

# Dynamically weighted Evolutionary Ordinal Neural Network for solving an Imbalanced Liver Transplantation Problem

M. Dorado-Moreno<sup>a,d</sup>, M. Pérez-Ortiz<sup>b</sup>, P.A. Gutiérrez<sup>a</sup>, R. Ciria<sup>c</sup>, J. Briceño<sup>c</sup>, C. Hervás-Martínez<sup>a</sup>

<sup>a</sup>Department of Computer Science and Numerical Analysis, University of Córdoba, Córdoba, Spain

<sup>b</sup>Department of Quantitative Methods, Universidad Loyola Andalucía, Córdoba, Spain

<sup>c</sup>Liver Transplantation Unit, Reina Sofía Hospital, Córdoba, Spain

<sup>d</sup>Corresponding author. Tel 957 218 349. E-mail: i92domom@uco.es

---

## Abstract

**Objective:** Create an efficient decision-support model to assist medical experts in the process of organ allocation in liver transplantation. The mathematical model proposed here uses different sources of information to predict the probability of organ survival at different thresholds for each donor-recipient pair considered. Currently, this decision is mainly based on the Model for End-stage Liver Disease, which depends only on the severity of the recipient and obviates donor-recipient compatibility. We therefore propose to use information concerning the donor, the recipient and the surgery, with the objective of allocating the organ correctly.

**Methods and materials:** The database consists of information concerning transplants conducted in 7 different Spanish hospitals and the King's College hospital (United Kingdom). The state of the patients is followed up for 12 months. We propose to treat the problem as an ordinal classification one, where we predict the organ survival at different thresholds: less than 15 days, between 15 and 90 days, between 90 and 365 days and more than 365 days. This discretization is intended to produce finer-grain survival information (compared with the common binary approach). However, it results in a highly imbalanced dataset in which more than 85% of cases belong to the last class. To solve this, we combine two approaches, a cost-sensitive evolutionary ordinal artificial neural network (in which we propose to incorporate dynamic weights to make more emphasis on the worst classified classes) and an ordinal over-sampling technique (which adds virtual patterns to the minority classes and thus alleviates the imbalanced nature of the dataset).

**Results:** The results obtained by our proposal are promising and satisfactory, considering the overall accuracy, the ordering of the classes and the sensitivity of minority classes. In this sense, both the dynamic costs and the over-sampling technique improve the base results of the considered ANN-based method. Comparing our model with other state-of-the-art techniques in ordinal classification, competitive results can also be appreciated.

**Conclusions:** The combination of the proposed cost-sensitive evolutionary algorithm together with the application of an over-sampling technique improves the predictive capability of our model in a significant way (especially for minority classes), which can help the surgeons make more informed decisions about the most appropriate recipient for an specific donor organ, in order to maximize the probability of survival after the transplantation and therefore the fairness principle.

**Keywords:** Liver transplantation, Ordinal Classification, Survival Analysis, Artificial Neural Networks, Imbalanced Classification

---

## 1. Introduction

Liver transplantation is an accepted treatment for patients who present end-stage liver disease. However, transplantation is restricted by the lack of suitable donors, where this imbalance between supply and demand results in significant waiting list deaths. With the objective of approaching this problem, several techniques have been proposed to find a better system to prioritize recipients on the waiting list. The first attempt at developing a system is the Donor Risk Index (DRI) [1], which establishes the quantitative risk associated to the surgery considering only donor information. The opposite methodology, and the pillar of the current allocation policy, is the Model for End-stage Liver Disease (MELD) [2], which only considers the severity of the recipient.

Nonetheless, the above-mentioned methods can not be con-

sidered good predictors of graft failure after transplantation, because they only take into account either characteristics of donors or recipients (but not both). Rana *et al.* [3] created a scoring system (SOFT) to predict recipient survival three months after liver transplantation using information of both donor and recipient. P. Dutkowski *et al.* recently proposed a balance of risk score (BAR) [4] also based on donor and recipient characteristics. A rule-based system is used to determine graft survival one year after transplantation in [5, 6] using donor, recipient and transplanted organ characteristics. These studies show that artificial neural networks perform better at this task than the rest of scores, showing very good predictive capabilities at a higher time threshold. In this sense, new trends in biomedicine have considered the use of machine learning to approach different problems related to patient survival [7, 8, 9, 10] resulting in remarkable applications for science

and medicine [11, 12], which motivates their study in future work. Note that generally, donors are assigned to the candidates under the greatest-risk according to MELD. Clearly, this policy does not allow the liver transplant team to match the donor to the recipient according to principles of fairness and benefit, which could lead to a risk of unconscious gaming when trying to match marginal donors to urgent candidates [13].

This paper considers a liver transplantation dataset with data from 1406 surgeries performed in seven Spanish transplant units and the King’s College Hospital (UK). As stated before, this dataset includes characteristics of donors, recipients and surgery, with the aim of developing and constructing a global system for predicting graft survival, by means of artificial intelligence techniques. The classification problem is addressed using an ordinal classification point of view since the classes are ordered taking into account the time leading up to liver failure (in case of failure). The classes considered for the dataset are: 1) graft loss before the first 15 days, 2) loss between 15 days and 89 days, 3) loss between 90 days and 365 days, and 4) graft survival after 365 days. These intervals have been highlighted by experts as being the most pertinent in early graft loss. Several issues are usually taken into account in order to exploit the presence of this order when applying ordinal classification techniques. These issues can be all simplified in the following premise: the class ordering should be considered in all the processes for constructing the classification model, which involves the classification training strategy, the metrics to evaluate the performance of the classifiers and even the preprocessing of the dataset.

To clarify the difference between standard classification and ordinal classification, consider the problem of classifying villi injuries (Marsh classification) given the labels: {normal state, mild atrophy, marked atrophy, complete atrophy}. Clearly, an order among the categories can be appreciated, and there are also some misclassification errors that should be more penalized. For example, misclassifying a complete atrophy with a normal state should be far more penalized than misclassifying it as a marked atrophy. Since this is a common learning issue, several approaches have been proposed to tackle this paradigm in the domain of pattern recognition and machine learning over the years, the first work applying logistic regression dating back to 1980 [14]. This issue has generally been addressed by transforming ordinal scales into numeric values and solving the problem as one of standard regression or multinomial classification. However, as said, there are several problems within this approach: on the one hand, the fact that, without a priori knowledge, the distance between different classes is unknown, thus the assumption of equidistant labels when performing standard regression may not hold; on the other hand, as nominal classification does not consider this order information, misclassification errors are treated equally. Nonetheless, other works have approached the paradigm considering the order information by means of threshold methods [15, 16, 17], which are based on the idea that some underlying real-valued outcomes exist (also called latent variable), although they are unobservable.

The liver transplantation problem tackled in this paper presents an ordinal and highly imbalanced nature, which leads

to a difficulty in classifying minority classes, which are, in fact, the most important ones for this application (the ones associated to organ failure). Imbalanced classification has been the focus of many machine learning researchers in the past years, given the hindrance that it usually poses for the learning machine. This paradigm is usually approached using a cost-sensitive approach (i.e. modifying the classification method) or using a resampling strategy (over-sampling the minority class or under-sampling the majority one). In our case, because of the complexity of the dataset, we propose to combine these two common approaches in imbalanced learning: a cost-sensitive method and an over-sampling technique (both of which are focused on ordinal classification problems). More specifically, a new fitness function is proposed for an evolutionary ordinal classification algorithm [18], which dynamically updates the weights of the classes considering the worst classified class in each generation. This classifier is used in conjunction with a recently proposed algorithm for data over-sampling in ordinal domains [19]. Note that an ordinal version of this dataset has also been considered in a previous research [13]. However, in this paper, we improve the performance of our previously developed method by proposing a new evolutionary Artificial Neural Network (ANN) technique, complement the study with an additional experiment and study the best model obtained in the experiments.

The rest of the paper is organized as follows: Section 2 presents the materials used for this paper, introducing evolutionary ANNs and the dataset considered. Then, Section 3 explains the article proposal, including the model and the dynamic costs used, as well as the methods used for ordinal over-sampling. Section 4 presents and discusses the database and the experimental results. Finally, the last section summarizes the main contributions of the paper, outlining some conclusions.

## 2. Materials

This section presents the dataset used and provides a brief introduction about evolutionary ANNs.

