A Hybrid Cost-Sensitive Machine Learning and Optimization Models to Minimize the Resource-Constrained Classification Problem Costs *

Lior Rabkin[1], Matan Marudi[1], Itay Margolin[2], Ilan Cohen[1], and Gonen Singer[1]

[1] Bar-Ilan University, Faculty of Engineering, Ramat-Gan, 52900, Israel
lior.rabkin@live.biu.ac.il
[2] Intuit Inc.

Abstract. Classification tasks aiming to minimize misclassification costs that involve allocation of scarce resources are common in many real-world problems such as allocation of organ transplants to patients, budget allocations for direct advertising, and classification of machines that need maintenance when there is a maintenance capacity limit. We propose a comprehensive analytic framework for scenarios that, in addition to including multi-class classification problems with misclassification costs, also have constraints on the number of classified samples of classes due to resource limitations. To classify samples under the constraints, the framework uses a probability matrix generated by a trained cost-sensitive classifier as the input for an optimization model with a minimum cost objective and resource allocation constraints. To illustrate its effectiveness and applicability, the framework with a cost-sensitive neural network was applied in the context of a medical resources allocation case study. The proposed framework performs significantly better than the alternative common approach with a cost-insensitive classifier. Our results show that the proposed framework is capable of providing an effective limited-resource allocation for misclassification cost problems.

Keywords: misclassification costs · cost-sensitive learning · resource allocation · optimization model · resource constraints

1 Introduction

In many classification tasks, misclassification costs derived from the consequences of errors, are involved [4,25,7]. Usually, misclassification errors in these tasks are asymmetric. In disease severity classification, for example, incorrectly classifying ill patients as healthy may put these patients at risk of losing their life, much heavier consequences than classifying healthy patients as ill, which may lead to unnecessary treatment and waste of medical resources.

* Supported by Intuit Inc.

In resource allocation problems, classification tasks usually face practical constraints regarding the number of predicted samples per class, because of limited resources [13,8,1]. Directly applying the classification model on new data (test data) that are independent of the training data will not guarantee compliance with the resource constraint. In the disease severity classification example, the doctor may have a limited amount of monitoring devices to detect deterioration of the disease (whether a patient is in early or late stage of the disease). Ideally, these devices should be allocated such that a high proportion of patients in the early stages of the disease would be identified. A healthcare center, however, may prioritize identification of patients who are in a later stage of the disease using a known high success rate of recovery to benefit from the efficiency of allocating a limited number of medical resources such as surgical rooms and teams. These types of applications usually assume a multi-class classification problem in which the target exhibits an ordering form, as in the example of disease severity classification task. When there is no accurate information about the misclassification costs' values, the distance between real and predicted classes, assuming ordering information between classes, can be used to reflect the misclassification errors [10,19,20,23,16,11].

This paper proposes a decision-making framework for multi-class classification problem with misclassification costs and resource constraints. Our approach combines a cost-sensitive algorithm and an optimization model that finds the optimal solution of the resource allocation.

2 Background and Literature Review

In this section, we review various approaches suggested by researchers for making classifiers cost sensitive. We then review related research studies that used classification algorithms for the limited resource allocation problem.

2.1 Cost-Sensitive Methods for Misclassification Cost Problems

First, we introduce definitions used throughout the paper. We denote a fully labeled training data set by $D = (X, Y)$. $X \in \mathbb{R}^{|S|\times|M|}$ is a set of samples S, with a set of features M. 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 F, meaning that Y contains $|F|$ different values. $\mathcal{M}$ is a classification model that returns probability matrix $\hat{Y} \in (0,1)^{|S|\times|F|}$, 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., $\hat{y}_s = \arg\max_i P(i|x_s)$.

Real-world classification tasks usually involve misclassification costs that result from different types of misclassification errors. In these classification tasks, which are known as misclassification cost problems, the performance of a classifier is usually assessed by the expected misclassification cost and not by its accuracy, which is a common criterion for classification problems without costs [5]. Over the years, many cost-sensitive methods have been proposed to address classification tasks with different costs incurred as a result of different types of misclassification errors. These methods can be categorized into three main approaches (see [21]):

1. Data level approach: These methods modify the class distribution in the training data without modifying the learning process. Changing the training data usually refers to pre-training modification of the distribution of classes using an over- or under-sampling distribution of classes [5,15] or weighting the samples [27,9].
2. Algorithmic level approach: These methods change the learning process by directly incorporating the misclassification costs into the classifier training so that they are biased towards classes with high misclassification costs [14,24,2].
3. Decision level approach: These methods assign samples to classes based on the insensitive classifier's outputs, such that a minimum expected cost achieved. Such methods usually refer to post-training modification of the outputs by introducing the misclassification costs when a decision on classification of a sample to a class is being made [27,18].

