COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 133 (2016) 183–193

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

A new decision support model for preanesthetic evaluation

Olivier Sobrie [a,c], *Mohammed El Amine Lazouni* [b], *Saïd Mahmoudi* [a], *Vincent Mousseau* [c], *Marc Pirlot* [a,*]

[a] *Faculté Polytechnique, Université de Mons, rue de Houdain 9, B-7000 Mons, Belgium*
[b] *Biomedical Engineering Laboratory, Faculty of Technology, Abou Bekr Belkaid University of Tlemcen, BP 230 – 13000 Chetouane Tlemcen, Algeria*
[c] *CentraleSupélec, Grande Voie des Vignes, 92290 Châtenay-Malabry, France*

ARTICLE INFO

Article history:
Received 27 November 2015
Received in revised form
27 May 2016
Accepted 31 May 2016

Keywords:
Multiple criteria decision support
Preanesthesia evaluation
ASA score
Classification
MR-Sort

ABSTRACT

Background and objective: The principal challenges in the field of anesthesia and intensive care consist of reducing both anesthetic risks and mortality rate. The ASA score plays an important role in patients' preanesthetic evaluation. In this paper, we propose a methodology to derive simple rules which classify patients in a category of the ASA scale on the basis of their medical characteristics.
Methods: This diagnosis system is based on MR-Sort, a multiple criteria decision analysis model. The proposed method intends to support two steps in this process. The first is the assignment of an ASA score to the patient; the second concerns the decision to accept—or not—the patient for surgery.
Results: In order to learn the model parameters and assess its effectiveness, we use a database containing the parameters of 898 patients who underwent preanesthesia evaluation. The accuracy of the learned models for predicting the ASA score and the decision of accepting the patient for surgery is assessed and proves to be better than that of other machine learning methods. Furthermore, simple decision rules can be explicitly derived from the learned model. These are easily interpretable by doctors, and their consistency with medical knowledge can be checked.
Conclusions: The proposed model for assessing the ASA score produces accurate predictions on the basis of the (limited) set of patient attributes in the database available for the tests. Moreover, the learned MR-Sort model allows for easy interpretation by providing human-readable classification rules.

1. Introduction

Within the anesthesiology practice, evaluating the patient's physical health status is an important issue in preoperative assessment before surgery. A commonly used system to determine the patient's health status is the ASA physical classification system, developed by the American Society of Anesthesiologists (ASA). It consists of classifying patients in one of the six categories going from "healthy person" to "brain-dead

* *Corresponding author.* Department of Mathematics and Operational Research, Université de Mons, Rue de Houdain 9, 7000 Mons, Belgium. Tel.: +32 65 374682; fax: +32 65 374689.
E-mail address: marc.pirlot@umons.ac.be (M. Pirlot).
http://dx.doi.org/10.1016/j.cmpb.2016.05.021

person whose organs are being removed for donor purposes". The assignment of a patient to an ASA category does not rely on precise, objective rules. Therefore, variability and inconsistencies in the assignment of ASA score have been observed and studied (see Ref. [1] and references therein). In order to support doctors in their preanesthetic evaluation of the patient, a number of computer aided systems have been devised.

In this paper, we propose a new method, based on Multiple Criteria Decision Analysis (MCDA) models, aiming to predict the ASA score.

The anesthesiology scientific literature is not very rich in works related to computer supported preoperative patient evaluation. In Refs. [2, 3], for instance, several machine learning algorithms were used to predict the ASA score and other preoperative assessments. A drawback of machine learning methods is that they are not easily interpretable by doctors. The link they establish (through the learning process) between the patient's characteristics and his/her classification is encrypted in the parameters of the learned model in an opaque manner. For instance, with a neural network model, such as the multilayer perceptron, it is not possible to understand the patient classification by looking at synaptic weights and activation functions, even more if the model involves many variables.

In this paper, we advocate the use of more interpretable models, as those used in MCDA. With such models, the classification of a patient with given characteristics can be described as the application of explicit rules. This feature is important for enhancing the confidence of the practitioner in a computerized support system.

The model we propose to predict the ASA score of patients described by classical medical parameters is called MR-Sort [4] and belongs to the family of outranking methods [5,6]. We learn the parameters of the MR-Sort model on the basis of a database containing the description of 898 patients. This database records the description of incoming patients at the medical consultation for preoperative assessment in three private clinics and two public hospitals from Western Algeria. Besides the patients' attributes and measured medical parameters (see Table 1 and Appendix S1: Appendix A below) the database also records the ASA score as assessed by the anesthesiologist at the consultation as well as his/her decision of accepting/refusing the patient for surgery. For all patients, the planned surgery has a low to medium risk level (ophthalmic, ENT, dermatological surgery). Data were collected in the period from February 1, 2011 to October 31, 2012. Note that only four of the six ASA score classes are represented in the database. Patients with ASA scores 5 or 6 do not appear since the hospitals from which the data were collected are not included in an organ donation network.

Table 1 – List of attributes taken into account in the prediction of the ASA score in Ref. [3]. Their domain of variation, measurement unit and the direction of increasing preference are specified.

Attribute	Domain (Unit)	Direction
Age	[0–105] (year)	min.
Diabetic	{0, 1}	min.
Hypertension	{0, 1}	min.
Respiratory failure	{0, 1}	min.
Heart failure	{0, 1}	min.
Heart rate	[55–123] (bpm)	max. min.
Heart rate steadiness	{0, 1}	max.
Pacemaker	{0, 1}	min.
Atrioventricular block	{0, 1}	min.
Left ventricular hypertrophy	{0, 1}	min.
Oxygen saturation	[43–100] (%)	max.
Blood glucose level	[0.5–3.8] (g/l)	max. min.
Systolic blood pressure	[9–20.5] (cm Hg)	min.
Diastolic blood pressure	[5–13] (cm Hg)	min.

As just mentioned, the database also contains the record of decisions made regarding acceptance/refusal for surgery. In general, such a decision is not made only by taking into account the patients' characteristics and physical status. It primarily depends on the adequacy of each anesthetizing location considering the complexity of the case. The ASA Relative Value Guide (ASA RVG) [7] basic units can be used to assess the maximum degree of complexity of cases that can be dealt with at a given location by an anesthesia group [8] (see Refs. [9–14] for a detailed discussion). Moreover, hospitals exhibit a certain degree of diversity in their surgical procedures (which can be quantified [15]). For the particular case of the 898 patients in the database, the procedures in the five hospitals are very similar, and the considered types of surgery have similar risk degree. Therefore, the physical status of the patient is likely to be predominant in the decision to accept or not the patient for surgery. We thus also learned the parameters of a MR-Sort model to predict this decision on the basis of the patients' ASA score and two other attributes. We emphasize that such a model can only be valid in a particular anesthetizing location and for a surgery of a specified complexity degree. Changing the location and/or the type of surgery requires adapting the model's parameter values (while the methodology remains valid).

