate its capability to produce competitive accuracy results.

4.1. Implementation of COCF on tabular dataset

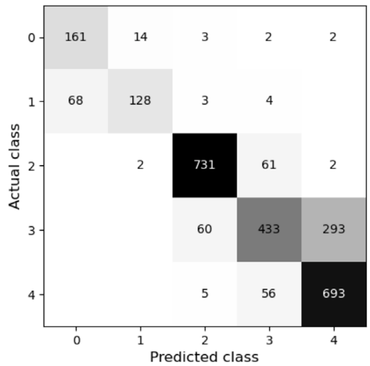
Datasets, experimental setup and methods. COCF with the ordinal decision tree and ordinal random forest algorithms was applied to a publicly available dataset, which can be accessed at https://www.kaggle.com/datasets/imsparsh/churn-risk-rate-hackerearth-ml.

The dataset contains personal information about users, including but not limited to browsing behavior, historical purchase data, and demographic information. Each user is assigned one among five discrete predictive values on an ordinal scale from 0 to 4, which estimates their likelihood of churning at any given time. A value of 0 represents a low probability of churning, while 4 indicates a high probability of churning. The objective is to forecast the churn score based on the provided dataset's features. To prepare the data for analysis, we performed data engineering, which involved the removal of instances with incomplete data and the removal of features that have a single value or a minimal number of distinct values, thereby having a limited impact on the target variable. Following this preprocessing stage, the dataset contained 13,609 training instances and 2,721 test instances.

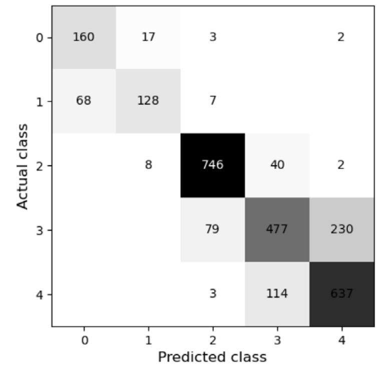
In our experiments, we explore several scenarios involving constraints. Several scenarios focus on users with a maximum risk of churning equal to 4 (Class 4), and can potentially be retained only with a phone call by team leaders. Here, the constraint is related to the number of team leaders available to make these calls. Another scenario focus on users with a risk of churning equal to 3 (Class 3), representing unstable users who can potentially be retained with a simple phone call by employees. All those scenarios assume a single constraint of a specific resource type. We also explore a scenario that involves constraints on two types of resources. In this case, limitations are placed on users with a risk of churning equal to 2 and 4 simultaneously. Class 2 users, who are less likely to churn, are used as training calls for new employees. However, calls to Class 4 users, who have a higher likelihood of churning, are overseen by team leaders. It is important to note that both new employees and team leaders are constrained resources.

The ordinal and non-ordinal decision tree-based algorithms were trained on the training dataset. To select the optimal hyperparameters, we employed a cross-validation strategy. The training dataset was partitioned into 5 groups. In each of the 5 iterations, the framework was trained on 4 of these groups and validated on the fifth. The hyperparameters that yielded the lowest mean cost across these 5 validation runs were then chosen for the subsequent step. Subsequently, the entire framework was applied to the test data, resulting in classification results for each instance. These classifications were made while adhering to the specified constraints and aimed at minimizing the overall cost.

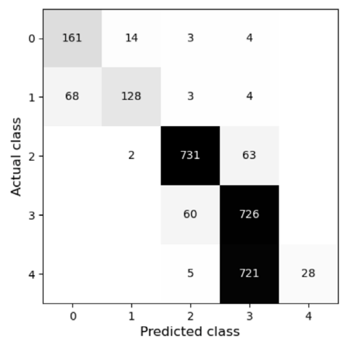
Results. To emphasize the significance of employing an ordinal model in conjunction with an optimization model, we offer a comprehensive comparison of the actual labels versus the predicted labels through confusion matrices for both the proposed COCF and CNOCF with the ordinal decision tree-based algorithms and their non-ordinal counterparts respectively. Fig. 3 displays these matrices, where the upper matrices depict results without a resource constraint, while the lower matrices represent the outcomes when a constraint is applied to class 4 (i.e., $n_4 =$



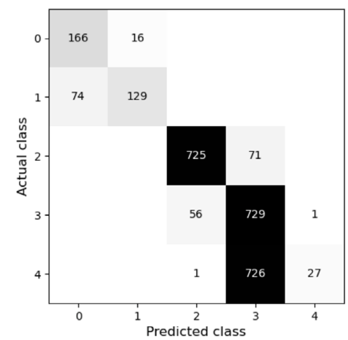
(a) CNOCF using decision tree on the tabular dataset W/o constraints



(b) COCF using ordinal decision tree on the tabular dataset W/o constraints



(c) CNOCF using decision tree on the tabular dataset with $n_4 = 1\%$



(d) COCF using ordinal decision tree on the tabular dataset with $n_4 = 1\%$

Fig. 3. Confusion matrices of the actual labels versus the predicted labels for COCF and CNOCF on the tabular dataset, without a resource constraint (upper matrices) and when applying 1% constraint on class 4.

1%). In the lower matrices, the fifth column (corresponding to class 4) sums to 28, representing the resource limitation of 1% of the test dataset.

When examining the confusion matrices, it can be observed that when using COCF, there are fewer cases with significant errors, as indicated by a distance of two steps or more from the diagonal (representing correct classification), compared to when applying the CNOCF. This behavior indicates that COCF takes into account ordinality during the machine learning phase, resulting in lower cost outcomes compared to CNOCF. For example, when applying the constraint to class 4 (the lower matrices), it can be observed that in the non-ordinal approach, there are 5 instances classified as class 2 instead of class 4, compared to only 1 instance in the ordinal model. Additionally, among the instances classified as class 3, there are instances with an actual class from each of the other classes. This phenomenon occurs because a significant number of instances were classified as class 3 as a result of the limited number of instances to be classified into class 4.