Assume a misclassification cost matrix $C \in \mathbb{R}^{|F|\times|F|}$, where $C_{i,j}$ is the cost associated with incorrectly classifying a sample to class i instead of the actual label j. According to the objective of misclassification cost problem, a sample $x_s \in X$ should be classified into the class i that minimizes the expected loss (cost in our case) at Eq. 1:

$$\mathcal{L}\left(x_s,\ i\right) = \sum_j P(j|x_s) C_{i,j}. \tag{1}$$

Thus, the optimal predicted class that minimizes the expected loss is $\hat{y}_s = \arg\min_i \mathcal{L}(x_s,\ i)$. The decision level approach assigns samples to classes based on the probability outputs $\hat{Y}$ of an insensitive classifier such that the loss function $\mathcal{L}$ is minimized. In our research, we propose a hybrid framework that combines both an algorithmic level approach and a decision level approach. Thus, the probability outputs of cost-sensitive classifiers are used as inputs for the optimization model that minimizes the loss function $\mathcal{L}$.

2.2 Machine Learning and Optimization Models for Classification Tasks with Constraints

Recently, several research studies proposed integrating machine learning methods and mathematical modeling formulation to solve classification tasks with limited resource allocation. Usually, the machine learning classification phase generates local probability outputs for each sample in the data set reflecting the probabilities of successful classification for every class, while the mathematical model seeks to optimize an objective function, while meeting global constraints and decision maker requirements. In [17], a framework that integrates interpretable machine learning models together with mathematical programming was proposed for handling allocation of applications to a limited number of open positions, while aiming to achieve high recruitment success (candidate qualification and match with low probability of turnover). The machine learning model was used to generate probabilities of successful recruitment per employee and position; the mathematical modeling took into account the constraint (limited number of open positions) and the diversity of the workforce among different departments. Another study [13] proposed a resource allocation scheme for classification of workload tasks with similar resource requirements. Here the aim was to efficiently allocate available resources while reducing energy consumption by minimizing the number of active physical machines. A solution that combines classification models with an optimization model to allocate limited resources to identify customer groups with different characteristics for marketing of a new product application while minimizing regret is proposed in [26]. In [6], a combination of a machine learning algorithm and optimization model is proposed to improve the efficiency of hospital operating rooms. The researchers used the machine learning algorithms to predict the uncertain recovery duration of surgical patients, which was then used as input for an optimization model meant to calculate optimal scheduling. The research study of [8] proposed an approach that introduces diagnosis-related group classifications into a mixed-integer programming-based resource allocation model that increase the utilization of resources such as operating rooms and beds.

None of the previous studies sought to solve misclassification cost and limited resource allocation problems. Recent pioneering research did develop an adaptive learning approach that considers resource constraints and misclassification costs for the binary classification problem [1]. The solution of the resource allocation of the optimization model for the binary case, however, is trivial, and thus so are the resource allocation decisions determined by the classifier's probability matrix.

The main contributions of the paper are:

1. *Methodology* We suggest a hybrid framework for solving multi-class classification problems with misclassification costs and resource constraints. The hybrid framework utilizes both algorithmic level and decision level approaches.
2. *Modeling* The suggested model can handle settings in which the resource's demand changes over time. The resource allocation literature has, in general, focused on optimizing specific objectives in a static environment where the resource's demand is known. In an enormous amount of real-world cases, however, uncertainty is inherently involved, and the resource's demand can change over time. For example, the number of patients who come to an emergency department and regarding whom staff have to decide whether or not to hospitalize, changes over time.
3. *Practicability* A case study of medical resource allocation with misclassification costs demonstrates the applicability of the proposed hybrid framework and its superiority over the common decision level approach, which uses a cost-insensitive classifier with an optimization model.

The rest of the paper is organized as follows. In Section 3, we introduce the proposed framework for solving a classification problem with misclassification costs and resource constraints. Section 4 presents the numerical experiments and results when evaluating and benchmarking the proposed framework against the commonly used approach. Section 5 concludes the paper.

3 Methods

3.1 Hybrid Framework for a Classification Problem with Misclassification Costs and Resource Constraints

This section introduces a hybrid cost-sensitive (Hyb_CS) framework for minimizing the misclassification costs of the resource-constrained classification problem, by utilizing both decision level and algorithmic level approaches. A key property of our framework is the generation of the probability matrix $\hat{Y}$ by a cost-sensitive classifier (algorithmic level approach), that can be used as input for our optimization model (decision level approach), aiming to achieve minimum costs while satisfying the resource constraints. Figure 1 presents the proposed framework consisting of two phases. In the first phase, a cost-sensitive machine learning algorithm (CS_ML) is trained using a training data set. In each epoch, the generated models are applied on a validation data set. The output of this phase, the probability matrix, is used in the second phase as input for the optimization model (decision level approach). The optimization model finds the predicted label of each sample, aiming to achieve minimum cost while satisfying the resource constraints. We select a cost-sensitive classifier based on the costs achieved after the optimization model phase on the validation data, to be applied over the test data as presented in Figure 2.