The obtained MR-Sort models provide synthetic rules, expressed in terms of patients' medical variables, accounting for the practice of a community of physicians in assessing the ASA score and deciding on acceptance for surgery (the latter in specific locations and for specific types of surgery).

This paper is organized as follows. We start with an overview of the existing anesthesiology scientific literature related to the decision support methods used in preoperative patient classification. The third section describes the MCDA method used in this paper, namely MR-Sort, as well as the algorithms that allow us to learn the model's parameters based on assignment examples. The fourth section details the experimental results obtained with the MR-Sort method in order to predict the patients' ASA score and the decision of acceptance to surgery. Section 5 discusses the results; conclusions appear in Section 6. There are two appendices. The first describes the patients' attributes or characteristics encoded in the database. The second shows 10 examples of MR-Sort models that predict the patients' ASA score with high accuracy, based on different subsets of patients' attributes.

2. Literature review

2.1. *Multiple criteria decision analysis in medicine*

MCDA methods have been used in medicine for various applications going from cancer diagnosis and treatment [16,17], to the selection of technologies in health care settings [18]. Many medical applications are based on the Analytic Hierarchy Process (AHP) method. An overview of existing applications using AHP is presented in Ref. [19]. Few articles report on applying other MCDA models like the ELECTRE methods. In Ref. [20], the ELECTRE TRI-C sorting procedure is used in the context of assisted reproduction. Couples are assigned to categories which correspond to the number of embryos that have to be transferred back to the uterus of the woman in order to obtain a single pregnancy. To the best of our knowledge, multiple criteria decision analysis was never used for ASA score determination while this score is widely used by anesthesiologists for preanesthetic evaluation.

An advantage of most MCDA methods as compared to machine learning algorithms is that they lean themselves more easily to interpretation. Therefore, such methods are prone to be used in decision contexts in which *explaining* the recommendation is important [21,22]. Such a feature could contribute to increasing the trust of the physician in a computer-aided preanesthesia assessment.

2.2. Decision support systems for anesthesia

Glance et al. [23] proposed a probabilistic model for evaluating the surgical mortality. The objective of this work is to predict the patient mortality after a non-cardiac surgery, in order to reduce the operation risks. This system calculates the risk score in an empirical way by using three descriptors, which are the ASA score; the type of surgery, either high or intermediate risk; and whether or not the surgery is urgent. A huge database involving 298,772 patients was gathered from different hospitals between 2005 and 2007, resulting as follows:

- patients with a risk score under 5 had a mortality risk under 0.5%;
- patients with a risk score between 5 and 6 had a mortality risk between 1.5% and 4%;
- patients with a risk score over 6 had a mortality risk of more than 10%.

An automatic system capable of predicting the operation anesthetic risk has been developed in Ref. [24]. This system assesses three classification techniques based on supervised learning. The assessment has been done by using the three following classification techniques, implemented in the WEKA toolbox: classification and regression trees systems, neural networks and Bayesian naive classification. The database used contained 362 patients evaluated on 37 descriptors.

Relying on the work of Wolters et al. [25], an algorithm for classifying patients in different risk classes of postoperative complications based on their ASA score and other preoperative risk factors is available at Ref. [26].

In Ref. [27], a new model was developed in order to predict the operative risk for the patients based on the ASA classification. A database composed of 1936 patients built on the input of two hospitals was used to predict the operative risks by fitting a logistic regression model.

However the ASA score has proven to be insufficient to predict patient mortality and morbidity in certain types of surgery [28,29]. Recent studies [30,31] have shown that the principal predictor of operative risk is the procedure and that the ASA physical status is the second most important factor.

The ASA scale encodes the patients' health status in one of the following 6 categories:

ASA 1: Healthy person,
ASA 2: Mild systemic disease,
ASA 3: Severe systemic disease,
ASA 4: Severe systemic disease that is a constant threat to life,
ASA 5: A moribund person who is not expected to survive without the surgery,
ASA 6: A declared brain-dead person whose organs are being removed for donor purposes.

The smaller the ASA score, the better the health status of the patient and the lower the risk of a surgery.

Zuidema et al. [32] use decision programming to reproduce the ASA score of 14,349 patients assessed by a heterogeneous group of "anesthesia caregivers". The system takes as input the answers to a structured questionnaire involving 22 questions.

Several supervised machine learning algorithms were tested in Ref. [2] in order to predict the ASA score. Among them, the Support Vector Machine (SVM) algorithm yielded the best predictions.

In Ref. [3], a computer aided diagnosis system aiming to help doctors in pre-anesthesia examination is proposed. Five supervised machine learning techniques are used: SVM, radial basis function (RBF), C4.5 decision tree classifier, K-nearest neighbor (KNN) and multilayer perceptron (MLP). An additional method used in this work is a majority voting rule which consists of assigning the patient to the category in which it has been assigned to by a majority of the 5 other machine learning algorithms. A database containing 898 patients evaluated on multiple attributes was used to train and evaluate the classifiers. The paper aims to offer computer-aided support to physicians relative to four preanesthetic examination issues, namely, determining the patient's ASA score; determining whether or not the patient is accepted for surgery; selecting the type of anesthetic method, either general or local; and determining whether the patient's tracheal intubation is easy or difficult. For each method and each issue, the validation was performed by splitting randomly the dataset in a learning and a test set (cross validation).

In the present paper, we focus on the prediction of the ASA score and the patient acceptance or refusal for surgery. We recall the main results of Ref. [3] regarding these two cases. The attributes taken into account in order to learn a model predicting the ASA score are given in Table 1. A detailed description of the attributes is provided in Appendix S1: Appendix A. Table 1 specifies the range of each attribute and whether a higher value contributes to a better ASA score ("max") or a lower value contributes to a better ASA score ("min") or else, extreme values contribute to a worse ASA score ("max min"). Although other attributes such as sickle cell anemia or malignant hyperthermia sensitivity could have been relevant for assessing the ASA score, the only attributes available in the database are those displayed in Table 1 and therefore they are the only ones used in the current study. Regarding the determination of patient

Table 2 – List of attributes taken into account in the prediction of acceptance or refusal of a patient for surgery score in Ref. [3]. Their domain of variation, measurement unit and the direction of increasing preference are specified.

Attribute	Domain (Unit)	Direction
ASA score	[0–4]	min.
Cerebrovascular accident	[0–2]	min.
Myocardial infarction	[0–2]	min.