### 2.1. Dataset description

A multi-centred retrospective analysis was made of 7 Spanish Liver Transplant units. Recipient, donor and surgery characteristics were reported at the time of transplant. Patients undergoing partial, split or living-donor liver transplantation and patients undergoing combined or multi-visceral transplants were excluded from the study. Liver Transplantation units were homogeneously distributed throughout Spain. The dataset constructed has 634 patterns (donor-recipient pairs) corresponding to the years 2007 and 2008. **The proportion of combined transplant was 2.3% in both cohorts. The proportion of partial grafts was 0,9% and 9,1% in the Spanish and British cohort, respectively. The few cases of combined transplantation was liver and kidney which, in several series, has not been reported to decrease the outcome of the liver graft. We decided to include the partial graft but not the living donor grafts because of two main reasons. Firstly, living donation is performed in a**

very well-selected manner, with excellent donors and with no cold ischaemia time. These grafts could be a tremendous bias. Besides, only Kings College and not the Spanish cohort performs right-lobe adult-to-adult liver transplantation. Secondly, partial adult grafts (pediatrics were excluded) have very similar results to whole grafts (see European Liver Transplantation Registry; ELTR 2016). The only ones with worse outcomes are the full-left-full-right splits, which are not included in these cohorts (only extended rights). Probably, combined liver transplants (which, in fact, are very few) may not be ruled by any allocation program, as they have specific D-R matching rules. However, extended right lobe splits should account as whole grafts and, thus, be allocated accordingly.

In addition, the dataset was completed with information about donor-recipient pairs from the King’s College Hospital (London), to perform a supranational study of donor-recipient allocation in liver transplantation. To obtain a similar number of patterns, only reported pairs of recipients over eighteen years of age between January 2002 and December 2010 were included. Thus, a dataset containing 858 English donor-recipient pairs was collected. In order to merge the datasets, several variables were selected, 16 recipient variables, 17 donor variables and 5 surgically related variables, as shown in Table 1.

All patients were followed from the date of transplant until death, graft loss or completion of the first year after the liver transplant. The final dataset was comprised of 1406 patterns.

To solve the donor-recipient matching problem, the cumulative frequency of cases according to the dependent variable is represented in Figure 1. As can be checked, the dataset shows a significant degree of imbalance. The Imbalance Ratio (IR) is a measure of the imbalance degree of a class with respect to the rest of classes and is defined as:

$$IR(q) = \frac{\sum_{j \neq q} N_j}{Q \cdot N_q} \quad (1)$$

where  $Q$  is the number of classes and  $N_q$  is the number of patterns in class  $q$ . The IRs for the different classes are  $IR(C_1) = 4.43$ ,  $IR(C_2) = 5.12$ ,  $IR(C_3) = 5.12$  and  $IR(C_4) = 0.04$ .

## 2.2. Artificial neural networks and evolutionary algorithms

ANNs were created in the 50s, and are still under study due to their high computational power. In [20], it is demonstrated that they are universal approximators, i.e., they are capable of approximating any continuous function with a single hidden layer, provided they have enough neurons. This and other characteristics make them a very attractive option for researchers in the field of artificial intelligence. Furthermore, they have been widely used in applications for medicine such as [21, 22, 23, 24].

ANNs need a training phase in order to adjust their weights, and thus, correctly model the output function. Generally, to train them, gradient descent algorithms (such as Backpropagation [25] or irProp [26]) are used, but there are other alternatives, such as Evolutionary Algorithms (EAs). EAs are computationally more expensive than a gradient descent approach, but they are also much more flexible when training an ANN [27].

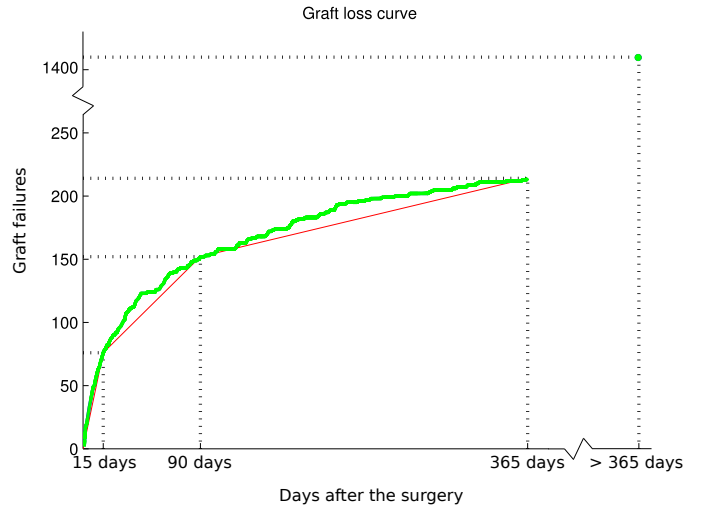


Figure 1: Graph showing the cumulative frequency of graft loss. Note that a regression analysis is not feasible because of the high number of organs which survived the 365 day threshold (for which, we do not have a label).

With EAs [28], we can perform any operation on the neural network, e.g., deleting, modifying or adding a link, deleting or adding new neurons, etc. Moreover, we can also establish a fitness function depending on the aspect of the problem considered. These features make EAs able to optimize both the structure and weights of the networks [29, 30] at the same time, focusing on those aspects of the problem that we pay interest to. Many researchers have shown that EAs perform well for global searching, because they are capable of finding promising regions in the whole search space, while gradient descent methods perform a local search.

Finally, a widely used technique is the hybridization of EAs with gradient descent ones [31] in order to find a first optimal performing a global search, and then, optimize it using a local search procedure.

## 3. Methods

This section describes the ordinal ANN model, the classification algorithm proposed and the over-sampling techniques used to deal with the imbalanced nature of the data.

The goal in ordinal classification is to assign an input vector  $\mathbf{x}$  to one of  $J$  discrete classes  $C_j$ ,  $j \in \{1, \dots, J\}$ , where there exists a given ordering among the labels  $C_1 < C_2 < \dots < C_J$ , < denoting this order information. Hence, the objective is to find a prediction rule,  $C : \mathcal{X} \rightarrow \mathcal{Y}$ , by using an i.i.d. training sample  $X = \{\mathbf{x}_i, y_i\}_{i=1}^N$ , where  $N$  is the number of training patterns,  $\mathbf{x}_i \in \mathcal{X}$ ,  $y_i \in \mathcal{Y}$ ,  $\mathcal{X} \subset \mathbb{R}^k$  is the  $k$ -dimensional input space, and  $\mathcal{Y} = \{C_1, C_2, \dots, C_J\}$  is the label space.

### 3.1. Proposed ordinal ANN classifier

An adaptation of the Proportional Odds Model (POM) [14] to artificial neural networks is used in this work. POM is the extension of binary logistic regression for dealing with ordered categories. Our adaptation is based on two elements: the first one is a second hidden linear layer with only one node whose

Table 1: Main characteristics of the dataset: features considered, number of patterns and classes, etc.