Section 3.2 presents a cost-sensitive neural network that is used in the experimental study. A detailed description of the optimization model is described in Section 3.3.

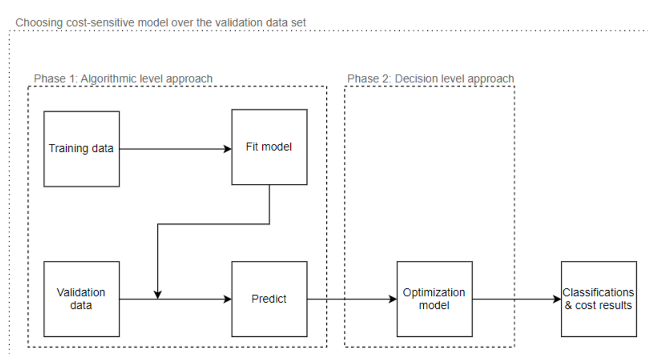


Fig. 1: Schematic of the Hyb_CS training process framework

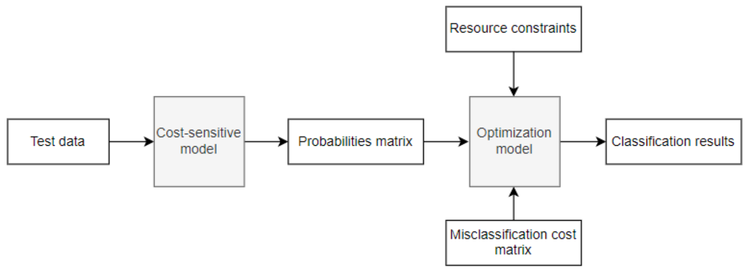


Fig. 2: Schematic process of applying the chosen cost-sensitive model to test data

3.2 Cost-Sensitive Neural Network

Many research studies used an algorithmic level approach to introduce the misclassification costs into the training of a neural network, in order to reduce the cost of misclassified samples or to handle the class imbalanced problem [22,12,2,3]. In our research, we use the proposed neural network with ordinal loss from [3] as a cost-sensitive learning model in phase 1 of the Hyb_CS framework (see Fig 1. An ordinal loss is proposed with a penalty cost matrix representing the costs predicting a sample belongs to class i when the observed class is j. The ordinal loss approach was used in [3] for knee osteoarthritis severity classifications, which is also used in our paper as explained in Section 4. Using the cost matrix C, as the ordinal matrix proposed in [3], and the probabilities matrix accepted by the softmax layer, the ordinal loss is presented in Eq. 2,

$$OL = \sum_{\forall x_s} \sum_{\forall i \in F} C_{i,y_s} \hat{Y}_{s,i}, \tag{2}$$

This ordinal loss function takes into account the estimated probabilities of the samples for each one of the possible classes $\hat{Y}$ and the misclassification costs associated with these predictions compared to the actual labels.

3.3 Resource-Constrained Optimization Model

We now describe the optimization model used in phase two of the proposed Hyb_CS framework shown in Figure 1. The model receives the probability matrix $\hat{Y}$ generated from the cost-sensitive learning model trained in the first phase of the proposed framework and a misclassification cost matrix $C \in \mathbb{R}^{|F| \times |F|}$. Our assumption is that available resources for each class $i \in F$, $\mathcal{B}_i$, limits the number of samples in the considered data set S' (test data set) classified as belonging to this class. The learning process is conducted on the training data set, and the model is validated on a validation data set that usually has different number of samples than the test data set. Thus, we project the resource constraints of the test data set onto the training and validation data sets by calculating the maximal percent of all classifications of the considered data set related to class i, i.e., $n_i = \frac{\mathcal{B}_i}{|S'|}$. In this way, we use the resource constraint for each class i for the training and validation processes as a product of the data set size and n_i. We define a constraint vector $N \in (0,1)^{|F|}$ as input to the optimization model, where these n_i's values relate to all classes.

The output of the optimization model is $R \in \{0,1\}^{|S|\times|F|}$, a binary matrix, where $R_{s,i} = 1$ represents a sample s classified as belonging to class i. The model forces each sample to be classified to one class. Table 1 presents the notations using in describing the problem.