Table 3 – Average classification accuracy of the test set when 70% of the examples of the dataset are used as learning set for the prediction of ASA score and acceptance or refusal for surgery.

Learning algorithm	ASA score	Acceptance/Refusal
SVM	0.8752	0.9142
C4.5	0.9154	0.9012
KNN	0.8468	0.9085
MLP	0.8927	0.9292
RBF	0.8333	0.8981
Majority voting	0.9259	0.9407

acceptance or refusal for surgery, three attributes, displayed in Table 2, are taken into account.

Table 3 shows the average classification accuracy (proportion of patients correctly classified) of the test set for the prediction of the ASA score and acceptance/refusal of patient for surgery when 70% of the dataset is used as learning set. The method returning the best results for both cases is the majority voting rule. It restores 93.59% of the assignments for ASA score prediction and 94.07% of the assignments for the acceptance or refusal for surgery.

3. Method

3.1. *Majority rule sorting model*

In order to classify patients for pre-anesthetic examination and accept or refuse them for surgery, we have been searching for a method using a model that is interpretable by doctors and opt for MR-Sort, a MCDA sorting procedure that aims at assigning each alternative of a set, evaluated on multiple criteria, to a category selected among a set of pre-defined and ordered categories. The method has been characterized axiomatically in Refs. [33, 34]. We consider in this paper the MR-Sort model without veto. A complete description of the model can be found in Refs. [4, 35].

We recall the assignment rule for a model composed of 2 categories, C^1 and C^2 and n attributes. Categories are ordered, so that C^2 is preferred to C^1. In MR-Sort, the attributes, also called criteria, should be monotonic, i.e. the preference for an object increases or decreases as a monotonic function of the attribute value. Each criterion has an importance that may vary. This importance is modeled through the use of weights: an important criteria has a bigger weight than a less important one. We denote by w_j the weight associated to the criterion j. Usually,

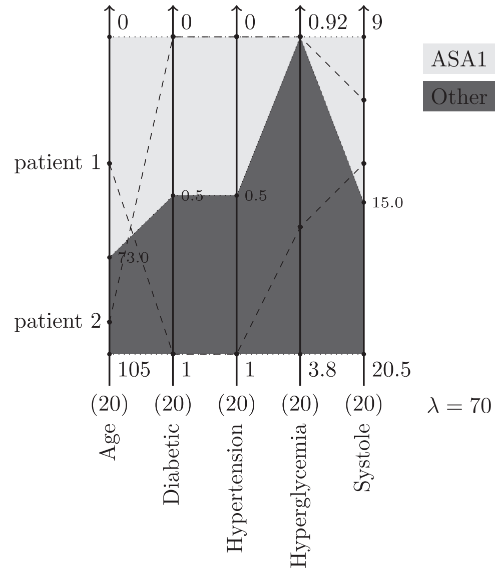

Fig. 1 – Criteria and categories of a fictive MR-Sort model used to determine the patient's ASA score.

the weights sum up to a fixed value, for instance 100, i.e. $\sum_{j=1}^{n} w_j = 100$.

To assign an object a either to C^1 or to C^2, the model compares the object to the profile b^1 delimiting the two categories. The object is assigned to C^2 if it is considered at least as good as the profile b^1. Otherwise it is assigned to category C^1. To be considered at least as good as b^1, an object should have performances at least as good as b^1 on a weighted majority of criteria. The weighted majority is reached when the sum of weights in favor of the object a is equal to or greater than a threshold λ. If the majority is reached, a is assigned to C^2; otherwise it is assigned to C^1. Formally, we express the assignment rule as follows[1]:

$$a \in C^1 \Leftrightarrow \sum_{j: a_j \geq b_j^1} w_j < \lambda,$$

$$a \in C^2 \Leftrightarrow \sum_{j: a_j \geq b_j^1} w_j \geq \lambda.$$

Note that these formulas are valid in case all criteria are of the type "the more the better".

As an example, consider that the model represented in Fig. 1 is used to determine whether a patient has an ASA score equal to 1 or not. The model is built such as each criterion has the same importance: the 5 criteria have equal weights. The majority threshold is set to 70. With this setting, a patient is assigned to category ASA 1 if he/she is at least as good as the profile delimiting the category on 4 criteria. As an example, patient 1 represented in Fig. 1 is not assigned to category ASA 1 because he/she has worse performance than the profile delimiting the class ASA 1 from the class ASA 2 on 3 criteria. Indeed he is diabetic, he suffers from hypertension and his blood glucose

[1] These formulas apply when the criteria are of the type "the more the better". They can be easily adapted to the case "the less the better".

level is higher than 0.92 g/l. On the contrary, patient 2 is assigned to ASA 1 because his performances are at least equal as or better than the one of the profile on 4 criteria. Patient 2 is not diabetic, he does not suffer from hypertension, his blood glucose level is not higher than 0.92 g/l and his systolic blood pressure is lower than 15 cm Hg. This coalition of 4 criteria represents a sufficient majority to assign patient 2 to class ASA 1 since it corresponds to 80% of the weights.

For a MR-Sort model composed of p categories, an object a is assigned to a category C^h if the two following conditions are met:

1. a is equal to or better than b^{h-1} on a weighted majority of criteria;
2. a is not equal to or better than b^h on a weighted majority of criteria.

Formally, the assignment rule writes:

$$a \in C^h \Leftrightarrow \sum_{j:a_j \geq b_j^{h-1}} w_j \geq \lambda \quad \text{and} \quad \sum_{j:a_j \geq b_j^{h}} w_j < \lambda. \qquad (1)$$

A MR-Sort model composed of p categories and n criteria involves the elicitation of $pn + 1$ parameters, i.e. n weights, $(p - 1)n$ profiles evaluations and a majority threshold. Eliciting these parameters by explicitly questioning an expert is not easy as the model involves many parameters. Experts often prefer to provide examples of assignments instead of explicitly eliciting the model parameters. That is why, in the past years, several papers have been devoted to the learning of MR-Sort model parameters on the basis of assignment examples. Mixed integer programs are proposed in Refs. [4, 36] in order to learn partially or globally the parameters of a MR-Sort model. However these algorithms are not efficient enough for the cases we want to deal with in this paper because they cannot handle large data sets.

In Ref. [35], a heuristic algorithm has been presented in order to learn all the parameters of an MR-Sort model based on a large set of assignment examples. The input of the algorithm is a set of assignment examples described by their vector of performances. The output is a MR-Sort model that tends to be compatible with as many examples as possible.