Fig. 4 presents the impact of the optimization model on the machine learning model's classifications for scenarios without resource constraints and when a constraint is applied to class 3 (i.e. $n_3 = 3\%$). The figure shows the percentage of samples whose cost, derived from the decision tree (DT) and ordinal decision tree (ord_DT) classifications, was either improved, left unchanged, or worsened by the optimization model, denoted 'pos', 'equal' and 'neg', respectively. The left pair of bars depict the results of the decision tree and ordinal decision tree for the misclassification costs problem without constraints. The right pair of bars illustrate the results for the misclassification cost problem under a resource constraint of $n_3 = 3\%$. As expected in the resource constraint scenario, the percentage of 'neg' samples was higher than

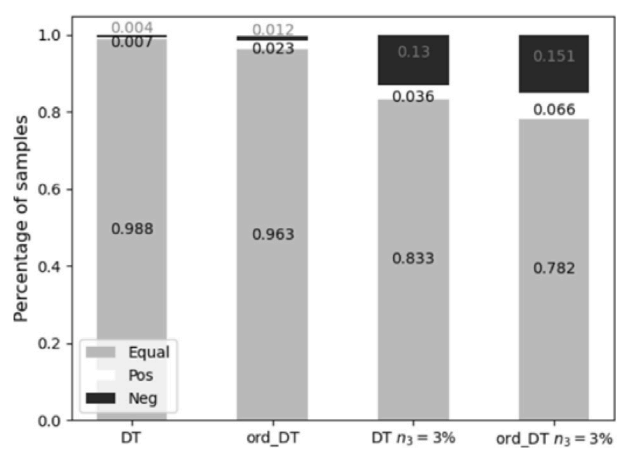
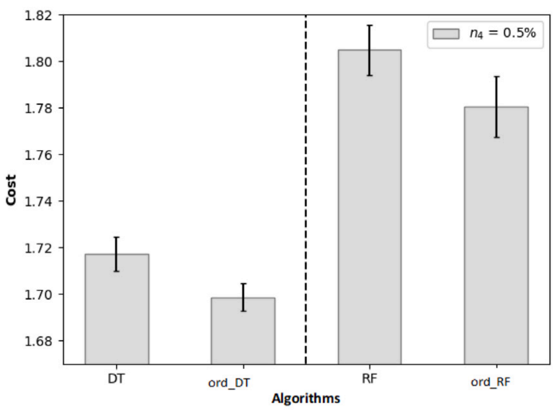
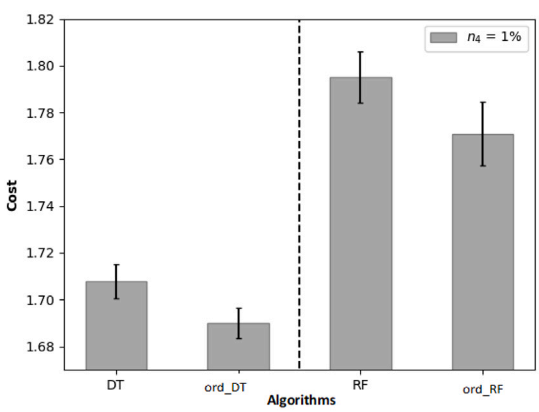


Fig. 4. Percentage of samples whose cost, derived from the decision tree and ordinal decision tree classifications, were improved, left unchanged, or worsened by the optimization model, denoted by 'pos', 'equal' and 'neg', respectively. The left pair of bars presents the problem without resource constraints, and the right pair of bars presents the problem under a resource constraint.

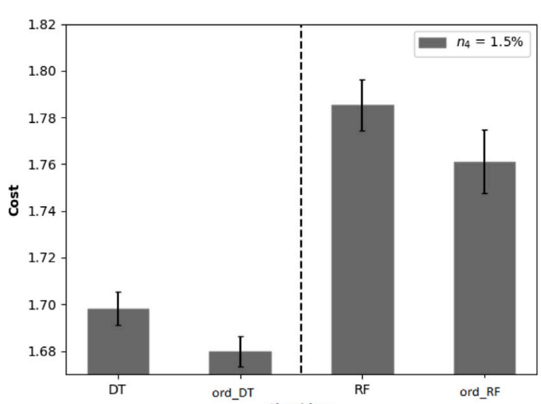
the percentage of 'pos' instances, unlike the unconstrained problem. This is because classification changes are necessary to comply with the resource constraint, even if it results in an increase in cost. It can be observed that the percentage of 'pos' instances in the ordinal decision tree was higher at 83% compared to the percentage of 'pos' instances in the decision tree (i.e., 6.6% compared to 3.6%). This observation



(a) $n_4 = 0.5\%$



(b) $n_4 = 1\%$



(c) $n_4 = 1.5\%$

Fig. 5. Cost and standard deviation results for the ordinal decision tree and ordinal random forest, compared to their non-ordinal counterparts for the Class 4 constraint scenarios of $n_4 = 0.5\%$, 1%, and 1.5%.

is logical as the ordinal decision tree, incorporating the ordinality of classes during the learning phase, has a higher probability of positive improvement.

Fig. 5 presents the mean and standard deviation of the cost results for each model, based on 5 cross-validation runs, under the Class 4 constraint scenarios of $n_4 = 0.5\%, 1\%$, and 1.5%. It can be observed that the overlap between the standard deviations for each ordinal decision tree-based model and its non-ordinal counterpart is minimal, indicating a consistent finding that the cost is lower in the ordinal models.