Notation	Size	Meaning
S	|S|	Set of samples
F	|F|	Set of classes
C	|F|×|F|	Cost matrix
N	|F|×|1|	Constraint vector
$\hat{Y}$	|S|×|F|	Probability matrix
R	|S|×|F|	Result vector

Table 1: Notations in the resource-constrained optimization model

We now can model the resource allocation problem as follows:

$$\min \quad \sum_{s=1}^{|S|}\sum_{i=1}^{|F|} R_{s,i} \sum_{j=1}^{|F|}(C_{i,j} \cdot \hat{Y}_{s,j}) + \gamma \cdot \sum_{i=1}^{|F|} v_i \cdot \sum_{s=1}^{|S|} R_{s,i} \tag{3}$$

$$\text{s.t.} \quad \sum_{s=1}^{|S|} R_{s,i} \leq n_i \cdot |S|, \quad \forall i \in F \tag{4}$$

$$\sum_{i=1}^{F} R_{s,i} = 1, \quad \forall s \in S \tag{5}$$

Our goal is to minimize a cost function comprising two parts. The first one represents the misclassification costs and the second one expresses the costs associated with resource allocation (e.g., time spent by a resource, price of resource allocation). The price of assigning a resource to a sample that is classified to label i is denoted as v_i. We use $\gamma \in [0,1]$ to balance the two cost function components. We define two constraints in the optimization model. The first constraint in Eq. 4 ensures that the number of samples that are classified as label i is lower than the resource constraint, reflected by percent of the total number of classifications. The second constraint in Eq. 5 ensures that each sample is assigned only one label. An optimal solution for the R matrix is achieved by minimizing the objective function in Eq. 3, while satisfying the constraints.

4 Experimental Study

In the following, we provide an experimental investigation of the proposed Hyb_CS, and compared its results with the common approach of using an insensitive classifier with a decision level approach (denoted herein as Dec_CS). We use a VGG-19 neural network with cross-entropy loss function as a cost-insensitive classifier and a VGG-19 neural network with ordinal loss function, described in Section 3.2, as the cost-sensitive classifier in the first phase of the proposed hybrid framework.

To understand the reasons for the different performances of the proposed hybrid framework verus the common approach, we investigate: 1) the change in classifications and cost improvement obtained from the machine learning model in the first phase, by the optimization model without the resource constraints, 2) the change in classifications and cost improvement obtained from the machine learning model in the first phase, by the optimization model with resource constraints, and 3) the effect of the constraint level on the classifications and costs.

4.1 The Data

Hyb_CS was applied to a publicly available data set obtained from the Osteoarthritis Initiative (OAI), and available at https://nda.nih.gov/oai/. Knee Osteoarthritis (OA) is the most common cause of activity limitation and physical disability among older adults. Early detection and treatment may slow down its progression so streamlining these medical procedures are very important. A total of 4796 images of knee bilateral posterior-anterior fixed flexion radiographs of 4796 participants exist in the data set. In our research, we used a total of 4130 pairs of knee joints, acquired from the OAI repository and made available by [3] after a prepossessing stage. All knee X-ray images were randomly divided into training, validation, and test data sets at a ratio of 7:1:2. The goal of this experiment was to classify the patients into five classes representing different OA severity levels and allocate the patients to limited treatment resources, according to classification results. We assume that assigning a patient to a wrong treatment leads to a misclassification cost.

4.2 Experimental Design

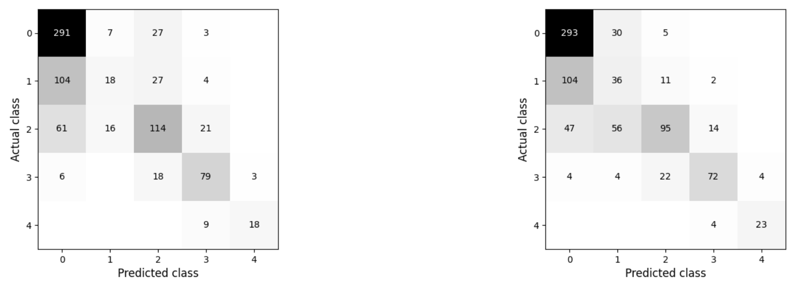
We used the proposed hybrid framework in two experiment setups with different objectives. In these experiments, we assume that prices for resources assignment do not exist, $v_i = 0, \forall i \in F$ and used a symmetric and ordinal cost matrix C, with $\forall i, j : C_{i,j} = 2 \cdot |j - i| + 1$, such that larger deviation between actual and predicted classes reflect errors with more serious consequences. For consistency, we used the same 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 previous epochs and without a cost improvement higher than 2.5% in the next five epochs.

Setup 1 – Misclassification Costs Problem Without Resource Constraints This experiment was designed so that we could investigate the behavior and performance of Hyb_CS, combining both decision level and algorithmic level approaches, compared to the common approach of applying only a decision level approach on the probability matrix accepted by the cost-insensitive learning model. The goal of this experiment was to classify the patients into OA severity levels such that minimum misclassification costs will be expended, while assuming unlimited resources.

Setup 2 – Misclassification Costs Problem with Resource Constraints This experiment was designed to investigate the behaviour and performance of Hyb_CS compared to the common approach, while integrating the resource constraints into the optimization model. Scarce resources allocation is supposed to increase the cost compared to the unlimited resources problem. In this experiment we investigated whether using the cost-sensitive learning model (instead of the cost-insensitive learning model) reduces the amount of changes in the classifications obtained from the machine learning by the optimization model and leads to better results. In this setup, we used two different scenarios. The first scenario had a constraint on the number of treatments for patients with severity level 3, $\mathcal{B}_3 = 50$, i.e., $n_3 = 3\%$ assuming $|S'| = 1656$ patients in the considered data set (test data set). The second scenario has a constraint on the number of treatments for patients with severity level 4, $\mathcal{B}_4 = 8/17/25$, i.e., $n_4 = 0.5\%/1\%/1.5\%$.