Number of patterns: 1406    Number of classes: 4    Number of features: 38		
Attribute name	Type	Value
Recipient		
Age	Numeric	[18,76]
Gender	Binary	0 = male; 1 = female
Body mass index	Numeric	[14,68.3]
Diabetes mellitus	Binary	0 = absence; 1 = presence
Arterial hypertension	Binary	0 = absence; 1 = presence
Dialysis at transplant	Binary	0 = absence; 1 = presence
Etiology	Nominal	0 = virus C cirrhosis; 1 = alcohol; 2 = virus B cirrhosis; 3 = fulminant hepatic failure; 4 = primary biliary cirrhosis; 5 = primary sclerosing cholangitis; 6 = others
Portal thrombosis	Ordinal	0 = no; 1 = partial; 2 = complete
Waiting list time	Numeric	[0,1978]
MELD (inclusion)	Numeric	[1,46]
MELD (at transplant)	Numeric	[5,50]
TIPS at transplant	Binary	0 = absence; 1 = presence
Hepatorrenal syndrome	Binary	0 = absence; 1 = presence
Upper abdominal surgery	Binary	0 = absence; 1 = presence
Pretransplant status performance	Nominal	0 = at home; 1 = hospitalized; 2 = hospitalized in ICU; 3 = hospitalized in ICU with mechanical ventilation
Cytomegalovirus	Binary	0 = absence; 1 = presence
Donor		
Age	Numeric	[10,86]
Gender	Binary	0 = male; 1 = female
Body mass index	Numeric	[14.38,53.35]
Diabetes mellitus	Binary	0 = absence; 1 = presence
Arterial hypertension	Binary	0 = absence; 1 = presence
Cause of exitus	Nominal	0 = brain trauma; 1 = cerebral vascular accident; 2 = anoxia; 3 = deceased vascular after cardiac death; 4 = others
Hospitalization length in ICU	Numeric	[0,58]
Hypotension episodes	Binary	0 = absence; 1 = presence
High inotropic drug use	Binary	0 = absence; 1 = presence
Creatinine plasma level	Numeric	[0.1,9.5]
Sodium plasma level	Numeric	[98,187]
Aspartate transaminase level	Numeric	[1,1090]
Alanine aminotransferase plasma level	Numeric	[2,1400]
Total bilirubin	Numeric	[0.06,4.2]
Hepatitis B	Binary	0 = absence; 1 = presence
Hepatitis C	Binary	0 = absence; 1 = presence
Cytomegalovirus	Binary	0 = absence; 1 = presence
Operative factors		
Multi-organ harvesting	Binary	0 = no; 1 = yes
Combined transplant	Binary	0 = no; 1 = yes
Complete or partial graft	Binary	0 = no; 1 = yes
Cold ischemia time	Ordinal	0 = <6h.; 1 = 6-12h.; 2 = >12h.
ABO compatible transplant	Binary	0 = no; 1 = yes

The end-point variable is the time leading up to liver failure: 1) before 15 days, 2) between 15 days and 3 months, 3) between 3 months and a year and 4) no graft failure presented after the first year.

All nominal and ordinal variables are transformed into binary ones, resulting in a set of 55 variables.

inputs are the non-linear transformations of the first hidden layer. The task of this node is to project the values into a line, in order to impose an order. Apart from this one-node linear layer, an output layer is included with one bias for each class, whose objective is to set the optimum thresholds to classify the patterns in the class they belong to.

The structure of our model is shown in Figure 2 which has two main parts. The part at the bottom is formed by sigmoidal unit (SU) neurons, where  $\mathbf{x} = (x_1, \dots, x_k)$ , is the vector of input variables, and  $k$  is the number of variables in the database.  $\mathbf{W} = \{\mathbf{w}_1 \dots \mathbf{w}_M\}$  is the matrix of weights of the connections from the input nodes to the hidden SU nodes (for each neuron,  $\mathbf{w}_j = \{w_{j0}, w_{j1}, \dots, w_{jk}\}$ ,  $w_{j0}$  being the bias of the neuron), and  $\sigma$  is the sigmoidal function.

The upper part of the figure shows a single node in the second hidden layer, which is the one that performs the linear transformation of the POM model. Its result,  $f(\mathbf{x}, \theta)$ , where  $\theta = \{\mathbf{W}, \beta_1, \dots, \beta_M\}$ , is connected, together with a second bias, to the output layer, where  $J$  is the number of classes, and  $\beta_0^0, \dots, \beta_0^{J-1}$  are the thresholds for the different classes. These

$J - 1$  thresholds are able to separate the  $J$  classes, but they have to fulfill the order constraint shown in the figure. Finally, the output layer obtains the outputs of the model,  $f_j(\mathbf{x}, \theta, \beta_0^j)$ , for  $j \in \{1, \dots, J - 1\}$ . These outputs are transformed into a probability ( $g_j(\mathbf{x}, \theta, \beta_0^j)$ ), using the POM model structure.  $g_j(\mathbf{x}, \theta, \beta_0^j)$  is the probability that each pattern has to belong to the different classes, and the class with the greatest probability is the one selected by the neural network. More details about this neural network model and the corresponding model equations can be checked in [32].

### 3.2. Training algorithm and proposed fitness function

In this section, the hybrid algorithm used to optimize the structure and weights of the ANN is described. The aim of the proposed algorithm is, firstly, to design an ANN with optimal structure, and secondly, to calculate the optimal weights for the problem being solved. It is a hybrid EA, in which, each  $G$  generations, the best ANN model is extracted and optimized by a gradient descent algorithm using the output error of the ANN. Later, that optimized ANN model is included in the next

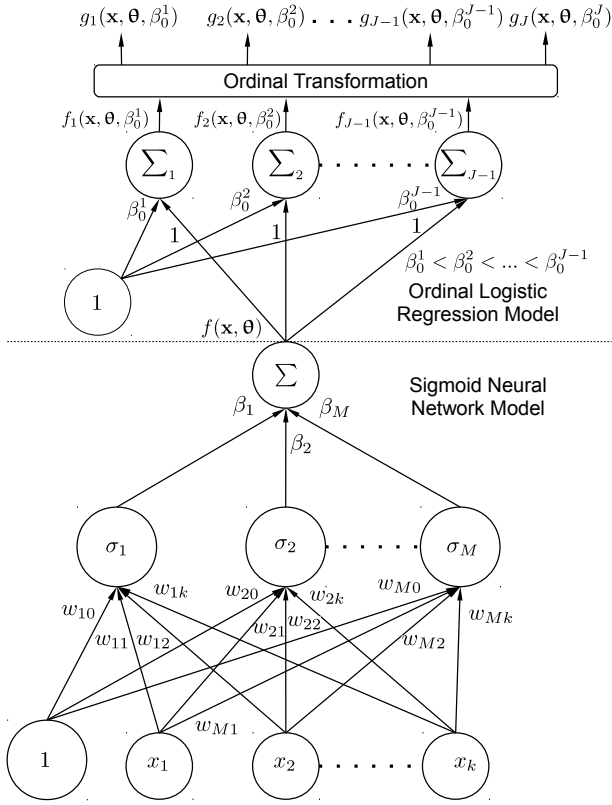


Figure 2: Neural Network model used for ordinal classification

generation of the EA. In this way, the algorithms hybridization follows a Darwinian principle [33]. The algorithm is based on the proposed in [34], where the main difference resides in the dynamic treatment of the class weights.

As this problem is of unbalanced ordinal nature, both algorithms (the EA and the gradient descent mechanism), must be specifically adapted, so the following modification of the Mean Squared Error (MSE) function is proposed, both for the fitness evaluation of the EA and for the gradient descent error:

$$l(\theta, \beta_0) = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^J IM(C_j) \cdot c(y_i, j) \cdot (g_j(\mathbf{x}, \theta, \beta_0^j) - y_i^{(j)})^2, \quad (2)$$

where several symbols must be clarified.  $\mathbf{y}_i = (y_i^{(0)}, y_i^{(1)}, \dots, y_i^{(J)})$  is a 1-of- $J$  encoding vector of the label from pattern  $\mathbf{x}_i$  (i.e.  $y_i^{(j)} = 1$  if the pattern  $\mathbf{x}_i$  belongs to class  $j$ , and 0 otherwise),  $y_i$  is the corresponding index (i.e.  $y_i = \arg_j(y_i^{(j)} = 1)$ ),  $\beta_0 = (\beta_0^1, \dots, \beta_0^{J-1})$  and  $\theta$  are the vector of biases and the vector of parameters of the ranking function, respectively, and  $c(y_i, j)$  is an ordinal cost function in the following form:

$$c(y_i, j) = \begin{cases} (J/2)(J-1), & \text{if } y_i = j, \\ |y_i - j|, & \text{if } y_i \neq j. \end{cases}$$

This cost function gives higher importance to the errors of the classifier as the number of classes between predicted and true

ranks increases. For example, if a pattern is predicted to be a member of class  $C_1$  and its true rank is class  $C_4$ , the absolute difference would be 3, which is a high cost. On the other hand, if a pattern is predicted to be in class  $C_3$  and its true rank is class  $C_4$ , the difference would be only 1, a considerably smaller error than in the first case. This modification of the error function helps us reduce the magnitude of the errors of the classifier in the ordinal scale, e.g. avoiding that patterns of class " $< 15days$ " are classified as " $> 365days$ ". The value  $(J/2)(J-1)$  corresponds to the maximum sum of the costs for a class (i.e. the sum of the costs for the extreme classes), in order to force the error of correctly classifying a pattern to be 0. Moreover, we must also consider the imbalance dataset problem, so we consider a dynamic cost ( $IM(C_j)$ ). These dynamic weights substitute the static weights proposed in [34], where the weights are fixed beforehand based on the number of patterns per class.