Table 2
Cost and accuracy results obtained by applying COCF and CNOCF on validation datasets with a five-fold cross-validation, utilizing the ordinal decision tree and non-ordinal decision tree models respectively. The results of the best framework are highlighted in bold for each set of constraints.

Set of constraints	CNOCF with DT		COCF with ord_DT	
	Cost	Acc	Cost	Acc
No constraints	1.456	0.779	**1.438**	**0.785**
$n_4 = 0.5\%$	1.717	0.647	**1.699***	**0.651**
$n_4 = 1\%$	1.708	0.652	**1.69***	**0.656***
$n_4 = 1.5\%$	1.698	0.657	**1.68***	**0.66**
$n_4 = n_2 = 0.5\%$	2.19	0.411	**2.17***	**0.415***
$n_3 = 3\%$	1.614	0.705	**1.598***	**0.707**

* Paired t-test significance at p-value<0.05.

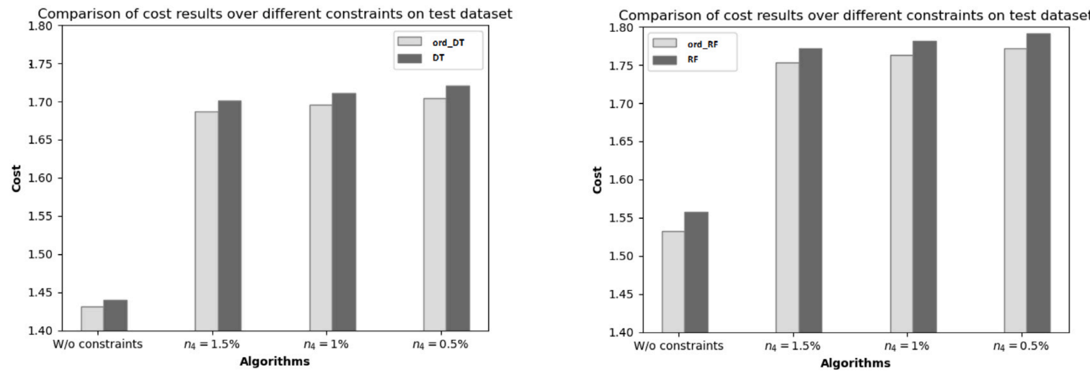
Table 3
Cost and accuracy results obtained by applying COCF and CNOCF on test dataset, utilizing the ordinal decision tree and non-ordinal decision tree models respectively, in the first phase of the frameworks. For each set of constraints, the results of the best framework are shown in bold.

Set of constraints	CNOCF with DT		COCF with ord_DT	
	Cost	Acc	Cost	Acc
No constraints	1.44	0.789	**1.431**	0.789
$n_4 = 0.5\%$	1.721	0.647	**1.704**	**0.648**
$n_4 = 1\%$	1.711	0.652	**1.695**	**0.653**
$n_4 = 1.5\%$	1.701	0.657	**1.687**	0.657
$n_4 = n_2 = 0.5\%$	2.206	0.404	**2.186**	**0.407**
$n_3 = 3\%$	1.636	0.697	**1.619**	**0.698**

Tables 2 and 3 present the cost and accuracy results for both the COCF with an ordinal decision tree (ord_DT) and CNOCF with a non-ordinal decision tree (DT), under various constraints for the validation datasets with five-fold cross-validation and test dataset respectively. For each set of constraints, the results of the best framework are shown in bold. For the validation datasets, statistically significant differences at a p-value < 0.05 based on paired t-test are indicated by *. In all 6 scenarios, we observed that the COCF with an ord_DT outperformed the CNOCF with the DT on the validation dataset, both in terms of cost and accuracy. In 5 of these 6 scenarios, COCF outperformed the CNOCF in terms of cost with a statistically significant difference (p-value< 0.05). For the test dataset, in all 6 scenarios, COCF outperformed the CNOCF in terms of cost.

Similar to Tables 2 and 3, Tables 4 and 5 present the cost and accuracy results for both the COCF with an ordinal random forest (ord_RF) and CNOCF with a non-ordinal random forest (RF), under various constraints for the validation datasets with five-fold cross-validation and test dataset respectively. For each set of constraints, the results of the best framework are shown in bold. For the validation datasets, statistically significant differences at a p-value < 0.05 based on paired t-test are indicated by *. In all 5 scenarios, we observed that the COCF with an ord_RF outperformed the CNOCF with a RF regarding cost and accuracy with a statistically significant difference (p-value< 0.05) on the validation dataset. For the test dataset in Table 5, in all 5 scenarios, COCF outperformed the CNOCF in terms of cost and in 4 scenarios in terms of accuracy.

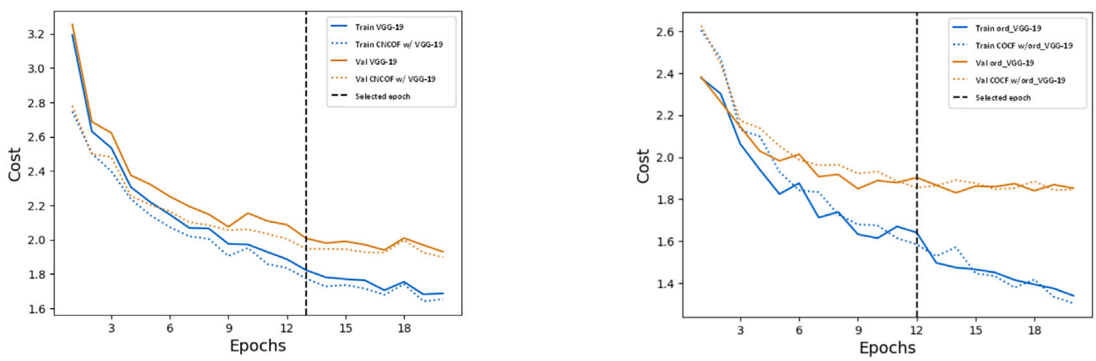
Fig. 6(a) compares the costs obtained by the COCF and CNOCF on a test dataset using the ordinal decision tree (ord_DT) and non-ordinal decision tree (DT) from Table 3. In Fig. 6(b), a comparison of the costs obtained by the COCF and CNOCF on a test dataset is presented, using the ordinal random forest (ord_RF) and non-ordinal random forest (RF) from Table 5. It can be observed that the COCF with the ordinal algorithms achieved lower costs than the CNOCF with their non-ordinal counterparts, resulting in an average cost reduction of 1%.