4.3 Results and Discussion

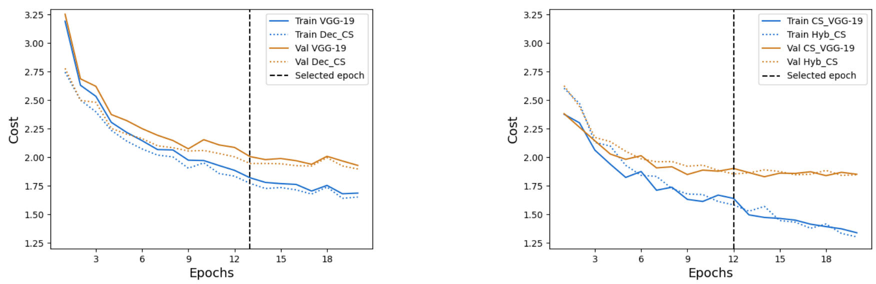
Setup 1 – Misclassification Costs Problem Without Resource Constraints The probability matrices of the Cost-Sensitive VGG-19 (CS_VGG-19) and VGG-19 were used to assign the images to the class with the highest probability, following the maximum likelihood principle. Figure 3 presents the confusion matrices of the actual labels versus the predicted labels for the applied classifiers. It can be observed that when using the CS_VGG-19, the amount of misclassifications decreases in each row when moving away from the diagonal, a phenomenon that does not occur when applying the VGG-19. For example, it can be seen in the left confusion matrix in Figure 3 that VGG-19 classified 61 samples as class 0 instead of the actual class, class 2, compared to 16 samples classified as class 1 instead of the actual class, class 2.



(a) VGG-19 (b) CS_VGG-19

Fig. 3: Confusion matrices of the actual labels versus the predicted labels for VGG-19 (left) and CS_VGG-19 (right)

The misclassification cost of the training and validation data sets run over 20 epochs is presented in Figure 4 after the classifier phase (VGG-19 or CS_VGG-19) and after applying the optimization model. As expected, when applying a VGG-19 classifier (insensitive-cost learner), a cost improvement was observed in the left graph for both training and validations data sets, after applying the optimization model (decision level approach), which aimed to minimize the misclassification cost using the obtained probability matrix. When applying the CS_VGG-19 classifier (algorithmic level approach), however, the optimization model did not contribute to a further improvement in the cost result. In this experiment with unlimited resources, the algorithmic level approach yields better cost results than the decision level approach. Applying the optimization model (decision level approach) to the probability matrix obtained from the cost-sensitive learner does not improve the cost results. The black dashed lines in the graphs represent the selected epochs according to the stopping mechanism described in Section 4.2. The trained models at these epochs were saved and used for the test data sets as described in Figure 2.



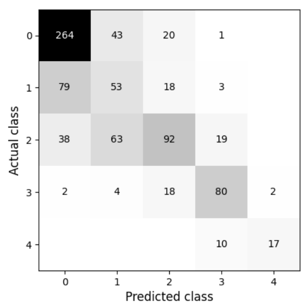
(a) VGG-19 compared to Dec_CS (b) CS_VGG-19 compared to Hyb_CS

Fig. 4: The misclassification cost of the training and validation data sets run over 20 epochs after the classification phase (VGG-19 or CS_VGG-19) and after applying the optimization model (Dec_CS or Hyb_CS). The black dashed lines represent the stopping epoch.

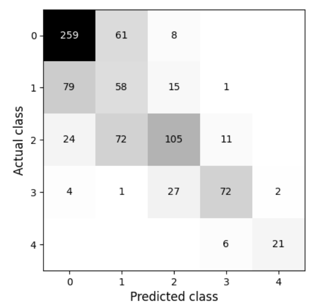
Setup 2 – Misclassification Costs Problem with Resource Constraints In this setup, since resource constraints exist, an optimization model is applied to achieve minimum misclassification costs while satisfying the resource constraints. Whereas in the previous setup, the optimization model does not contribute to a further improvement in the cost of CS_VGG-19 and, therefore, its use is superfluous, in this setup the optimization model must be applied to ensure compliance with the resource constraints.

A constraint on the number of treatments for patients with OA severity level 3 Figure 5 presents the confusion matrices of the actual labels versus the predicted labels for the common decision level approach with VGG-19 and Hyb_CS, 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).