In this paper, we propose to initialize the costs for each class inversely proportional to the number of patterns per class:

$$IM(C_j) = \frac{1 - \frac{N_j}{N}}{\sum_{j=1}^J 1 - \frac{N_j}{N}} = \frac{N - N_j}{N(J-1)}, \quad j = 1, \dots, J,$$

where  $N_j$  corresponds to the number of patterns of class  $j$ . This corresponds to the normalized imbalance ratio of the class. This will give higher cost to the classes with less patterns, penalizing their misclassification higher than the misclassification of patterns from the majority class.

There has been some attempts to modify  $MSE$  to be applied in imbalanced problems like [35], where they set a fixed cost based on the overlap ratio and class distribution ratio obtaining competitive results. In our case, these costs are updated every  $G$  generations, based on the minimum sensitivity of the best model for the worst and the best classified classes (with indexes  $p$  and  $q$ , respectively), in this way, the costs adapt to the current classification performance, instead of being fixed by some initial parameters:

$$p = \arg_j \min(S_j), q = \arg_j \max(S_j), \quad j = 1, \dots, J, \\ IM(C_p)' = IM(C_p)(1 - \alpha) \\ IM(C_q)' = IM(C_q)(1 + \alpha)$$

where  $p$  and  $q$  the indexes of the worst and best classified classes, respectively, and  $\alpha$  is a value in the range  $[0, 1]$  which is set by the user to configure the updating rate and  $S_j$  is the sensitivity of the  $j$ -th class for the best model of the population (i.e. the ratio of patterns correctly predicted as belonging to the  $j$ -th class). These sensitivities are obtained as:

$$S_j = \frac{100}{N_j} \sum_{i=1}^{N_j} (I(y_i^* = y_i)), \quad j = 1, \dots, J, \quad (3)$$

where  $y_i^*$  is the label predicted by the model,  $y_i$  is the correct label,  $N_j$  is the number of patterns in the  $j$ -th class and  $I(\cdot)$  is the indicator function, returning 0 if the condition is false and 1 otherwise.

The idea is to make the EA focus on the class which has more problems to be correctly classified. This can be considered as an evolutionary (or dynamic) cost that will force the

misclassified class, however small it may be, to have a strong impact on the fitness function. A flowchart of the algorithm is shown in Figure 3, where *LS ind.* denotes local search individual, *EA ind.* evolutionary algorithm individual, the maximum number of generations is equal to 150 and the value of  $G$  is 50 generations.

First, the ANNs population of the EA is initialized randomly and the mutation operators, both structural (optimize weights) and parametric (optimize structure), are applied. Every generation, the population is updated, maintaining the 10% of the best individuals for the next generation, thus being an elitist algorithm. Every  $G = 50$  generations, the best individual of the population is taken and it is optimized by the iRProp+ [36] algorithm, a local search method. In this way, the derivatives of the weighted error (see Equation 2) are used to further improve the model (please, check more details about this algorithm and the corresponding derivatives in [27, 32]).

Once the best individual is optimized by the gradient descent algorithm, that model is included again in the population. The best evolutionary and gradient descent solutions (Best Evo and Best iRProp+, respectively) are recorded. Then, the EA continues searching for optimal ANN solutions for the problem until the next gradient descent optimization. Finally, in the last generation, the best models from all the reports obtained during the evolutionary process is obtained and returned as the final solution.

The implementation of the ANN model and the training algorithm was based on NNEP, a Java framework for neural network evolution, which can be found in KEEL environment<sup>1</sup>.

### 3.3. Ordinal over-sampling

As stated before, the ordinal nature of the classes should be taken into account even for preprocessing the data. In this sense, three over-sampling techniques in the context of ordinal classification are used in this paper as a preprocessing step for the dataset [19]. The main intuition behind this method is that the ordering structure of the classes can be exploited when generating new synthetic patterns. A synthetic pattern in this case corresponds to a virtual donor-recipient pair, which we create to balance the class distribution and make the classifier pay more attention to minority classes. These virtual pairs are created using the information of other pairs, so the new synthetic pattern is not totally virtual, but rather based on the combination of two donors and two recipients. The order is exploited considering a neighborhood graph between the classes, which aims to capture the underlying manifold of the ordinal labelling space. New patterns are generated on the paths that preserve the ordinal structure of this manifold and create a spatial continuity on the input space.

Roughly speaking, the over-sampling techniques considered make use of different graph operations to create a representative data-driven manifold that exploits the order of the classes. The strategies tested in this paper are:

- Ordinal Graph-based Over-sampling via Neighbourhood Information using a probability function for the intra-class edges (OGO-NI). In this case, the manifold structure is constructed based on a simple neighborhood analysis, and new synthetic patterns are created using a probabilistic approach.
- Ordinal Graph-based Over-sampling via Shortest Paths using a probability function for the intra-class edges (OGO-SP). The manifold is constructed using the previous mentioned neighborhood analysis, but it is refined computing the shortest paths in the graph.
- Ordinal Graph-based Over-sampling via Interior Shortest Paths (OGO-ISP). The technique for constructing the graph is the same than in OGO-SP but, in this case, the probabilistic approach for constructing new patterns is also used.

Further details about these methods can be consulted in [19]. The software used for performing oversampling was developed in MATLAB and can be downloaded from our research group website<sup>2</sup>.

## 4. Results

This section presents the experimental part of this paper, where nine different classification methodologies have been compared for the purpose of predicting survival time post-transplant.

### 4.1. Evaluated methodologies

To tackle this donor-recipient allocation problem, the methodology described in Section 3.2 has been referred to as Dynamic IMbalanced Ordinal neural NETWORK (DIM-ORNET). The experiments in this paper have been designed to compare several methodologies with this proposal:

- Ordinal neural NETWORK (ORNET). The original algorithm presented in [18], where the fitness function is the one presented in Section 3.2 without considering the cost  $IM(C_j)$ .
- IMbalanced Ordinal neural NETWORK (IM-ORNET). In this case the fitness function is the one presented in Section 3.2, without considering the dynamic update of the costs (i.e. the method discussed in [34]).
- Support Vector Classifier using the one-versus-one approach (SVC) [37], a state-of-the-art nominal classification methodology.
- Tree methodologies, which are gaining popularity in medical applications, because of their interpretability and good performance. Specifically, we considered Random Forests (RF) [38] and Gradient Boosted Trees (GBT) [39].