(a) Decision tree

(b) Random forest

Fig. 6. Comparison of COCF and CNOCF cost results on a test dataset using (a) decision tree and ordinal decision tree models, and (b) random forest and ordinal random forest models, across different sets of constraints.



(a) VGG-19 compared to CNOCF with VGG-19

(b) ordinal VGG-19 compared to COCF with ordinal VGG-19

Fig. 7. The misclassification costs in scenarios without resource constraints for both the training and validation datasets computed over 20 epochs following the classification phase (using either VGG-19 or ordinal VGG-19) and after applying the optimization model (CNOCF with VGG-19 or COCF with ordinal VGG-19). The stopping epoch is indicated by the black dashed lines.

Table 4
Cost and accuracy results obtained by applying COCF and CNOCF on validation datasets, utilizing the ordinal random forest and non-ordinal random forest models respectively. The results of the best framework are highlighted in bold for each set of constraints.

Set of constraints	CNOCF with RF		COCF with ord_RF	
	Cost	Acc	Cost	Acc
No constraints	1.585	0.724	**1.534***	**0.744***
$n_4 = 0.5\%$	1.805	0.61	**1.78***	**0.618***
$n_4 = 1\%$	1.795	0.615	**1.77***	**0.623***
$n_4 = 1.5\%$	1.785	0.62	**1.761***	**0.628***
$n_3 = 3\%$	1.757	0.654	**1.7***	**0.671***

* Paired t-test significance at p-value<0.05.

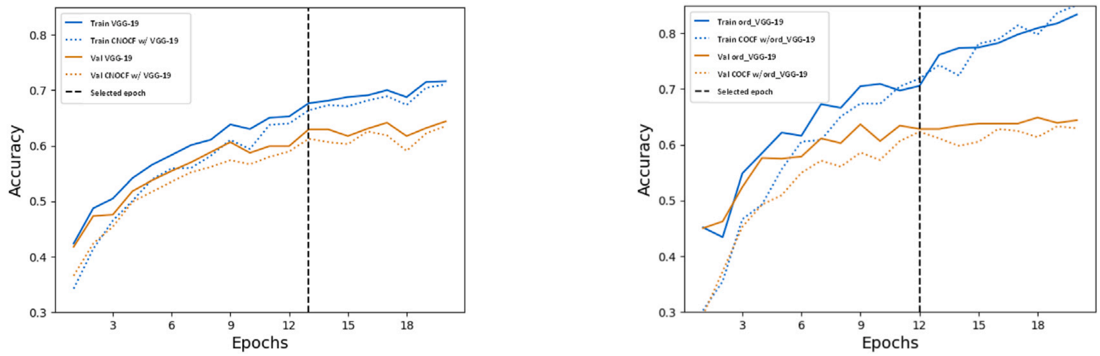
Table 5
Cost and accuracy results obtained by applying COCF and CNOCF on test dataset, utilizing the ordinal random forest and non-ordinal random forest models respectively. For each set of constraints, the results of the best framework are shown in bold.

Set of constraints	CNOCF with RF		COCF with ord_RF	
	Cost	Acc	Cost	Acc
No constraints	1.558	0.742	**1.532**	**0.75**
$n_4 = 0.5\%$	1.792	0.621	**1.772**	**0.629**
$n_4 = 1\%$	1.781	0.626	**1.763**	**0.634**
$n_4 = 1.5\%$	1.772	0.631	**1.753**	**0.639**
$n_3 = 3\%$	1.737	**0.665**	**1.726**	0.658

4.2. Implementation of COCF on image dataset

Datasets, experimental setup and methods. In this part, COCF was applied with an ordinal neural network of VGG-19 type, as presented in article (Chen et al., 2019), on a publicly available dataset obtained from the Osteoarthritis Initiative (OAI), and available at https://nda.nih.gov/oai/. Knee Osteoarthritis (OA) is a frequent cause of limited activity and physical disability in the elderly population. Early identification and treatment can potentially slow down its progression, thereby optimizing the necessary medical interventions. A total of 4796 images of knee bilateral posterior–anterior fixed flexion radiographs of 4796 participants exist in the dataset. In our research, we used a total of 4130 pairs of knee joints, acquired from the OAI repository and made available by Chen et al. (2019) after a reprocessing stage. All knee X-ray images were randomly divided into training, validation, and test datasets at a ratio of 7:1:2. The aim of this experiment was to categorize patients into five classes, each representing different levels of OA severity, and then assign these patients to limited treatment resources based on the classification results. It is assumed that the cost of incorrectly assigning a patient to treatment is influenced by the distance between the actual class and the predicted class, as represented by the ordinal cost matrix O.