3.2. *MR-Sort for the prediction of ASA score and patient acceptance or refusal for surgery*

Compared to other machine learning algorithms, the MR-Sort rule can be more easily interpreted. It is possible to describe the model as a set of simple rules. In the present paper, we use the MR-Sort algorithm elaborated in Ref. [35] to learn the parameters of MR-Sort models predicting the ASA score of a patient and whether or not he/she is accepted for surgery. To address the two cases, we reuse the dataset of Ref. [3] which is composed of 898 patients. Table 4 shows the distribution of the patients of the data set among the first four ASA classes. No patient has an ASA score above 4 and a majority of them has an ASA score below 3. Table 5 shows the proportion of patients accepted and refused for surgery.

The ASA score of a patient is determined based on 14 attributes (see Table 1). The acceptance or refusal for surgery is determined based on 3 attributes (see Table 2) including the patient's ASA score.

Table 4 – ASA data set: number of patients per ASA score.

ASA score	Number of instances (proportion in percents)
ASA 1	211 (23%)
ASA 2	396 (44%)
ASA 3	239 (27%)
ASA 4	52 (06%)

Table 5 – ASA data set: number of patients accepted and refused.

Patient status	Number of instances (proportion in percents)
Accepted	762 (85%)
Refused	136 (15%)

As using a MR-Sort model requires attributes which are monotonic, it implies that some attributes in Table 1 have to be modified in order to obtain a monotonic scale for each attribute. Indeed, attributes "Heart rate" and "Blood glucose level" are not monotonic. The preference on these attributes increases and then decreases as a function of the attribute value. As an example, a person with a heart rate of 70 beats per minute (bpm) is preferred to someone who has a heart rate of 50 bpm and to someone with a heart rate of 100 bpm. In order to have criteria for which the preference either increases or decreases as a function of its value, these two attributes are split in four sub-attributes: "Bradycardia", "Tachycardia", "Hypoglycemia" and "Hyperglycemia". Table 6 lists the four criteria and whether they should be maximized or minimized.

For instance, the blood glucose level is split into two criteria: Hypoglycemia and Hyperglycemia. The Hypoglycemia criterion measures the severity of hypoglycemia (whenever the patient's blood glucose level is below 0.92 g/l) and similarly for the Hyperglycemia criterion. Therefore, a hyperglycemic patient is assessed as very good w.r.t. to the hypoglycemic criterion and conversely.

Attributes used to determine the acceptance or refusal of a patient for surgery (Table 2) are all monotonic. There is no need to transform any of them.

4. Results

4.1. *Quality of ASA score and acceptance prediction using MR-Sort*

To assess whether or not MR-Sort gives better results than other machine learning algorithms, we perform a cross validation

Table 6 – Attributes split in two in order to determine the ASA score with a MR-Sort model.

Attribute		Domain (Unit)	Direction
Heart rate	Bradycardia	[50–70] (bpm)	max.
	Tachycardia	[70–123] (bpm)	min.
Blood glucose level	Hypoglycemia	[0.5–0.92] (g/l)	max.
	Hyperglycemia	[0.92–3.8] (g/l)	min.

Table 7 – Prediction of the ASA score: average classification accuracy of the learning and test sets for different sizes of learning set (30%, 50%, 70% of the dataset).

		Learning set	Test set
CA	30%	0.9862 ± 0.0064	0.9469 ± 0.0124
	50%	0.9829 ± 0.0053	0.9553 ± 0.0101
	70%	0.9810 ± 0.0045	0.9615 ± 0.0129
AUC	30%	0.9958 ± 0.0029	0.9830 ± 0.0067
	50%	0.9950 ± 0.0022	0.9858 ± 0.0053
	70%	0.9943 ± 0.0021	0.9878 ± 0.0053

on the dataset and compare the results to the ones obtained in Ref. [3]. The cross validation is done by using successively 30%, 50%, 70% of the database as learning set and the rest as test set. The split between learning and test alternatives is done at random. For a given size of learning and test sets, the cross validation is repeated 100 times, each time with different learning and test sets.

The comparison of machine learning algorithms with the MR-Sort algorithm is done by measuring two indices:

1. classification accuracy (*CA*): it corresponds to the proportion of alternatives correctly assigned among the total number of alternatives;
2. area under the curve (*AUC*): it corresponds to the probability that a classifier will assign an alternative chosen at random from a lower class lower than another alternative chosen at random from an upper class [37–39].

First, we assess the ability of the MR-Sort algorithm to return a model that is compatible with the highest number of examples. The results are given in Table 7. Their variability is assessed by the standard deviation of the results on the 100 randomly drawn learning and test sets. Compared to results obtained with other machine learning algorithms, given in Table 3, we observe that the classification accuracy with MR-Sort is significantly better. MR-Sort enables to gain almost 4% in classification accuracy as compared to the majority voting algorithm used in Ref. [3] (see Table 3). We also note that the value of the area under the curve is quite high which means that the model can efficiently discriminate alternatives from different classes.

The same experiment is done for the prediction of the patient acceptance or refusal for surgery. Table 8 shows the results obtained with the MR-Sort algorithm. We observe that the model can restore at least 92% of the examples. Compared to each of the five supervised classification techniques used in Ref. [3], MR-Sort has similar performances, but it is about 2% less efficient than the majority voting rule applied to the results of the five techniques. Regarding the area under the curve, we note that the method is less efficient than for the prediction of the ASA score.[2]

Table 8 – Prediction of patient acceptance/refusal for surgery using three attributes for different sizes (30%, 50%, 70%) of the learning set.

		Learning set	Test set
CA	30%	0.9268 ± 0.0121	0.9207 ± 0.0097
	50%	0.9252 ± 0.0084	0.9241 ± 0.0092
	70%	0.9259 ± 0.0055	0.9235 ± 0.0129
AUC	30%	0.7604 ± 0.0377	0.7509 ± 0.0162
	50%	0.7521 ± 0.0235	0.7513 ± 0.0246
	70%	0.7536 ± 0.0148	0.7507 ± 0.0346

We will see in Section 5 that replacing the ASA score by all the attributes that were used to predict it considerably improves the performance of MR-Sort in the prediction of acceptance for surgery. In the rest of this section, however, we keep investigating the models and data analyzed in Ref. [3] in order to fully compare our results with theirs.

4.2. *Explaining predictions and interpretability*

Machine learning algorithms often operate as black boxes. It is difficult for the user to interpret the resulting models. Compared to machine learning algorithms, MR-Sort is a model whose parameters can be interpreted in order to explain the assignments. In this subsection, we use the full ASA data set to learn MR-Sort models restoring as many examples as possible. We select one of the learned models and describe it.