<sup>1</sup><http://keel.es/>

<sup>2</sup><http://www.uco.es/grupos/ayrna/GB0forOR>

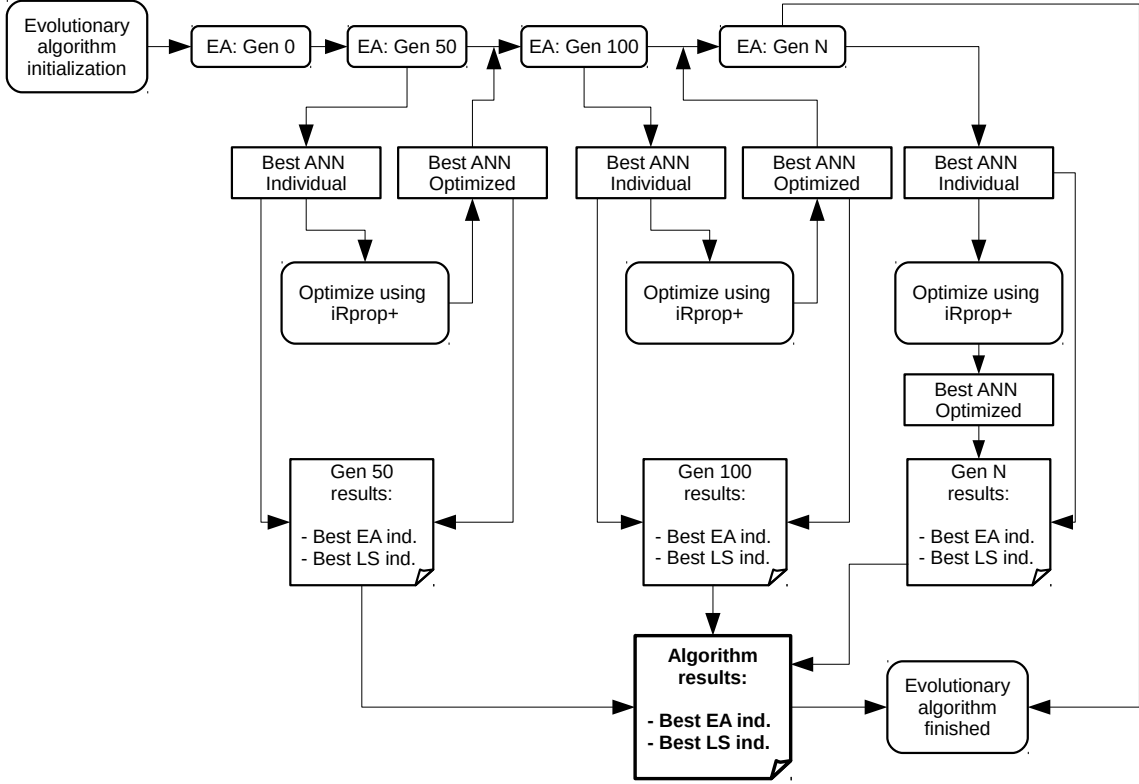


Figure 3: Flowchart of the proposed EA.

- Proportional Odds Model (POM) [14], i.e. a standard ordinal logistic regression method.
- Extreme Learning Machine for Ordinal Regression (ELMOR) [40], which corresponds to the adaptation to ordinal classification of a very popular non-iterative neural network training method.
- Error Correcting output codes with the utility cascade model and Support Vector Machines (ECSVM), which is the method previously proposed in [13] for tackling the same classification problem.
- Kernel Discriminant Learning for Ordinal Regression (KDLOR) [15], which applies discriminant learning in the context of ordinal classification.
- REDuction scheme for ordinal classification using Support Vector Machines (REDSVM) [41], which approaches ordinal classification based on an augmented binary classification problem, using SVMs.

As can be seen, seven of the eight methods considered for comparison are ordinal, although only ECSVM considers the imbalanced nature of the dataset, given the low amount of works that deal simultaneously with both problems.

#### 4.2. Evaluation metrics

Several measures can be considered for evaluating ordinal classifiers. The most common ones in machine learning are the Mean Absolute Error (*MAE*) and the Mean Zero-one Error (*MZE*) [42], being  $MZE = 1 - Acc$ , where *Acc* is the accuracy or correct classification rate. However, these measures may not be the best option, for example, when measuring performance in the presence of class imbalance [43], and/or when the costs of different errors vary markedly. The accuracy (*Acc*) is defined by:

$$Acc = \frac{100}{N} \sum_{i=1}^N (I(y_i^* = y_i)),$$

where  $I(\cdot)$  is the zero-one loss function,  $y_i$  is the desired output for pattern  $\mathbf{x}_i$ ,  $y_i^*$  is the prediction of the model and  $N$  is the total number of patterns in the dataset. *Acc* values vary from 0 to 100, and it represents global performance in the classification task. However, it does not take the category order into account, and it is not recommended for imbalanced datasets.

The geometric mean of the sensitivities of each class (*GMS*) is an average of the percentages of correct classification individually obtained for each class:

$$GMS = \sqrt[J]{\prod_{j=1}^J S_j},$$

where the sensitivities are obtained using Equation 3.

The average mean absolute error (*AMAE*) [43] is the mean of *MAE* classification errors throughout the classes, where *MAE* is the average absolute deviation of the predicted class from the true class (in number of categories on the ordinal scale). It is able to mitigate the effect of imbalanced class distributions. Let  $MAE_j$  be the *MAE* for a given  $j$ -th class:

$$MAE_j = \frac{1}{N_j} \sum_{i=1}^{N_j} |O(y_i) - O(y_i^*)|, \quad 1 \leq j \leq J,$$

where  $O(C_j) = j$ ,  $1 \leq j \leq J$ , i.e.  $O(y_j)$  is the order of class label  $y_j$ . Then, the *AMAE* measure can be defined in the following way:

$$AMAE = \frac{1}{J} \sum_{j=1}^J MAE_j.$$

*MAE* values range from 0 to  $J - 1$ , as do those of *AMAE*.

These three performance metrics briefly summarize the most important aspects of confusion matrices: *Acc* giving an idea of the global performance, *GMS* focusing on the performance of minority classes and *AMAE* reflecting the mean magnitude of errors in the ordinal scale.

#### 4.3. Experimental setting

For evaluating the results, a stratified 10-fold technique has been used to divide the data, and the results have been taken as the mean and standard deviation of the measures for the test sets over the 10 obtained models.

The parameters for all methods (except for the ORNET methodologies) have been chosen using a nested 5-fold cross-validation over the training set (independent of the 10-fold technique). The final parameter combination was the one which obtained, in mean, the best average performance for the 5 validation sets of this nested 5-fold cross-validation, where the metric used was the *AMAE*. The test sets were never used during model selection.

The parameters used during model selection are now specified. The kernel selected for all the kernel methods (SVC, ECSVM, KDLOR and REDSVM) was the Gaussian one,  $K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{\sigma^2}\right)$ , where  $\sigma$  is the kernel width. For every tested kernel method, the kernel width was tuned within the range  $\sigma \in \{10^{-3}, 10^{-2}, \dots, 10^3\}$ , as well as the cost parameter associated to SVM-based methods,  $C \in \{10^{-3}, 10^{-2}, \dots, 10^3\}$ . For the ELMOR methodology, the sigmoidal units were used, and the number of hidden neurons was selected within the values  $H \in \{5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$ .

The *AMAE* metric was used for model selection for all these methods because cross-validating by *Acc* led to very poor results in this application (poor results in terms of the minority classes). Cross-validating by *GMS* does not take the ordinal magnitude of the errors into account, although it led to reasonable results for minority classes. *AMAE*, however, mixes both ideas, and although the *GMS* could be zero, minority classes are ordered in a better manner, which is the main purpose of

the application. Nonetheless, the use of *AMAE* or *GMS* usually leads to worse results for *Acc* (which is obvious, given the imbalanced nature of the data). Finally, it should also be mentioned that because of the definition of the set of thresholds in KDLOR (these are not optimized, they are set to the middle point between two consecutive and projected classes), this method usually obtains better results for minority classes and worse results for global *Acc*.

For the ORNET methodologies, the hidden layer was composed by sigmoidal basis functions, where the minimum number of hidden neurons was 20 and the maximum was set to 30. The maximum number of hidden neurons was always set to 10 units more to give the EA enough freedom to search for the optimum classifier. The population size and the number of generations was set to 500. For DIM-ORNET  $G$  is 500 and  $\alpha$  to 0.05.

Several works in the literature have considered the over-sampling of classes that present an imbalanced ratio (*IR*) higher than 1.5 [44, 45]. This threshold ( $IR = 1.5$ ) is used, given that it is suggested in the literature that it is the point from which the imbalance of the data starts to hamper the learning phase. We use this threshold to decide if a class is imbalanced with respect to the rest of classes. Once that we have decided this, we compute the number of necessary patterns to include, so that the *IR* of that class is lower than the initial threshold. This is, we add new patterns until we consider that the class is balanced. In this case, we over-sample the first three classes of the dataset, which present a pattern distribution of {69, 75, 64}, as opposed to the majority class which is composed of 1198 patterns.

#### 4.4. Results

As a first experiment and in order to analyze the necessity of considering both donor and recipient characteristics for this application, different versions of the original dataset have been created combining these blocks of features. In this way, we test the importance of each block and verify if all of them are relevant for the model, or if we can solve the problem avoiding some of the characteristics, making the model simpler.