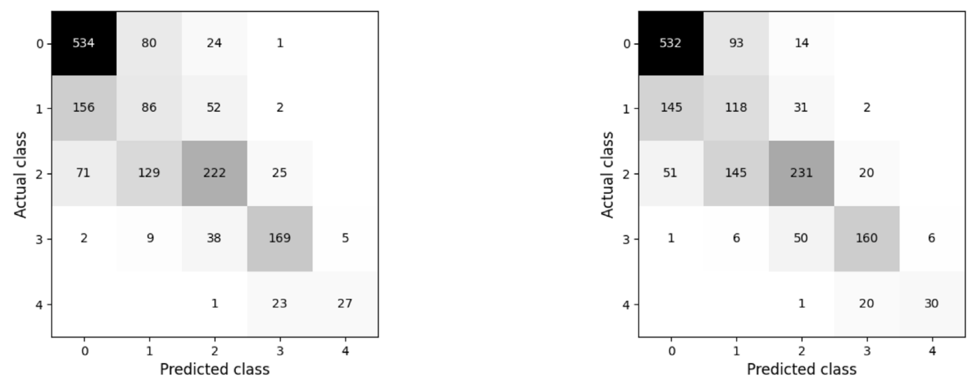
We used the exact epoch selection mechanism to find the "best" trained model in all experiments. The selected epoch is that with a lower obtained cost (after applying the optimization model) than all



(a) VGG-19 compared to CNOCF with VGG-19

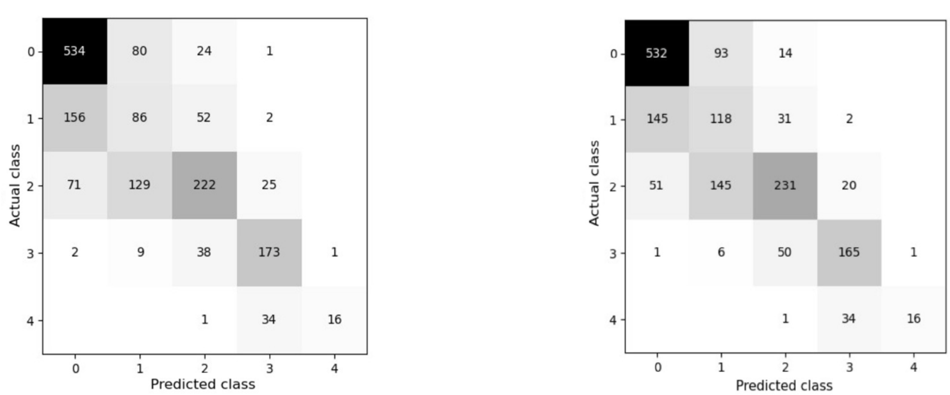
(b) ordinal VGG-19 compared to COCF with ordinal VGG-19

Fig. 8. The accuracy results in scenarios without resource constraints for both the training and validation datasets computed over 20 epochs following the classification phase (using either VGG-19 or ordinal VGG-19) and after applying the optimization model (CNOCF with VGG-19 or COCF with ordinal VGG-19). The stopping epoch is indicated by the black dashed lines.



(a) CNCOF using VGG-19 on the images dataset W/o constraints

(b) COCF using ordinal VGG-19 on the images dataset W/o constraints

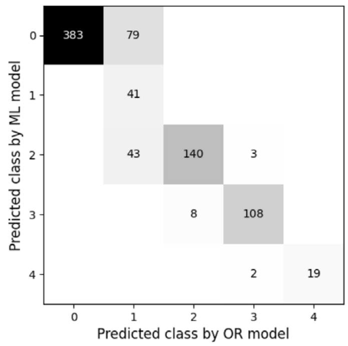


(c) CNOCF using VGG-19 on the images dataset with $n_4 = 1\%$

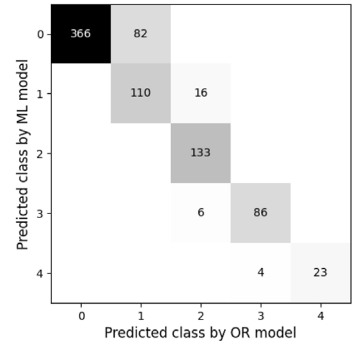
(d) COCF using ordinal VGG-19 in the images dataset with $n_4 = 1\%$

Fig. 9. Confusion matrices of the actual labels versus the predicted labels for CNOCF and COCF on the images dataset, without a resource constraint (upper matrices) and when applying 1% constraint on class 4.

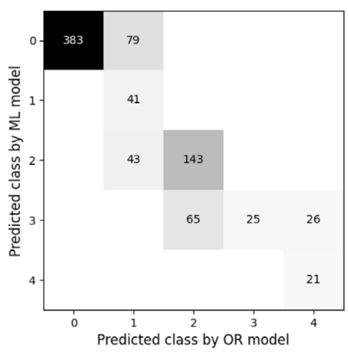
previous epochs and without a cost improvement higher than 2.5% in the following five epochs. In our experiments, we explored two different scenarios that imposed a constraint on the number of treatments. The first scenario is for patients with severity level 4, denoted as $B_4 = 17$. This corresponds to 1% of the patients in the considered dataset (test dataset), which has a total of $|S'| = 1656$ patients. This scenario mirrors a practical real-world situation where a healthcare center needs to prioritize patients in the later stages of the disease, optimizing the allocation of limited surgery rooms. The second scenario imposed a constraint on the number of treatments for patients with severity level 3, where $B_3 = 50$, representing $n_3 = 3\%$ of the patients in the test dataset. This situation mirrors a practical scenario wherein a doctor may have a limited number of monitoring devices to detect the deterioration of the disease.