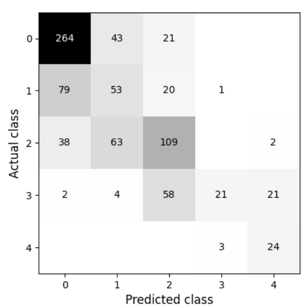
In Figure 5, the fourth column (representing severity level 3) of the lower matrices sums to 25, which is equal to 3% of the validation data set. In both the common decision level approach and Hyb_CS, the reduction in the number of correctly predicted samples was similar, i.e., 35 and 39 respectively, since samples were moved from predictions of severity level 3 to other severity levels, to meet the resource constraint. In other words, since the Hyb_CS cost results was better than the common decision level approach, for the misclassification cost problem without resource constraints, it is also yielded better results for the problem with the resource constraint.



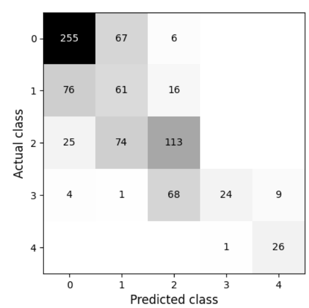
(a) Dec_CS



(b) Hyb_CS



(c) Dec_CS with $n_3 = 3\%$



(d) Hyb_CS with $n_3 = 3\%$

Fig. 5: Confusion matrices of the actual labels versus the predicted labels for Dec_CS and Hyb_CS, without a resource constraint (upper matrices) and when applying a 3%- constraint on severity level 3 (lower matrices)

Figure 6 shows the percentage of samples whose cost, derived from the VGG-19 and CS_VGG-19 classifications, which were improved, left unchanged, or worsened by the optimization model, denoted 'pos', 'neg' and 'equal', respectively. Note that the 'equal' category includes samples whose classifications were left unchanged and others that were changed by the optimization model; the latter change, however, resulted in the sample's classification having the same distance from the actual label as before the change. The two left bars represent the VGG-19 and CS_VGG-19 results for the misclassification costs problem without constraints. The two right bars present the results for the misclassification cost problem with resource constraints. It can be observed that the CS_VGG-19 'equal' parts are higher than the VGG-19 'equal' parts in both cases, i.e., with and without the constraints. This insight is rational since CS_VGG-19 takes into account the misclassification costs already in the learning phase. As expected in the resource constraint problem, the percentage of 'neg' samples was higher than the percentage of 'pos' examples in contrast to the unconstrained problem, because classification changes are required to satisfy the resource constraint, even at the expense of increasing the cost.

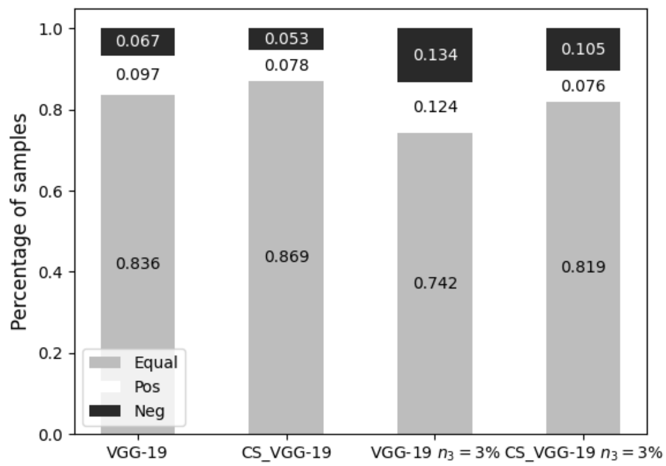


Fig. 6: Percentage of samples whose cost, derived from the VGG-19 and CS_VGG-19 classifications , were improved, left unchanged, or worsened by the optimization model, denoted by 'pos', 'neg' and 'equal', respectively. The two left bars present the problem without resource constraints and the two right bars with the resource constraints.

In each experiment, we selected the classification model according to the epoch selection mechanism described in Section 4.2 and applied these models over the test data sets to the yield costs (and accuracy) presented in Table 2. It can be seen that applying Hyb_CS to the misclassification cost problems, both with and without constraints, yielded better cost results with improvements of 4.2% and 4.5%, respectively, and improved accuracy results of 3.5% and 3.2%, respectively.

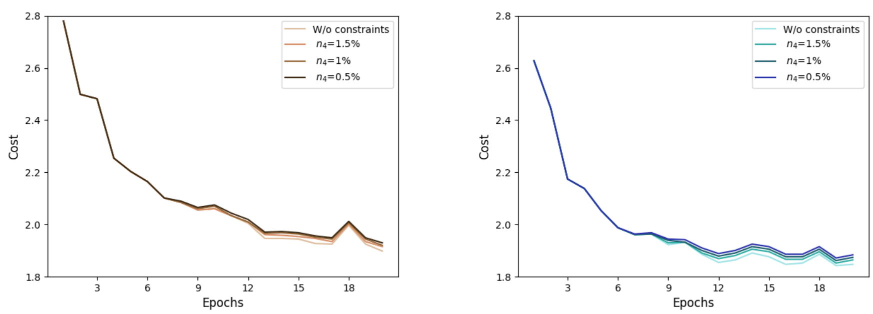
	W/o constraints		$n_3 = 3\%$	
Model	Dec_CS	Hyb_CS	Dec_CS	Hyb_CS
Cost	1.883	1.798	1.984	1.9
Accuracy	0.627	0.647	0.577	0.597