The proposed model has been trained and generalized with each of the new datasets, obtaining the test results shown in Table 2. Note that all the results included in this Table are obtained using the original version of the test sets, the over-sampling method being only applied to the training sets. The characteristics considered are those shown in Table 1, where All corresponds to all features, Donor and Recipient characteristics are clearly described in Table 2, and Surgery characteristics are those related to Operative factors. It can be seen that the results obtained with a single block of variables (both when using only donor or recipient characteristics) are not suitable for the problem, because they both classify nearly all the patterns as belonging to the majority class ( $GMS = 0\%$  and  $AMAE = 1.50$ , i.e. a trivial classifier, which however leads to a very high *Acc*). This effect is usually checked by considering the Area Under the ROC curve (AUC), but this evaluation metric is restricted to binary classification problems. There are some extensions of ROC to multi-class problems, such as [46, 47], which seem



Table 2: Average and standard deviation ( $Mean \pm SD$ ) of  $Acc$ ,  $GMS$  and  $AMAE$  results for the test sets

Method	Input Features	$Acc(\%)$	$GMS(\%)$	$AMAE$
DIM-ORNET	Original	$73.40 \pm 2.73$	$5.70 \pm 12.04$	$1.395 \pm 0.089$
DIM-ORNET	Donor	$82.84 \pm 1.67$	$0.00 \pm 0.00$	$1.497 \pm 0.058$
DIM-ORNET	Recipient	$81.73 \pm 1.73$	$0.00 \pm 0.00$	$1.476 \pm 0.050$
DIM-ORNET	Donor + Surgery	$79.8611 \pm 1.50$	$0.00 \pm 0.00$	$1.450 \pm 0.070$
DIM-ORNET	Recipient + Surgery	$78.47 \pm 1.82$	$0.00 \pm 0.00$	$1.413 \pm 0.111$

attractive because of the benefits of ROC analysis, but their exponential computational complexity as function of the number of classes is very restrictive [48, 49]. That being the reason why we consider  $GMS$  in this paper (complemented by  $AMAE$  to check the magnitude of the errors). It is clear how adding information about the surgery is necessary to improve the results, surpassing the  $AMAE = 1.50$  limit, which means that minority classes are starting to be differentiated from the majority one. In conclusion, observing the results corresponding to consider all features, which lead to a reasonable  $AMAE$  and a  $GMS$  different from 0%, all the variables analyzed in Section 2 are necessary to provide a suitable model to solve the organ allocation problem.

Once we have validated these three sources of information, we include an experiment that tests different algorithms for this problem. In this sense, Table 3 shows the results obtained for the different algorithms considered with the four versions of the dataset (the original one and three versions obtained by applying the three over-sampling methods to the different training sets, as presented in Section 3.3). Observing the results in Table 3, different methods can be emphasized depending on the evaluation metric considered. In this dataset, it is important to obtain a trade-off between all the metrics included, aiming at an accurate model that also respects the classification and order of minority classes. Note that a  $GMS = 0\%$  means that at least one class is always misclassified, and this is the case of SVC, POM, REDSVM and tree methodologies (RF and GBT), so they are not included in the table. If we only focus on  $Acc$ , the best model is obtained by ELMOR using the original dataset, but this result, examining  $GMS$  and  $AMAE$  scores, shows that nearly all the patterns are classified in the predominant class (the majority one). Note that most classifiers are in this case trivial ones (especially the ones not included in the table), showing the difficulty of the considered application.

It can be seen that the application of the different over-sampling techniques improves the  $GMS$  results and obviate trivial models in some cases. This is especially true when combined with the evolutionary ordinal neural network methodologies (ORNET, IM-ORNET and DIM-ORNET). It can also be seen that the results obtained by DIM-ORNET improve those obtained for the same problem in previous studies (ECSVM [13] and IM-ORNET [34]). Furthermore, the use of dynamic weights in the evolution improve the results in this dataset for both  $GMS$  and  $AMAE$ .

The best  $AMAE$  model is obtained by KDLOR using the

OGO-NI technique. The corresponding  $GMS$  score is 2.79% in average (which is, at least, higher than 0%), and whose  $AMAE$  shows that this classifier tends to respect the order of the label space (i.e., generally, it does not misclassify a pattern belonging to class 0 in class 3, for example) at the cost, however, of sacrificing  $Acc$ , whose result around 50% is not acceptable. In this way, the classifier is biased against predicting almost all patterns as belonging to intermediate classes (which indeed minimizes  $AMAE$ ). Note that the combination of ECSVM and the over-sampling approaches does not result in a satisfactory result. This is due to the formulation of this classifier, in which the imbalanced nature of the data is already taken into account, which is the reason why, in the original data, ECSVM performs better than in the over-sampling case (where the data is balanced by the introduction of synthetic examples). When combining DIM-ORNET with the OGO-ISP over-sampling technique, the model with the best trade-off is obtained (i.e. a relatively high  $Acc$ , the best result in  $GMS$  and an acceptable  $AMAE$ ).

In conclusion, it can be affirmed that the dynamic weight proposal improves the results of the base algorithm in most cases, obtaining a  $GMS$  higher than 5% considering the three over-sampling techniques and competitive  $Acc$  and  $AMAE$ .

#### 4.5. Interpretation of the best model obtained

For model interpretation, we select the best algorithm (DIM-ORNET with OGO-ISP) and we extract information from the best ANN model among the 10 folds. The results obtained in this case are the following:  $Acc = 70.14\%$ ,  $GMS = 33.34\%$  and  $AMAE = 1.068$ , which are very competitive.

To interpret the ANN model used, we consider the number of times that each variable has been used in the model and the mean of the absolute values of the associated ANN-weights. In this case, the variables that have a mean weight higher than the average and that are selected more times than the average are: body mass index (recipient), etiology (recipient, primary biliary cirrhosis), portal thrombosis (recipient, no), pretransplant status performance (recipient, hospitalized), diabetes mellitus (donor), hospitalization length in ICU (donor), hypotension episodes (donor), high inotropic drug use (donor) and hepatitis C (donor). It can be appreciated that both features concerning the recipient and the donor are used, which emphasizes the need of using both sources of information. Concerning the selected features, it can be seen that there are different conditions (primary biliary cirrhosis, portal thrombosis, diabetes

Table 3: Average and standard deviation ( $Mean \pm SD$ ) of  $Acc$ ,  $GMS$  and  $AMAE$  results for the test sets

Method	Preprocessing	$Acc(\%)$	$GMS(\%)$	$AMAE$
ORNET	Original	$71.84 \pm 3.15$	$0.00 \pm 0.00$	$1.393 \pm 0.078$
ORNET	OGO-NI	$66.24 \pm 3.05$	$5.03 \pm 10.74$	$1.396 \pm 0.116$
ORNET	OGO-SP	$66.74 \pm 3.21$	$4.4 \pm 9.27$	$1.431 \pm 0.089$
ORNET	OGO-ISP	$68.23 \pm 4.96$	$2.62 \pm 8.28$	$1.393 \pm 0.078$
IM-ORNET	Original	<i><math>84.85 \pm 0.91</math></i>	$0.00 \pm 0.00$	$1.501 \pm 0.002$
IM-ORNET	OGO-NI	$69.08 \pm 3.54$	$5.29 \pm 11.17$	$1.367 \pm 0.072$
IM-ORNET	OGO-SP	$71.13 \pm 2.99$	<i><math>7.47 \pm 12.08</math></i>	$1.354 \pm 0.113$
IM-ORNET	OGO-ISP	$66.95 \pm 6.09$	$2.66 \pm 8.41$	$1.421 \pm 0.092$
DIM-ORNET	Original	$73.40 \pm 2.73$	$5.70 \pm 12.04$	$1.395 \pm 0.089$
DIM-ORNET	OGO-NI	$68.68 \pm 3.29$	$7.01 \pm 11.37$	$1.356 \pm 0.116$
DIM-ORNET	OGO-SP	$68.61 \pm 3.50$	$5.29 \pm 11.14$	$1.383 \pm 0.121$
DIM-ORNET	OGO-ISP	$70.00 \pm 2.31$	<b><math>7.49 \pm 12.01</math></b>	$1.326 \pm 0.096$
ELMOR	Original	<b><math>85.21 \pm 0.27</math></b>	$0.00 \pm 0.00$	$1.500 \pm 0.000$
ELMOR	OGO-NI	$62.09 \pm 3.94$	$0.00 \pm 0.00$	$1.336 \pm 0.063$
ELMOR	OGO-SP	$64.58 \pm 2.87$	$2.13 \pm 6.72$	$1.294 \pm 0.101$
ELMOR	OGO-ISP	$61.39 \pm 8.81$	$0.00 \pm 0.00$	$1.322 \pm 0.087$
ECSVM	Original	$66.28 \pm 4.08$	$2.28 \pm 7.20$	$1.333 \pm 0.109$
ECSVM	OGO-NI	$83.00 \pm 1.67$	$0.00 \pm 0.00$	$1.493 \pm 0.057$
ECSVM	OGO-SP	$83.00 \pm 2.16$	$0.00 \pm 0.00$	$1.503 \pm 0.044$
ECSVM	OGO-ISP	$83.00 \pm 1.25$	$0.00 \pm 0.00$	$1.503 \pm 0.041$
KDLOR	Original	$50.50 \pm 3.96$	$5.22 \pm 11.05$	$1.355 \pm 0.159$
KDLOR	OGO-NI	$54.56 \pm 4.31$	$2.79 \pm 8.81$	<b><math>1.206 \pm 0.064</math></b>
KDLOR	OGO-SP	$54.70 \pm 4.85$	$0.00 \pm 0.00$	$1.240 \pm 0.078$
KDLOR	OGO-ISP	$54.27 \pm 5.33$	$3.05 \pm 9.63$	<i><math>1.224 \pm 0.085</math></i>

The best result is in **bold** face and the second best result is in *italics*.

mellitus, hypotension episodes and hepatitis C) which are important for the prediction of organ survival post-transplant.