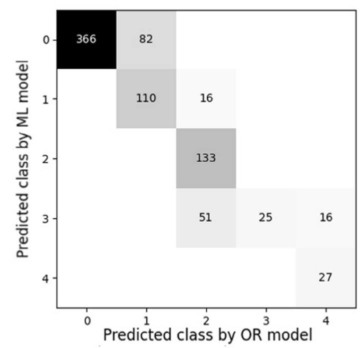
(a) CNOCF using VGG-19 without constraints



(b) COCF using ordinal VGG-19 without constraints



(c) CNOCF using VGG-19 with $n_3 = 3\%$



(d) COCF using ordinal VGG-19 with $n_3 = 3\%$

Fig. 10. Transition matrices of the predicted labels for VGG-19 and ordinal VGG-19 versus the predicted labels for CNOCF with VGG-19 and COCF with ordinal VGG-19, without a resource constraint (upper matrices) and when applying a 3% constraint on severity level 3 (lower matrices).

Results. In this section, we present the results of COCF with ordinal VGG-19 to showcase the versatility of using COCF across different types of datasets and with various ordinal algorithms. Figs. 7 and 8 illustrate the influence of the optimization model on the machine learning model's classifications in scenarios without resource constraints. As expected, when utilizing a VGG-19 classifier, Fig. 7(a) demonstrates a cost improvement for both the training and validation datasets after applying the optimization model. However, for the ordinal VGG-19 classifier, it is observed from Fig. 7(b), that the optimization model did not contribute to a further enhancement in the cost results. Fig. 8(a) presents that for both the training and validation datasets, the accuracy is lower when using the optimization model (which aims to minimize cost) compared to using only the VGG-19 classifier. In Fig. 8(b), it is observed that when using the ordinal VGG-19 classifier, the optimization model did not significantly affect the accuracy results, particularly on the training dataset, similarly to the cost results in Fig. 7(b). The black dashed lines in the graphs indicate the selected epochs determined by the stopping mechanism. The models trained at these epochs were saved and utilized for the test datasets.

Fig. 9 displays the confusion matrices comparing the actual labels with the predicted labels for COCF vs. CNOCF, both without resource constraints (upper matrices) and when a constraint is applied to limit the number of treatments for patients with severity level 4, denoted as $n_4 = 1\%$ (lower matrices). The fifth column (representing severity level 4) of the lower matrices sums to 17, which corresponds to 1% of the patients in the considered dataset (test dataset). Using these confusion matrices, we conducted further analyses to assess whether the utilization of the ordinal learning model, in contrast to the non-ordinal learning model, results in improved classification outcomes with reduced costs. These improved starting conditions are advantageous for the subsequent optimization model, which refines the classification to adhere to the constraint, ultimately leading to better results. From the upper matrices, it can be observed that while CNOCF with the VGG-19 achieved an accuracy of 62.7% and a cost of 3118, COCF with the ordinal VGG-19 achieved an accuracy of 64.7% and a cost of 2978, representing an 4.5% cost reduction. It is also worth noting that when using COCF with the ordinal VGG-19, there were 75 cases with significant errors, as indicated by a distance of two steps or more from the diagonal (representing correct classification), compared to 110 when applying CNOCF with the non-ordinal VGG-19. With the constraint in place, several instances initially classified as class 4 were changed to class 3 to meet the constraint. Specifically, 19 instances were altered when applying COCF, with 5 of them correctly classified as Class 3. On the other hand, 15 instances were changed when applying CNOCF, with 4 of them correctly classified as class 3. Similar differences in results were observed in the lower matrices, with an accuracy of 62.3% and a cost of 3132 when applying CNOCF, and an accuracy of 64.1% and a cost of 2996 (representing a 4.3% cost reduction) when applying COCF.

In Fig. 4, we explored for the tabular experiment the impact of the optimization model on the machine learning model's classifications via the calculation of the percentage of improved, unchanged, and worsened changes. In this experiment, we will investigate in detail the number of changes between each pair of classes. Fig. 10 presents the transition matrices of the predicted labels of the machine learning algorithm versus the predicted labels after the optimization model, without a resource constraint (upper matrices) and when applying a constraint on the number of treatments for patients with severity level 3, i.e., $n_3 = 3\%$ (lower matrices). The fourth column (representing severity level 3) of the lower matrices sums to 25, equaling 3% of the validation data set. These matrices show how the optimization models influence the machine learning results. According to Fig. 10, constraining matrices (a) and (b) results in matrices (c) and (d), respectively, which alters the classification of instances by categorizing them as 2 or 4 to conform to the restriction on class 3. It can be observed that the number of changes between classes by the optimization model is 25% and 29% higher when using the non-ordinal VGG-19 compared to when using the ordinal VGG-19 for the scenarios without constraints and with the constraint of $n_3 = 3\%$, respectively. This insight is rational since ordinal VGG-19 takes into account the ordinality of classes already in the learning phase.

5. Conclusion

We introduce a framework that integrates an ordinal machine learning algorithm with an optimization model to address ordinal classification problems under resource constraints. We illustrated the equivalence between the formulation of the resource allocation problem into samples and the transportation problem, enabling the utilization of established transportation heuristics for our solution. Moreover, we demonstrated that the characteristics of the problem formulation facilitate the identification of a solution with integer variables. Our experiments demonstrate that this combined framework, which utilizes both an ordinal machine learning algorithm and an optimization model, consistently outperforms its non-ordinal counterpart in various constraint scenarios. This is shown in the case of both tabular and image datasets, where we employ ordinal decision tree-based models and ordinal neural networks. These experiments consistently demonstrate the superiority of our proposed framework with ordinal models, resulting in an average cost reduction of 1% for the ordinal decision tree-based models and 4.4% for the ordinal neural networks. The experiments also revealed that, as ordinal classifiers consider the ordinality of classes during the learning phase, the optimization model did not contribute to a further improvement in the cost results in the scenario without resource constraints. Additionally, there were much lower classification changes when using the ordinal classifiers compared to the non-ordinal classifiers. Future research direction can explore ensemble algorithms combining ordinal and non-ordinal models to improve cost results. Another potential research direction could explore the applicability of the proposed framework to other ordinal classification problems involving diverse datasets, including text, audio, or EEG signals (Haba et al., 2023). It would also be interesting to explore the utilization of the classification changes performed by the optimization model to implement adaptive learning of the ordinal learning classifier, which has recently been investigated to address local training-test class distribution mismatch and resource constraints in a binary classification problem (Shifman et al., 2023; Volk et al., 2023; Volk and Singer, 2023).