4.2.1. *Reducing the number of attributes*

The heuristic algorithm we use for learning a MR-Sort model [40] has two main characteristics. First, it is a randomized algorithm that evolves a population of models with the aim of improving them iteratively. The initial step of the algorithm execution consists of an initialization of the models in the population. Second, each iteration in this improvement process involves a phase in which the current profiles are modified followed by a phase in which the criteria weights are adjusted.

To simplify the model as much as possible, we identify attributes having little influence. Therefore, we run 100 times the randomized algorithm starting with populations of 20 MR-Sort models that are also generated in a randomized manner. The algorithm is configured to run 20 times the heuristic improving the profiles and 20 times the heuristic adjusting the weights. At the end of each run of the algorithm, we record the model restoring the highest number of assignments. In the 100 models eventually obtained, we observe that several criteria are not often used i.e. their weight is set to zero in the learned model. The histogram in Fig. 2 shows how many times each criterion was discarded among the 100 MR-Sort models. With 16 criteria, we observe that “Bradycardia”, “Tachycardia” and “Hyperglycemia” are three attributes that are discarded in more than 75% of the models. This shows that taking these criteria into account is not crucial, in view of the other criteria.

To simplify the model, we apply a leave-one-out procedure, removing one attribute at a time from the model. For a data set involving 16 attributes, we repeat the experiment 16 times, each time with another subset of 15 attributes and we compute the average classification accuracy of the models for the different subsets. Finally, we remove the attribute leading to the least decrease in classification accuracy. The same procedure is repeated for 15 criteria and so on.

[2] Note that the AUC index in Table 7 is multiclass and computed as an average of binary class AUCs [39]. Therefore it can be interpreted as an average probability and compared to the binary classification AUC in Table 11.

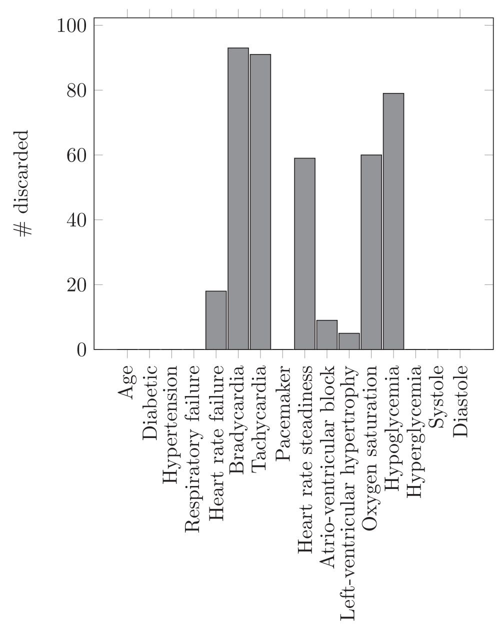

Fig. 2 – Frequency with which each criterion has been discarded among the 100 best models corresponding to 100 runs of the randomized algorithm.

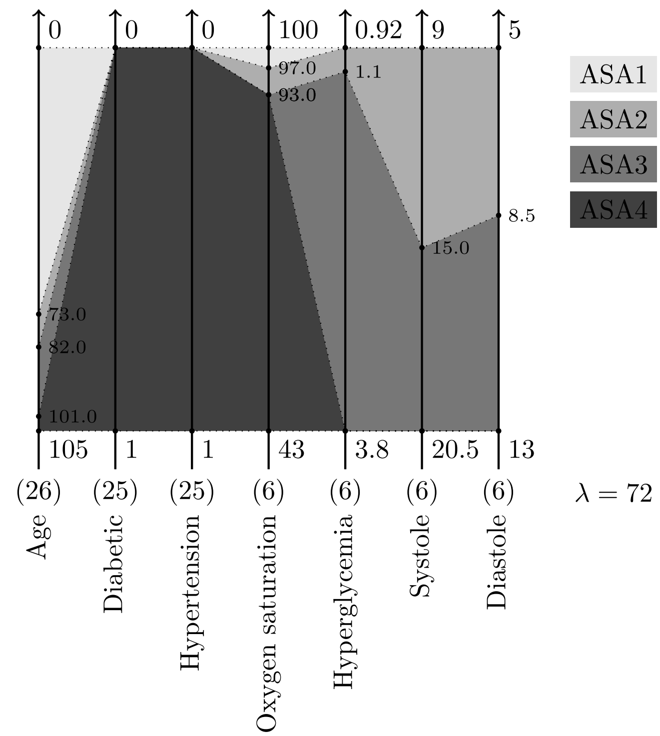

Fig. 4 – MR-Sort model for the prediction of the ASA score. Values in parentheses below the axis are the criteria weights in the model.

However, we did not apply this principle mechanically. Although some criteria may appear as redundant when many criteria are used, they may be viewed by experts as key factors in the determination of the ASA score. From Fig. 2, we observe that "Oxygen saturation", for instance, is considered a dispensable criterion in 60% of the best models when all 16 criteria are used. In spite of this, we chose to keep this criterion among the explanatory variables in the process of iterative elimination of criteria. The reason for this was that it is considered as important by experts, i.e. in our case, by the anesthesiologists working in the 5 hospitals from Western Algeria who provided the cases recorded in the database. Oxygen saturation was the only criterion that was "forced", in the initial steps of the elimination process, to remain in the set of explanatory variables on the basis of expert advice.

Fig. 3 shows the evolution of the average classification accuracy (CA) and area under the curve (AUC) when reducing from 16 to 5 the number of attributes in the model.

We observe that the classification accuracy and area under the curve slowly decrease when attributes are removed. The loss becomes more important when the model passes from 8 to 7 attributes and the classification accuracy declines with more than one percent. The area under the curve remains stable up to 7 criteria. It decreases with more than one percent when the number of attributes goes from 7 to 6. Passing from 6 to 5 criteria results in a decrease of more than 3 percent of the AUC.

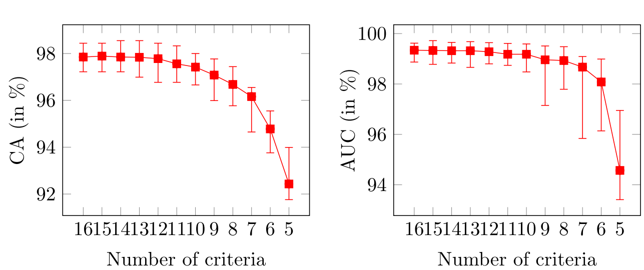

Fig. 3 – Evolution of the classification accuracy (CA) and area under the curve (AUC) when using 16 to 5 attributes.