Two of the features selected also belong to the factors of expanded criteria donors (donors with extreme values of age, days in the intensive care unit (ICU), inotrope usage, body mass index (BMI) and cold ischemia time), characteristics that result in an increased risk of recipient and/or graft losses compared to the risk associated with the use of livers from non-extended criteria donors [50, 51].

Moreover, the variables that are selected fewer times than the mean and that also present a mean weight lower than the mean are: gender (recipient), dialysis at transplant (recipient), waiting list time (recipient), MELD at transplant (recipient), hepatorenal syndrome (recipient), upper abdominal surgery (recipient), cytomegalovirus (recipient), gender (donor), alanine aminotransferase plasma level (donor), cytomegalovirus (donor), multi-organ harvesting (donor) and combined transplant (donor).

Finally, to conclude with the interpretability of the model, it must be noted that, as this is a threshold based model, the output of the ANN model for every transplant can be observed, giving us information about the robustness of the result. If the output of a pattern is far enough from the class border (threshold), the medical expert can assume the result is more robust, while in the opposite case, where the output of a pattern is close to a

class border, that pattern could jump to the next (or previous, depending on which border the pattern is) class easily. That also helps the medical expert to take the proper decisions.

## 5. Discussion and conclusions

This paper tackles the problem of developing an organ allocation system in liver transplantation using machine learning. To support the medical expert with finer-grain information, the classification problem is considered from an ordinal classification point of view, where 4 classes are used, discretizing the time leading up to organ failure. More specifically, the dataset is composed of donor-recipient pairs of 7 transplantation units in Spain and one hospital in the United Kingdom, and considers characteristics of the donor, the recipient and the surgery. The classification problem is tackled by an ordinal evolutionary artificial neural network that takes into account the imbalanced nature of the data by the definition of a dynamic fitness function. This method is considered in conjunction with an ordinal over-sampling approach. The good synergy of these strategies widely used for imbalanced classification is shown in the results of our paper.

From the results in the experiments of this paper, several conclusions can be extracted. First of all, some of the methodologies have been discarded, which, due to the complexity of the problem considered, obviated one of the classes in the dataset,

most of them obtaining trivial classifiers. This usually results in an outstanding *Acc* score, given that 85% of the patterns belong to the majority class, while these classifiers are useless for prediction. Moreover, it can be observed that the use of an ordinal over-sampling technique improves the results in imbalanced domains with respect to the original dataset (not only in the proposed methodology, but also in the other methodologies tested). In this case, our proposal of combining ordinal over-sampling (to palliate the unbalance of the pattern distribution) with the idea of adding a dynamic fitness to the EA (which helps the model to give more importance to the worst classified class at the moment) results in the best results. As said, the proposal (DIM-ORNET) also improves the results with respect to previous works, obtaining a good balance in performance for all the metrics considered (concerning the overall classification, the classification of minority classes and the ordering of the data).

As stated before, the results obtained by the complete set of used methodologies indicate the complexity of the problem dealt within this study in which the proposal was able to obtain reasonably good results.

This model could be tested in the practice analyzing the result of each transplant performed (i.e. without changing the current policy but analyzing if the model predicts the survival accurately in a real world scenario). Note that this approach does not intend to substitute the medical professional, but rather provide him/her with an accurate prediction model which serves as a decision-support system. Moreover, our model can also be complemented with the MELD score, to also consider other factors, such as the severity of the recipient.

Future research lines comprise the acquisition of more data to create a supranational dataset and the simulation of the model in a more controlled environment to analyze its behavior.

## 6. Acknowledgements

This work has been partially subsidized by the TIN2014-54583-C2-1-R project of the Spanish Ministerial Commission of Science and Technology (MINECO), FEDER funds, the P11-TIC-7508 project of the “Junta de Andalucía” (Spain), the PI-0312-2014 project of the “Fundación pública andaluza progreso y salud” (Spain) and the PI15/01570 project (Spain, “Proyectos de Investigación en Salud”).

We would also like to thank Astellas Pharma Company for their partial support in the data acquisition

## References

- [1] S. Feng, N. Goodrich, J. Bragg-Gresham, D. Dykstra, J. Punch, M. DeRoy, S. Greenstein, R. Merion, Characteristics associated with liver graft failure: The concept of a donor risk index, *American Journal of Transplantation* 6 (4) (2006) 783–790.
- [2] P. Kamath, W. Kim, The Model for End-stage Liver Disease (MELD), *Hepatology* 45 (3) (2007) 797–805.
- [3] A. Rana, M. A. Hardy, K. J. Halazun, D. C. Woodland, L. E. Ratner, B. Samstein, J. V. Guarrera, R. S. Brown, J. C. Emond, Survival outcomes following liver transplantation (SOFT) score: a novel method to predict patient survival following liver transplantation, *American Journal of Transplantation* 8 (12) (2008) 2537–46.
- [4] P. Dutkowski, C. Oberkofler, K. Slankamenac, M. Puhan, E. Schadde, B. Millhaupt, A. Geier, P. Clavien, Are there better guidelines for allocation in liver transplantation? A novel score targeting justice and utility in the model for end-stage liver disease era, *Annals of Surgery* 254 (5) (2011) 745–753.
- [5] M. Cruz-Ramírez, C. Hervás-Martínez, J. Fernández-Caballero, J. Briceño, M. de la Mata, Multi-Objective Evolutionary Algorithm for Donor-Recipient Decision System in Liver Transplants, *European Journal of Operational Research* 222 (2) (2012) 317–327.
- [6] J. Briceño, M. Cruz-Ramírez, M. Prieto, M. Navasa, J. O. de Urbina, R. Orti, M.-A. Gómez-Bravo, A. Otero, E. Varo, S. Tomé, G. Clemente, R. B. nares, R. Bárcena, V. Cuervas-Mons, G. Solórzano, C. Vinaixa, A. Rubn, J. Colmenero, A. Valdivieso, R. Ciria, C. Hervás-Martínez, M. de la Mata, Use of artificial intelligence as an innovative donor-recipient matching model for liver transplantation: Results from a multicenter spanish study, *Journal of Hepatology* 61 (5) (2014) 1020 – 1028.
- [7] N. Peek, C. Combi, R. Marin, R. Bellazi, Thirty years of artificial intelligence in medicine (aime) conferences: a review of research themes, *Artificial Intelligence in Medicine* 65 (2015) 61–73.
- [8] F. Jimenez, G. Sanchez, J. M. Juarez, Multi-objective evolutionary algorithms for fuzzy classification in survival prediction, *Artificial Intelligence in Medicine* 60 (2014) 197–219.
- [9] V. Van Belle, K. Pelckmans, S. Van Huffel, J. Suykens, Support vector methods for survival analysis: a comparison between ranking and regression approaches, *Artificial Intelligence in Medicine* 53 (2011) 107–118.
- [10] M. Cruz-Ramírez, C. Hervás-Martínez, J. Fernández, J. Briceño, M. de la Mata, Predicting patient survival after liver transplantation using evolutionary multi-objective artificial neural networks, *Artificial Intelligence in Medicine* 58 (2013) 37–49.
- [11] M.-H. Tseng, H.-C. Liao, The genetic algorithm for breast tumor diagnosis - the case of dna viruses, *Applied Soft Computing* 9 (2) (2009) 703 – 710.
- [12] C.-J. Su, C.-Y. Wu, Jade implemented mobile multi-agent based, distributed information platform for pervasive health care monitoring, *Applied Soft Computing* 11 (1) (2011) 315 – 325.
- [13] M. Pérez-Ortiz, M. Cruz-Ramírez, M. Ayllón-Terán, N. Heaton, R. Ciria, C. Hervás-Martínez, An organ allocation system for liver transplantation based on ordinal regression, *Applied Soft Computing* 14 (A) (2014) 88 – 98.
- [14] P. McCullagh, Regression models for ordinal data, *Journal of the Royal Statistical Society, Series B (Methodological)* 42 (2) (1980) 109–142.
- [15] B.-Y. Sun, J. Li, D. D. Wu, X.-M. Zhang, W.-B. Li, Kernel discriminant learning for ordinal regression, *IEEE Transactions on Knowledge and Data Engineering* 22 (2010) 906–910.
- [16] P. McCullagh, J. A. Nelder, *Generalized Linear Models*, 2nd Edition, Monographs on Statistics and Applied Probability, Chapman & Hall/CRC, 1989.
- [17] W. Chu, S. S. Keerthi, Support vector ordinal regression, *Neural Computation* 19 (3) (2007) 792–815.
- [18] M. Dorado-Moreno, P. A. Gutiérrez, C. Hervás-Martínez, Ordinal classification using hybrid artificial neural networks with projection and kernel basis functions, in: 7th International Conference on Hybrid Artificial Intelligence Systems (HAIS2012), 2012, p. 319330.
- [19] M. Pérez-Ortiz, P. Gutiérrez, C. Hervás-Martínez, X. Yao, Graph-based approaches for over-sampling in the context of ordinal regression, *Knowledge and Data Engineering, IEEE Transactions on* 27 (5) (2015) 1233–1245.
- [20] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, *Neural Networks* 2 (5) (1989) 359–366.
- [21] C. M., J. C. Fernández, F. F., Memetic evolutionary multi-objective neural network classifier to predict graft survival in liver transplant patients, in: Proceedings of the 13th annual conference companion on Genetic and evolutionary computation, GECCO '11, ACM, New York, NY, USA, 2011, pp. 479–486.
- [22] J. Kaldersstam, P. Edén, P. Bendahl, C. Strand, M. Ferno, M. Ohlsson, Training artificial neural networks directly on the concordance index for censored data using genetic algorithms, *Artificial Intelligence in Medicine* 58 (2013) 125–132.
- [23] R. Houthoof, J. Ruysinck, et al, Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores, *Artificial Intelligence in Medicine* 63 (2015) 191–207.