CRediT authorship contribution statement

Lior Rabkin: Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft. **Ilan Cohen:** Supervision, Methodology, Writing – review & editing. **Gonen Singer:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We shared links to the public datasets used.

References

Abukasis, Danit Shifman, Cohen, Izack, Xian, Xiaochen, Huang, Kejun, Singer, Gonen, 2022. Adaptive learning for the resource-constrained classification problem. arXiv preprint arXiv:2207.09196.

Akan, Oguzhan, Verma, Abhishek, 2022. Machine learning models for customer churn risk prediction. In: 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference. UEMCON, IEEE, pp. 0623–0628.

Babu, Md Ashraful, Ahmmed, Md Mortuza, Salim, Zahir Rayhan, Babu, Md Shohel, Abdul, Mohammad, 2020. A Brief Overview of the Classical Transportation Problem. Xi'an University of Architecture & Technology.

Chaudhuri, Arindam, De, Kajal, Chatterjee, Dipak, Mitra, Pabitra, 2013. Trapezoidal fuzzy numbers for the transportation problem. arXiv preprint arXiv:1307.1893.

Chen, Pingjun, Gao, Linlin, Shi, Xiaoshuang, Allen, Kyle, Yang, Lin, 2019. Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss. Comput. Med. Imaging Graph. 75, 84–92.

Dikkers, Harmen, Rothkrantz, Leon, 2005. Support vector machines in ordinal classification: An application to corporate credit scoring. Neural Network World 15 (6), 491.

Eshghali, Masoud, Kannan, Devika, Salmanzadeh-Meydani, Navid, Esmaieeli Sikaroudi, Amir Mohammad, 2023. Machine learning based integrated scheduling and rescheduling for elective and emergency patients in the operating theatre. Ann. Oper. Res. 1–24.

Ford, Jr., Lester Randolph, Fulkerson, Delbert Ray, 1956. Solving the transportation problem. Manage. Sci. 3 (1), 24–32.

Frumosu, Flavia Dalia, Khan, Abdul Rauf, Schiøler, Henrik, Kulahci, Murat, Zaki, Mohamed, Westermann-Rasmussen, Peter, 2020. Cost-sensitive learning classification strategy for predicting product failures. Expert Syst. Appl. 161, 113653.

Gartner, Daniel, Kolisch, Rainer, Neill, Daniel B., Padman, Rema, 2015. Machine learning approaches for early DRG classification and resource allocation. INFORMS J. Comput. 27 (4), 718–734.

Gutiérrez, Pedro Antonio, Perez-Ortiz, Maria, Sanchez-Monedero, Javier, Fernandez-Navarro, Francisco, Hervas-Martinez, Cesar, 2015. Ordinal regression methods: survey and experimental study. IEEE Trans. Knowl. Data Eng. 28 (1), 127–146.

Haba, Rotem, Singer, Gonen, Naftali, Sara, Kramer, Mordechai R., Ratnovsky, Anat, 2023. A remote and personalised novel approach for monitoring asthma severity levels from EEG signals utilizing classification algorithms. Expert Syst. Appl. 223, 119799.

He, Deniu, 2022. Active learning for ordinal classification based on adaptive diversity-based uncertainty sampling. IEEE Access.

Heller, I., 1957. Constraint matrices of transportation-type problems. Nav. Res. Logist. Q. 4 (1), 73–76.

Juraev, Gayrat, Rakhimberdiev, Kuvonchbek, 2022. Mathematical modeling of credit scoring system based on the monge-kantorovich problem. In: 2022 IEEE International IOT, Electronics and Mechatronics Conference. IEMTRONICS, IEEE, pp. 1–7.

Kim, Gyeongho, Choi, Jae Gyeong, Ku, Minjoo, Lim, Sunghoon, 2023. Developing a semi-supervised learning and ordinal classification framework for quality level prediction in manufacturing. Comput. Ind. Eng. 181, 109286.

Korukoğlu, Serdar, Ballı, Serkan, 2011. An improved Vogel's approximation method for the transportation problem. Math. Comput. Appl. 16 (2), 370–381.

Lázaro, Marcelino, Figueiras-Vidal, Aníbal R., 2023. Neural network for ordinal classification of imbalanced data by minimizing a Bayesian cost. Pattern Recognit. 137, 109303.

Li, Li, Chen, Xiaoting, Zhu, Peng, 2024. How do e-commerce anchors' characteristics influence consumers' impulse buying? An emotional contagion perspective. J. Retail. Consum. Serv. 76, 103587.

Li, Hongjian, Ding, Shiwang, Zhang, Pengfei, Lai, Jun, 2018. Energy-efficient resource allocation strategy based on task classification in data center. In: 2018 3rd International Conference on Automation, Mechanical Control and Computational Engineering (AMCCE 2018). Atlantis Press, pp. 111–119.

Lima, Fernando D.S., Pereira, Francisco Lucas F., Chaves, Iago C., Machado, Javam C., Gomes, Joao Paulo P., 2020. Predicting the health degree of hard disk drives with asymmetric and ordinal deep neural models. IEEE Trans. Comput. 70 (2), 188–198.

Liu, Weiqiang, Ge, Tianshuo, Luo, Linkai, Peng, Hong, Xu, Xide, Chen, Yuangui, Zhuang, Zefeng, 2022. A novel focal ordinal loss for assessment of Knee Osteoarthritis severity. Neural Process. Lett. 54 (6), 5199–5224.