4.2.2. *MR-Sort model for the prediction of the ASA score*

In agreement with doctors, we choose to keep a model using 7 attributes in order to predict the ASA score of a patient. The attributes taken into account are "Age", "Diabetic", "Hypertension", "Oxygen saturation", "Hyperglycemia", "Systole" and "Diastole". The various algorithm runs provide several models having similar classification accuracy and area under the curve. Some of these models are given in Appendix S1: Appendix B.

Among the 100 best MR-Sort models obtained, we select the one represented in Fig. 4. This model proves able to restore the ASA score of 96.21% of the patients in the learning set, with an AUC equal to 98.48%. The confusion matrix displayed in Table 9 specifies, among the patients classified in a given ASA category (ASA 1, ASA 2, ASA 3 or ASA 4), the number of patients whose predicted ASA category is $\widehat{\text{ASA 1}}$, $\widehat{\text{ASA 2}}$, $\widehat{\text{ASA 3}}$,

Table 9 – Prediction of patients' ASA score: confusion matrix.

	$\widehat{\text{ASA 1}}$	$\widehat{\text{ASA 2}}$	$\widehat{\text{ASA 3}}$	$\widehat{\text{ASA 4}}$
ASA 1	202	9	0	0
ASA 2	11	382	3	0
ASA 3	6	5	228	0
ASA 4	0	0	0	52

or $\widehat{\text{ASA 4}}$. For instance, among the 211 patients in the ASA 1 category, 202 were correctly predicted by the model, while 9 were incorrectly classified $\widehat{\text{ASA 2}}$ (none were predicted $\widehat{\text{ASA 3}}$ or $\widehat{\text{ASA 4}}$).

Using MR-Sort, the patient attribute vectors are compared to the profiles delimiting the categories in ascending order, i.e. the comparison begins with the profile delimiting the category ASA 4 from ASA 3. To be assigned to a category, a patient should be at least as good as the lower profile of that category and not as good as its upper profile. In the model given in Fig. 4, a patient is as good as the profile if his/her attribute values are at least as good as those of the profile on each criterion of one of the following four criteria coalitions:

1. {Age, Diabetic, Hypertension};
2. {Age, Diabetic, Hyperglycemia, Oxygen saturation, Systole, Diastole};
3. {Age, Hypertension, Hyperglycemia, Oxygen saturation, Systole, Diastole};
4. {Diabetic, Hypertension, Hyperglycemia, Oxygen saturation, Systole, Diastole}.

A patient is assigned to a category better than ASA 4 if his/her parameters are as good as the parameters of the profile delimiting the category ASA 4 from ASA 3. In other words, a patient is assigned to a category strictly better than ASA 4 if he/she satisfies the three following conditions:

1. he/she is not older than 101;
2. he/she is not diabetic;
3. he/she does not suffer from hypertension.

The ASA score of a patient is also better than ASA 4 if he/she satisfies two of these conditions in conjunction with an oxygen saturation level at least equal to 93%. On the contrary, a patient who does not satisfy two of the three conditions listed above is always assigned to category ASA 4.

A patient is assigned to a category better than ASA 3 if he/she satisfies the three following conditions:

1. he/she is not older than 82;
2. he/she is not diabetic;
3. he/she does not suffer from hypertension.

The ASA score of a patient is also better than ASA 3 if he/she satisfies two of the above conditions in conjunction with:

1. an oxygen saturation level equal to or greater than 93%;
2. at most a low level of hyperglycemia characterized by a blood glucose level at most equal to 1.10 g/l;
3. a systole pressure level at most equal to 15 cm Hg;
4. a diastole pressure level at most equal to 8.5 cm Hg.

Finally, a patient is always classified in category ASA 1 if the following three conditions are met:

1. he/she is not older than 73;
2. he/she is not diabetic;
3. he/she does not suffer from hypertension.

A patient is also assigned to ASA 1 if he/she satisfies two of these conditions in conjunction with:

1. an oxygen saturation level equal to or greater than 97%;
2. No hyperglycemia, characterized by a blood glucose level equal to or smaller than 0.92 g/l;
3. a systole pressure level equal to or smaller than 9.4 cm Hg;
4. a diastole pressure level equal to or smaller than 5.4 cm Hg.

4.3. *MR-Sort model for the prediction of patient acceptance/refusal for surgery*

The decision of accepting or refusing a patient for surgery is made on the basis of his/her parameter values on the three criteria listed in Table 2. As for the prediction of the ASA score, we use the full data set as learning set to obtain a MR-Sort model interpretable by doctors. We run the (randomized) algorithm a hundred times to obtain a set of models. By using the 898 patients in the database as learning set, we obtain an average classification accuracy equal to 0.9254 and an *AUC* equal to 0.7537. Many of the 100 best models obtained by running the algorithm repeatedly are identical. Fig. 5 shows one of the obtained models. In this particular case, it should be noted that the threshold value $\lambda = 100$ imposes unanimity and therefore the particular values of the weights are irrelevant.

According to this model, a patient is accepted for surgery if he/she fulfills the three following conditions:

1. his/her ASA score is better than ASA 4;
2. he/she has not been subject to a cerebrovascular accident.
3. he/she has not been subject to a myocardial infarction.

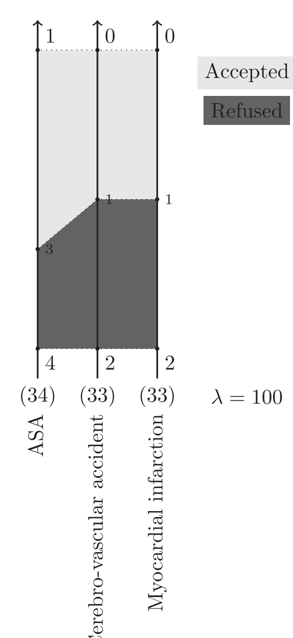

Fig. 5 – MR-Sort model for the prediction of the patient acceptance or refusal for surgery. Values in parentheses below the axis are the criteria weights in the model.

Table 10 – Prediction of patient acceptance/refusal for surgery: confusion matrix.

	$\widehat{\text{Accepted}}$	$\widehat{\text{Refused}}$
Accepted	762	0
Refused	67	69

The confusion matrix is given in Table 10. We note that this model has a bias toward "acceptance for surgery". All patients that should be accepted are correctly classified by the model. However, the model also accepts almost 50% of the patients that should be refused for surgery.

5. Discussion

The experimentations described in the previous section show that MR-Sort is more efficient than other machine learning algorithms for the prediction of the ASA score. As an additional advantage w.r.t. machine learning methods, MR-Sort models can be formulated as rules that are easily understandable by physicians. The rules issued from the learning algorithm can also be modified by the users, and the classification accuracy of the adapted rule can be assessed on the same dataset. For instance, doctors could consider that 100 would be a more reasonable age limit than 101 for the profile delimiting ASA 4 from ASA 3. By using the modified rule on the dataset, one can measure the consequences of such a change in terms of accuracy.