- [24] O. Fontenla-Romero, B. Guijarro-Berdinas, A. Alonso-Betanzo, V. Moret-Bonillo, A new method for sleep apnea classification using wavelets and feedforward neural networks, *Artificial Intelligence in Medicine* 34 (2003) 65–76.
- [25] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning representations by backpropagating errors, *Nature* 323 (1986) 533–536.
- [26] M. Riedmiller, H. Braun, A direct adaptive method for faster backpropagation learning: The RPROP algorithm, in: *Proceedings of the 1993 IEEE International Conference on Neural Networks*, San Francisco, CA, 1993, pp. 586–591.
- [27] P. A. Gutiérrez, C. Hervás-Martínez, F. J. Martínez-Estudillo, Logistic regression by means of evolutionary radial basis function neural networks, *IEEE Transactions on Neural Networks* 22 (2) (2011) 246–263.
- [28] X. Yao, Evolving artificial neural networks, *Proceedings of the IEEE* 87 (9) (1999) 1423–1447.
- [29] X. Yao, Y. Liu, A new evolutionary system for evolving artificial neural networks, *IEEE Transactions on Neural Networks* 8 (3) (1997) 694–713.
- [30] V. Maniezzo, Genetic evolution of the topology and weight distribution of neural networks, *IEEE Transactions on Neural Networks* 5 (1994) 39–53.
- [31] H. Ishibuchi, T. Yoshida, T. Murata, Balance between genetic search and local search in hybrid evolutionary multi-criterion optimization algorithms, *IEEE Transactions on Evolutionary Computation* 7 (2) (2003) 204–223.
- [32] P. A. Gutiérrez, P. Tiño, C. Hervás-Martínez, Ordinal regression neural networks based on concentric hyperspheres, *Neural Networks* 59 (2014) 51–60.
- [33] C. R. Darwin, *The origin of species*, New York: P.F. Collier & Son, 1959.
- [34] M. Dorado-Moreno, M. Pérez-Ortiz, M. D. Ayllón-Terán, P. A. Gutiérrez, C. Hervás-Martínez, Ordinal evolutionary artificial neural networks for solving an imbalanced liver transplantation problem, in: *Hybrid Artificial Intelligent Systems: 11th International Conference, HAIS 2016*, Springer International Publishing, 2016, pp. 451–462.
- [35] P. Vorraboot, C. Lursinsap, S. Rasmequan, K. Chinnasarn, A modified error function for imbalanced dataset classification problem, 2012, pp. 854–859.
- [36] C. Igel, M. Husken, Empirical evaluation of the improved rprop learning algorithm, *Neurocomputing* 50.
- [37] C.-W. Hsu, C.-J. Lin, A comparison of methods for multi-class support vector machines, *IEEE Transaction on Neural Networks* 13 (2) (2002) 415–425.
- [38] L. Breiman, Random forests, *Machine Learning* 45 (1) (2001) 5–32.
- [39] J. H. Friedman, Stochastic gradient boosting, *Computational Statistics & Data Analysis* 38 (4) (2002) 367–378.
- [40] W.-Y. Deng, Q.-H. Zheng, S. Lian, L. Chen, X. Wang, Ordinal extreme learning machine, *Neurocomputing* 74 (1–3) (2010) 447–456.
- [41] H.-T. Lin, L. Li, Reduction from cost-sensitive ordinal ranking to weighted binary classification, *Neural Computation* 24 (5) (2012) 1329–1367.
- [42] P. A. Gutiérrez, M. Pérez-Ortiz, J. Sánchez-Monedero, F. Fernández-Navarro, C. Hervás-Martínez, Ordinal regression methods: survey and experimental study, *IEEE Transactions on Knowledge and Data Engineering* 1 (2016) 127–146.
- [43] S. Baccianella, A. Esuli, F. Sebastiani, Evaluation measures for ordinal regression, in: *Proceedings of the Ninth International Conference on Intelligent Systems Design and Applications, ISDA'09*, 2009, pp. 283–287.
- [44] A. Fernández, S. García, M. J. del Jesus, F. Herrera, A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets, *Fuzzy Sets Syst.* 159 (18) (2008) 2378–2398.
- [45] F. Fernández-Navarro, C. Hervás-Martínez, P. Antonio Gutiérrez, A dynamic over-sampling procedure based on sensitivity for multi-class problems, *Pattern Recognition* 44 (8) (2011) 1821–1833.
- [46] D. J. Hand, R. J. Till, A simple generalisation of the area under the roc curve for multiple class classification problems, *Machine Learning* 45 (2) (2001) 171–186.
- [47] T. Fawcett, An introduction to ROC analysis, *Pattern Recognition Letters* 27 (2006) 861–874.
- [48] J. C. Fernández-Caballero, F. J. Martínez-Estudillo, C. Hervás-Martínez, P. A. Gutiérrez, Sensitivity versus accuracy in multiclass problems using memetic pareto evolutionary neural networks, *IEEE Transactions on Neural Networks* 21 (5) (2010) 750–770.
- [49] T. Landgrebe, R. P. Duin, Efficient multiclass ROC approximation by decomposition via confusion matrix perturbation analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30 (5) (2008) 810–822.
- [50] R. W. Busuttill, K. Tanaka, The utility of marginal donors in liver transplantation, *Liver Transplant* 9 (7) (2003) 651–663.
- [51] J. Briceño, G. Solorzano, C. Pera, A proposal for scoring marginal liver grafts, *Transplant International* 13 (2000) S249–S252.