Marudi, Matan, Ben-Gal, Irad, Singer, Gonen, 2022. A decision tree-based method for ordinal classification problems. IISE Trans. 1–20.

Nabi, Fizza Ghulam, Sundaraj, Kenneth, Lam, Chee Kiang, Palaniappan, Rajkumar, 2019. Characterization and classification of asthmatic wheeze sounds according to severity level using spectral integrated features. Comput. Biol. Med. 104, 52–61.

Pessach, D., Singer, G., Avrahami, D., Ben-Gal, H.C., Shmueli, E., Ben-Gal, I., 2020. Employees recruitment: A prescriptive analytics approach via machine learning and mathematical programming. Decis. Support Syst. 1132.

Rosati, Riccardo, Romeo, Luca, Vargas, Víctor Manuel, Gutiérrez, Pedro Antonio, Hervás-Martínez, César, Frontoni, Emanuele, 2022. A novel deep ordinal classification approach for aesthetic quality control classification. Neural Comput. Appl. 34 (14), 11625–11639.

Shapiro, Jeremy F., 1979. Mathematical programming: structures and algorithms.

Shifman, Danit Abukasis, Cohen, Izack, Huang, Kejun, Xian, Xiaochen, Singer, Gonen, 2023. An adaptive machine learning algorithm for the resource-constrained classification problem. Eng. Appl. Artif. Intell. 119, 105741.

Silva, Joselito Adriano da, Emi, Angélica Santos, Leão, Eliseth Ribeiro, Lopes, Maria Carolina Barbosa Teixeira, Okuno, Meiry Fernanda Pinto, Batista, Ruth Ester Assayag, 2017. Emergency severity index: accuracy in risk classification. Einstein (Sao Paulo) 15, 421–427.

Singer, Gonen, Anuar, Roee, Ben-Gal, Irad, 2020. A weighted information-gain measure for ordinal classification trees. Expert Syst. Appl. 152, 113375.

Singer, Gonen, Cohen, Izack, 2020. An objective-based entropy approach for interpretable decision tree models in support of human resource management: The case of absenteeism at work. Entropy 22 (8), 821.

Singer, Gonen, Marudi, Matan, 2020. Ordinal decision-tree-based ensemble approaches: The case of controlling the daily local growth rate of the COVID-19 epidemic. Entropy 22 (8), 871.

Singer, Gonen, Ratnovsky, Anat, Naftali, Sara, 2021. Classification of severity of trachea stenosis from EEG signals using ordinal decision-tree based algorithms and ensemble-based ordinal and non-ordinal algorithms. Expert Syst. Appl. 173, 114707.

Sweeney, Jillian C., Soutar, Geoffrey N., 2001. Consumer perceived value: The development of a multiple item scale. J. Retail. 77 (2), 203–220.

Tran, Son, Verhoeven, Peter, 2021. Kelly criterion for optimal credit allocation. J. Risk Financial Manag. 14 (9), 434.

Tularam, Abdul Sattar Soomro1 Gurudeo Anand, Bhayo, Ghulam Murtaa, 2014. A comparative study of initial basic feasible solution methods for transportation problems.

Volk, Ohad, Ratnovsky, Anat, Naftali, Sara, Singer, Gonen, 2023. Classification of tracheal stenosis with asymmetric misclassification errors from EMG signals using an adaptive cost-sensitive learning method. Biomed. Signal Process. Control 85, 104962.

Volk, Ohad, Singer, Gonen, 2023. An adaptive cost-sensitive learning approach in neural networks to minimize local training–test class distributions mismatch. Intell. Syst. Appl. 200316.

Wang, Tianhui, Liu, Renjing, Qi, Guohua, 2022. Multi-classification assessment of bank personal credit risk based on multi-source information fusion. Expert Syst. Appl. 191, 116236.

Wang, Jingan, Shuo, Meng, Wang, Lei, Sun, Fengxin, Pan, Ruru, Gao, Weidong, Shi, Kangjun, 2021. Objective evaluation of fabric smoothness appearance with an ordinal classification framework based on label noise estimation. Textile Res. J. 91 (3–4), 316–334.

Winston, Wayne L., 2004. Operations Research: Applications and Algorithm. Thomson Learning, Inc..

Wu, Xing, Li, Pan, Zhao, Ming, Liu, Ying, Crespo, Rubén González, Herrera-Viedma, Enrique, 2022. Customer churn prediction for web browsers. Expert Syst. Appl. 209, 118177.

Yıldırım, Pelin, Birant, Ulaş K., Birant, Derya, 2019. EBOC: Ensemble-based ordinal classification in transportation. J. Adv. Transp. 2019, 1–17.

Zhang, Huan, Jiang, Liangxiao, Li, Chaoqun, 2021. CS-ResNet: Cost-sensitive residual convolutional neural network for PCB cosmetic defect detection. Expert Syst. Appl. 185, 115673.

Zhang, Siyu, Liao, Peng, Ye, Heng-Qing, Zhou, Zhili, 2020. Multiple resource allocation for precision marketing. In: Journal of Physics: Conference Series, Vol. 1592. IOP Publishing, 012034.

Zhu, Peng, Miao, Chun, Wang, Ziwei, Li, Xiaotong, 2023. Informational cascade, regulatory focus and purchase intention in online flash shopping. Electron. Commer. Res. Appl. 101343.

Zhu, Genglin, Peng, Zixin, Li, Mu, Fan, Jiantao, Lai, Xinjun, 2022. Credit allocation considering loaner's credit risk and willingness for acceptance: A hybrid xgboost-topsis enabled optimization approach. In: 2022 International Conference on Artificial Intelligence, Internet and Digital Economy. ICAID 2022, Atlantis Press, pp. 985–995.