For the prediction of the decision of accepting or refusing a patient for surgery, we observe that MR-Sort yields results that are of similar quality, in terms of accuracy, as the five individual machine learning algorithms used in Ref. [2]. However, making the decision advocated by the majority of these five algorithms leads to a slightly superior prediction accuracy.

In order to improve the efficiency of MR-Sort for the prediction of patient acceptance or refusal for surgery, one might consider replacing the ASA score by the more detailed information contained in the 16 attributes that were used to determine it. We applied the MR-Sort algorithm to learn such a rule using this detailed description of the patients plus the two other criteria, i.e. "Cerebrovascular accident" and "Myocardial Infarction". The results of our experiments are displayed in Table 11. We observe that the CA and AUC are definitely improved by replacing the ASA criterion by the 16 criteria used to determine it. The CA is about 3 percent better than with the individual machine learning algorithms and about 1% better than with the majority voting rule, when the size of the learning set is 70% of the whole dataset. The most striking difference with the MR-Sort model obtained for patients described by only 3 criteria appears in the value of the AUC which increases from 75% with 3 criteria to 91% with 16 criteria.

Table 11 – Prediction of patient acceptance/refusal for surgery using 18 attributes for different sizes (30%, 50%, 70%) of the learning set.

		Learning set	Test set
CA	30%	0.9794 ± 0.0086	0.9347 ± 0.0156
	50%	0.9701 ± 0.0063	0.9475 ± 0.0113
	70%	0.9668 ± 0.0049	0.9525 ± 0.0133
AUC	30%	0.9672 ± 0.0272	0.9129 ± 0.0338
	50%	0.9486 ± 0.0267	0.9188 ± 0.0277
	70%	0.9281 ± 0.0277	0.9085 ± 0.0377

Of course, it is not surprising that using more criteria helps to improve the CA and the AUC. It is likely that one could reduce the number of these criteria to a number between 3 and 18 as we did for the prediction of the ASA score. This would allow to extract rules involving a reasonable number of criteria that would allow to predict acceptance/refusal to surgery with a satisfactory accuracy. Moreover, these rules could be understood and possibly adapted by physicians.

6. Conclusion

The development of medical computer-aided diagnosis systems is becoming nowadays a very active research field. Indeed, numerous artificial intelligence researchers are striving to propose interpretable, intelligent, automatic systems to help doctors in their routine clinical work.

The objective of this paper is to propose a multiple criteria decision aiding system based on the use of the MR-Sort model. This system enables to accurately assess the patient's ASA score, based on the list of patient attributes in the database used for the tests. Moreover, the underlying MR-Sort model allows for easy interpretation by providing human-readable classification rules. It should however be emphasized that the set of patient attributes available in the database is very limited as compared with the large set of attributes observed and taken into account in clinical practice by anesthesiologists. The study provides nevertheless strong evidence that MR-Sort models are a valuable approach to model such medical expert judgments. Further study should be pursued using a more extensive patients database.

The results obtained with MR-Sort show an improvement over machine learning algorithms as regards classification accuracy on the basis of the same patients database. For the ASA score, the prediction accuracy of the model reaches 96%, and the assignment to an ASA category can be described by simple rules. Regarding the decision of accepting the patient for surgery, the method allows to restore the actual decision for given anesthetizing locations and type of surgery with good accuracy (although not as good as for the ASA score (92%)). Considering more detailed patient description enables to improve the fit.

Further research, based on the same methodology, but taking into account the characteristics of the anesthetizing location and the complexity of surgery, could help in building MR-Sort models accounting for acceptance/refusal decisions in a more general context. Moreover, from a methodologic point of view, it would be relevant to study the possible role of vetoes for modeling the experts judgments [33,34], using veto inference [41].

Acknowledgment

The authors wish to thank Ms. Michèle Parmentier for her assistance and support in revising the English and two anonymous referees whose comments and remarks greatly helped us to improve the initial version of the manuscript.

Appendix. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2016.05.021.

REFERENCES

[1] M. Daabiss, American Society of Anaesthesiologists physical status classification, Indian J. Anaesth. 55 (2) (2011) 111–115.

[2] M.E.A. Lazouni, M. El Habib Daho, N. Settouti, M.A. Chikh, S. Mahmoudi, Machine learning tool for automatic ASA detection, in: Modeling Approaches and Algorithms for Advanced Computer Applications, 2013, pp. 9–16.

[3] M.A. Lazouni, M.A. Chikh, S. Mahmoudi, A new computer aided diagnosis system for pre-anesthesia consultation, J. Med. Imaging Health Inform. 3 (4) (2013) 471–479.

[4] A. Leroy, V. Mousseau, M. Pirlot, Learning the parameters of a multiple criteria sorting method, in: R. Brafman, F. Roberts, A. Tsoukiàs (Eds.), Algorithmic Decision Theory, vol. 6992, Lecture Notes in Artificial Intelligence, Springer Berlin, Heidelberg, 2011, pp. 219–233.

[5] J. Figueira, V. Mousseau, B. Roy, ELECTRE methods, in: J. Figueira, S. Greco, M. Ehrgott (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys, Springer Verlag, Boston, Dordrecht, London, 2005, pp. 133–162.

[6] B. Roy, The outranking approach and the foundations of ELECTRE methods, Theory Decis. 31 (1991) 49–73.

[7] American Society of Anesthesiologists' Relative Value Guide, American Society of Anesthesiologists, Park Ridge, IL, 2015.

[8] F. Dexter, E. Thompson, Relative value guide basic units in operating room scheduling to ensure compliance with anesthesia group policies for surgical procedures performed at each anesthetizing location, J. Am. Assoc. Nurse Anesth. 69 (2) (2001) 120–123.

[9] F. Dexter, R.E. Wachtel, J.C. Yue, Use of discharge abstract databases to differentiate among pediatric hospitals based on operative procedures—surgery in infants and young children in the state of Iowa, Anesthesiology 99 (2) (2003) 480–487.

[10] R.E. Wachtel, F. Dexter, Differentiating among hospitals performing physiologically complex operative procedures in the elderly, Anesthesiology 100 (6) (2004) 1552–1561.

[11] R.E. Wachtel, F. Dexter, Tactical increases in operating room block time for capacity planning should not be based on utilization, Anesth. Analg. 106 (1) (2008) 215–226.