Table 2: Cost and accuracy results on test data set applying Dec_CS and Hyb_CS to misclassification cost problems, both without and with resource constraints for patients with severity level 3

Effect of different levels of constraints on the number of available treatments
To evaluate the effect of the constraint on the Hyb_CS cost results, we calculated the misclassification costs for different constraint levels, $n_4 = 0.5\%/1\%/1.5\%$. It can be observed from Table 3 that the tighter the constraint, the greater the number of classification's changes performed by the optimization model. The misclassification cost of the validation data set over 20 epochs after applying a classification model in the first phase and after running an optimization model in the second phase for the different constraint levels is presented in Figure 7. It can be observed that the tighter the constraint, the higher the cost.

Model	W/o constraints	$n_4 = 1.5\%$	$n_4 = 1\%$	$n_4 = 0.5\%$
Dec_CS	0.163	0.171	0.176	0.18
Hyb_CS	0.131	0.143	0.148	0.151

Table 3: Mean number of classification's changes performed by the optimization model after applying Dec_CS and Hyb_CS for the different constraint levels



(a) Dec_CS (b) Hyb_CS

Fig. 7: The misclassification cost of the validation data set over 20 epochs after applying Dec_CS and Hyb_CS for the different constraint levels

Table 4 presents the costs and accuracy results obtained by applying the selected classification models in Dec_CS and Hyb_CS over the test data for different levels of resource constraints. It can be seen that applying Hyb_CS with different levels of resource constraints show improvements (between 3.6%-4.6%) in the cost results compared to the Dec_CS approach. Hyb_CS improvements in the accuracy results compared to the Dec_CS approach are in the range of 1.3%-3.2%.

	$n_4 = 0.5\%$		$n_4 = 1\%$		$n_4 = 1.5\%$	
Model	Dec_CS	Hyb_CS	Dec_CS	Hyb_CS	Dec_CS	Hyb_CS
Cost	1.9	1.832	1.892	1.809	1.886	1.802
Accuracy	0.618	0.626	0.623	0.641	0.625	0.645

Table 4: Cost and accuracy results for test data set after applying Dec_CS and Hyb_CS to misclassification cost problems with resource constraint levels of 0.5%/1%/1.5% for patients with severity level 4

5 Conclusion

We propose a novel hybrid framework that combines both algorithmic level and decision level approaches. The objective is to solve multi-class misclassification costs problems with resource constraints. We demonstrated via experiments that the optimization model did not contribute to a further improvement in the cost value when using a cost-sensitive classification model when resource constraints do not exist and, therefore, the model is superfluous. When resource constraints exist, the optimization model should be used to take into account the resource constraints. In this case, the proposed hybrid framework combining both a cost-sensitive learning algorithm and an optimization model yields better results than the common decision level approach of using a cost-insensitive learning algorithm with an optimization model. Our experimental results show that the proposed hybrid framework outperformed the common approach with a cost reduction in the range of 3.6%-4.6% for the misclassification problem with different levels of resource constraints, and 4.5% without resource constraints on the test data set. Similar accuracy improvements were observed in the range of 1.3%-3.5% for the misclassification problem with different levels of resource constraints, and 3.2% without resource constraints on the test data set. Future research directions that we are considering are: 1) utilizing the classification changes performed by the optimization model to implement adaptive learning of the cost-sensitive learning algorithm. 2) Investigate different combinations of cost-sensitive approaches such as data level and decision level approaches.

References

1. Danit Shifman Abukasis, Izack Cohen, Xiaochen Xian, Kejun Huang, and Gonen Singer. Adaptive learning for the resource-constrained classification problem. *arXiv preprint arXiv:2207.09196*, 2022.
2. Yuri Sousa Aurelio, Gustavo Matheus de Almeida, Cristiano Leite de Castro, and Antonio Padua Braga. Learning from imbalanced data sets with weighted cross-entropy function. *Neural Processing Letters*, 50(2):1937–1949, 2019.
3. Pingjun Chen, Linlin Gao, Xiaoshuang Shi, Kyle Allen, and Lin Yang. Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss. *Computerized Medical Imaging and Graphics*, 75:84–92, 2019.