[12] R.E. Wachtel, F. Dexter, D.A. Lubarsky, Financial implications of a hospital's specialization in rare physiologically complex surgical procedures, Anesthesiology 103 (1) (2005) 161–167.

[13] R.E. Wachtel, E.U. Dexter, F. Dexter, Application of a similarity index to state discharge abstract data to identify opportunities for growth of surgical and anesthesia practices, Anesth. Analg. 104 (5) (2007) 1157–1170.

[14] R.E. Wachtel, F. Dexter, B. Barry, C. Applegeet, Use of state discharge abstract data to identify hospitals performing similar types of operative procedures, Anesth. Analg. 110 (4) (2010) 1146–1154.

[15] F. Dexter, J. Ledolter, B.J. Hindman, Quantifying the diversity and similarity of surgical procedures among hospitals and anesthesia providers, Anesth. Analg. 122 (1) (2016) 251–263.

[16] K.J. Carter, N.P. Ritchey, F. Castro, L.P. Caccamo, E. Kessler, B.A. Erickson, Analysis of three decision-making methods: a breast cancer patient as a model, Med. Decis. Making 19 (1) (1999) 49–57.

[17] D. West, P. Mangiameli, R. Rampal, V. West, Ensemble strategies for a medical diagnostic decision support system: a breast cancer diagnosis application, Eur. J. Oper. Res. 162 (2) (2005) 532–551.

[18] R.L. Chatburn, F.P. Primiano, Decision analysis for large capital purchases: how to buy a ventilator, Respir. Care 46 (10) (2001) 1038–1053.

[19] M.J. Liberatore, R.L. Nydick, The analytic hierarchy process in medical and health care decision making: a literature review, Eur. J. Oper. Res. 189 (1) (2008) 194–207. ISSN 0377-2217.

[20] J.R. Figueira, J.A. Dias, S. Matias, B. Roy, M.J. Carvalho, C.E. Plancha, Electre Tri-C, a multiple criteria decision aiding sorting model applied to assisted reproduction, Int. J. Med. Inform. 80 (4) (2011) 262–273.

[21] D.A. Klein, Decision-Analytic Intelligent Systems: Automated Explanation and Knowledge Acquisition, L. Erlbaum Associates Inc, Hillsdale, NJ, USA, 1994.

[22] C. Labreuche, A general framework for explaining the results of a multi-attribute preference model, Artif. Intell. 175 (7–8) (2011) 1410–1448.

[23] L.G. Glance, S.J. Lustik, E.L. Hannan, T.M. Osler, D.B. Mukamel, F. Qian, et al., The surgical mortality probability model: derivation and validation of a simple risk prediction rule for noncardiac surgery, Ann. Surg. 255 (4) (2012) 696–702.

[24] S. Karpagavalli, K. Jamuna, M. Vijaya, Machine learning approach for pre-operative anaesthetic risk prediction, Int. J. Recent Trends Eng. 1 (2) (2009).

[25] U. Wolters, T. Wolf, H. Stützer, T. Schröder, ASA classification and perioperative variables as predictors of postoperative outcome, Br. J. Anaesth. 77 (2) (1996) 217–222.

[26] The Medical Algorithms Company. Model of Wolters et al for predicting postoperative complications from ASA classification and perioperative variables. http://www.medicalalgorithms.com/model-of-wolters-et-al-for-predicting-postoperative-complications-from-asa-classification-and-perioperative-variables. Accessed 11.10.15.

[27] A. Donati, M. Ruzzi, E. Adrario, P. Pelaia, F. Coluzzi, V. Gabbanelli, et al., A new and feasible model for predicting operative risk, Br. J. Anaesth. 93 (3) (2004) 393–399.

[28] K. Chijiiwa, K. Yamaguchi, H. Yamashita, Y. Ogawa, J. Yoshida, M. Tanaka, ASA physical status and age are not factors predicting morbidity, mortality, and survival after pancreatoduodenectomy, Am. Surg. 62 (9) (1996) 701–705.

[29] W.D. Owens, American Society of Anesthesiologists physical status classification system is not a risk classification system, Anesthesiology 94 (2001) 378.

[30] J.E. Dalton, A. Kurz, A. Turan, E.J. Mascha, D.I. Sessler, L. Saager, Development and validation of a risk quantification index for 30-day postoperative mortality and morbidity in noncardiac surgical patients, Anesthesiology 114 (6) (2011) 1336–1344.

[31] M.J.G. Sigakis, E.A. Bittner, J.P. Wanderer, Validation of a risk stratification index and risk quantification index for predicting patient outcomes: in-hospital mortality, 30-day mortality, 1-year mortality, and length-of-stay, Anesthesiology 119 (3) (2013) 525–540.

[32] X. Zuidema, R.C. Tromp Meesters, I. Siccama, P.L. Houweling, Computerized model for preoperative risk assessment, Br. J. Anaesth. 107 (2011) 180–185.

[33] D. Bouyssou, T. Marchant, An axiomatic approach to noncompensatory sorting methods in MCDM, I: the case of two categories, Eur. J. Oper. Res. 178 (1) (2007) 217–245.

[34] D. Bouyssou, T. Marchant, An axiomatic approach to noncompensatory sorting methods in MCDM, II: more than two categories, Eur. J. Oper. Res. 178 (1) (2007) 246–276.
[35] O. Sobrie, V. Mousseau, M. Pirlot, Learning a majority rule model from large sets of assignment examples, in: P. Perny, M. Pirlot, A. Tsoukiás (Eds.), Algorithmic Decision Theory, Lecture Notes in Artificial Intelligence, Springer, 2013, pp. 336–350.
[36] O. Cailloux, P. Meyer, V. Mousseau, Eliciting ELECTRE TRI category limits for a group of decision makers, Eur. J. Oper. Res. 223 (1) (2012) 133–140.
[37] Tom Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (8) (2006) 861–874.
[38] D.J. Hand, R.J. Till, A simple generalisation of the area under the ROC curve for multiple class classification problems, Mach. Learn. 45 (2) (2001) 171–186, http://dx.doi.org/10.1023/A:1010920819831.
[39] W. Waegeman, B. De Baets, L. Boullart, ROC analysis in ordinal regression learning, Pattern Recognit. Lett. 29 (1) (2008) 1–9.
[40] O. Sobrie, V. Mousseau, M. Pirlot, Learning the parameters of a multiple criteria sorting method from large sets of assignment examples, in: 77th Meeting of the EWG on MCDA, Rouen, France, 2013.
[41] O. Sobrie, V. Mousseau, M. Pirlot, New veto rules for sorting models, in: 20th Conference of the International Federation of Operational Research Societies, Barcelona, Spain, 2014.