4. Koen W De Bock, Kristof Coussement, and Stefan Lessmann. Cost-sensitive business failure prediction when misclassification costs are uncertain: A heterogeneous ensemble selection approach. *European Journal of Operational Research*, 285(2):612–630, 2020.
5. Charles Elkan. The foundations of cost-sensitive learning. In *International Joint Conference on Artificial Intelligence*, volume 17, pages 973–978. Lawrence Erlbaum Associates Ltd, 2001.
6. Michael Fairley, David Scheinker, and Margaret L Brandeau. Improving the efficiency of the operating room environment with an optimization and machine learning model. *Health care management science*, 22(4):756–767, 2019.
7. Flavia Dalia Frumosu, Abdul Rauf Khan, Henrik Schiøler, Murat Kulahci, Mohamed Zaki, and Peter Westermann-Rasmussen. Cost-sensitive learning classification strategy for predicting product failures. *Expert Systems with Applications*, 161:113653, 2020.
8. Daniel Gartner, Rainer Kolisch, Daniel B Neill, and Rema Padman. Machine learning approaches for early drg classification and resource allocation. *INFORMS Journal on Computing*, 27(4):718–734, 2015.
9. Nazeeh Ghatasheh, Hossam Faris, Ismail AlTaharwa, Yousra Harb, and Ayman Harb. Business analytics in telemarketing: cost-sensitive analysis of bank campaigns using artificial neural networks. *Applied Sciences*, 10(7):2581, 2020.
10. Pedro Antonio Gutiérrez, Maria Perez-Ortiz, Javier Sanchez-Monedero, Francisco Fernandez-Navarro, and Cesar Hervas-Martinez. Ordinal regression methods: survey and experimental study. *IEEE Transactions on Knowledge and Data Engineering*, 28(1):127–146, 2015.
11. Deniu He. Active learning for ordinal classification based on adaptive diversity-based uncertainty sampling. *IEEE Access*, 2022.
12. Salman H Khan, Munawar Hayat, Mohammed Bennamoun, Ferdous A Sohel, and Roberto Togneri. Cost-sensitive learning of deep feature representations from imbalanced data. *IEEE transactions on neural networks and learning systems*, 29(8):3573–3587, 2017.
13. Hongjian Li, Shiwang Ding, Pengfei Zhang, and Jun Lai. 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)*, pages 111–119. Atlantis Press, 2018.
14. Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.
15. Charles X Ling and Victor S Sheng. Cost-sensitive learning and the class imbalance problem. *Encyclopedia of machine learning*, 2011:231–235, 2008.
16. Matan Marudi, Irad Ben-Gal, and Gonen Singer. A decision tree-based method for ordinal classification problems. *IISE Transactions*, (just-accepted):1–20, 2022.
17. D. Pessach, G. Singer, D. Avrahami, H. C. Ben-Gal, E. Shmueli, and I. Ben-Gal. Employees recruitment: A prescriptive analytics approach via machine learning and mathematical programming. *Decision Support Systems*, 1132, 2020.
18. Victor S Sheng and Charles X Ling. Thresholding for making classifiers cost-sensitive. In *AAAI*, volume 6, pages 476–481, 2006.
19. Gonen Singer, Roee Anuar, and Irad Ben-Gal. A weighted information-gain measure for ordinal classification trees. *Expert Systems with Applications*, 152:113375, 2020.
20. Gonen Singer, Anat Ratnovsky, and Sara Naftali. Classification of severity of trachea stenosis from eeg signals using ordinal decision-tree based algorithms

and ensemble-based ordinal and non-ordinal algorithms. *Expert Systems with Applications*, 173:114707, 2021.
21. Ohad Volk and Gonen Singer. Adaptive cost-sensitive learning in neural networks for misclassification cost problems. *arXiv preprint arXiv:2111.07382*, 2021.
22. Shoujin Wang, Wei Liu, Jia Wu, Longbing Cao, Qinxue Meng, and Paul J Kennedy. Training deep neural networks on imbalanced data sets. In *2016 international joint conference on neural networks (IJCNN)*, pages 4368–4374. IEEE, 2016.
23. Hao Wu, Xi Zhang, and Jufeng Yang. Deep learning-based encrypted network traffic classification and resource allocation in sdn. *Journal of Web Engineering*, pages 2319–2334, 2021.
24. Zhenyu Wu, Yang Guo, Wenfang Lin, Shuyang Yu, and Yang Ji. A weighted deep representation learning model for imbalanced fault diagnosis in cyber-physical systems. *Sensors*, 18(4):1096, 2018.
25. Huan Zhang, Liangxiao Jiang, and Chaoqun Li. Cs-resnet: Cost-sensitive residual convolutional neural network for pcb cosmetic defect detection. *Expert Systems with Applications*, 185:115673, 2021.
26. Siyu Zhang, Peng Liao, Heng-Qing Ye, and Zhili Zhou. Multiple resource allocation for precision marketing. In *Journal of Physics: Conference Series*, volume 1592, page 012034. IOP Publishing, 2020.
27. H. Zhao. Instance weighting versus threshold adjusting for cost-sensitive classification. *Knowledge and Information Systems*, 15(3):321–334, 2008.