A new decision support model for a pre-anesthetic examination

Olivier Sobrie   Mohammed EL Amine Lazouni
Saïd Mahmoudi   Vincent Mousseau   Marc Pirlot

December 12, 2015

Abstract

In recent years, the principal challenges related to the field of anesthesia and intensive care consist of reducing both anesthetic risks and mortality rate, as well as providing better and more efficient assistance to Doctors Specialized in Anesthesia (DSA).

In this paper we propose a computer aided diagnosis system aiming to help doctors in the pre-anesthesia examination. The proposed approach is based on MR-Sort, a multiple criteria decision analysis method. In order to evaluate our algorithms, we use a local anesthetic database composed of 898 patients. The proposed method includes two steps. The first one is devoted to an automatic detection of ASA (American Society of Anesthesiologists) scores. In the second step, a decision making process is applied in order to accept or refuse the patient for surgery. The classification results obtained by using our system prove the reliability and the coherence of the proposed method. In addition, the MR-Sort technique used gives good results for the two steps and manifests a great application prospect for supporting clinic aided diagnosis.

We assess the MR-Sort method and compare it to other machine learning algorithms. In the paper, we also advocate the use of multiple criteria decision analysis to obtain a model that can be explained to doctors.

Keywords: *Multiple Criteria Decision Analysis, American Society of Anesthesiologists Scores, MR-Sort, Classification.*

1 Introduction

With regards to anesthesiology, assessing the patient's physical health state is a crucial step before deciding whether or not he can be accepted for surgery. This information allows DSA to identify the anesthesia type and determine the ease of the tracheal intubation. A commonly used system to determine the patient's health state 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 person" whose organs are being removed for donor purposes. Up to now, the classification of a patient does not rely on a well-established and objective method but on the subjective advice of one or several doctors. That is why we propose in this paper a new Multiple Criteria Decision Analysis (MCDA) method in order to determine the ASA score and the acceptance or refusal for the patient's surgery.

In medicine, Multiple Criteria Decision Analysis (MCDA) can be used for various applications going from cancer diagnosis to health care settings. Among medicine applications based on decision aid methods, we observe that many of them are using the Analytic Hierarchical Process method. Few of them are using other families of MCDA methods like outranking ones or methods based on additive value function.

In anesthesiology scientific literature, very few works related to preoperative patient classification were carried out. In [12, 13], this concern has been treated by using several machine learning algorithms. A drawback of such algorithms is that they are not easily interpretable by doctors. Indeed, these algorithms are used as blackboxes and are difficult to interpret. It is often difficult to understand the patient classification by referring to the algorithm parameters. For instance, with a neural network algorithm like the multilayer perceptron, it is difficult to understand the classification by looking at synapses values and activation functions, even more if the model involves a lot of variables.

As far as we know and according to our research in existing literature, there is no work dealing with multiple criteria decision analysis and ASA physical status classification.

In this paper, we advocate the use of more interpretable models, like the MCDA ones, as we assume that doctors prefer to be able to understand the model leading to the classification of a patient, which can be achieved by an easily interpretable model.

Besides, we propose the use of the MR-Sort MCDA procedure to determine the ASA score of patient evaluated on multiple criteria. We suggest to learn the parameters of MR-Sort models on basis of a large database containing information of about 898 patients evaluated on multiple criteria and who have been classified in one of the six ASA classes. The idea behind is to obtain a set of models that can be easily interpreted by doctors. We should remember that our database contains only four of the six ASA score classes, knowing that ASA scores 5 and 6 have not been collected because the plants from which we collected our data are not included within the donating organ ones.

This paper is organized as follows. After the introduction, we give an overview of the existing anesthesiology scientific literature related to the decision aided 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 to infer the parameters of this method. The fourth section details experimental results that have been obtained with the MR-Sort method in view of predicting the patients ASA score. The fifth section presents other experimental results obtained in view of predicting whether or not a patient can be accepted for a surgical operation. Finally, in the last section, a set of conclusions.

2 Literature review

2.1 Multicriteria Decision Analysis in medicine

In medicine, the decision aid methods are used for various applications going from cancer diagnosis and treatment [19, 4], to the selection of technologies in health care settings [5]. Many medical applications are based on Analytic Hierarchical Process (AHP). In [15] an overview of existing applications using AHP methods is presented. However, few articles are dealing with other MCDA methods like the ELECTRE ones. In [7], the ELECTRE TRI-C sorting procedure is used in the context of assisted reproduction, couples are assigned in 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, there are no works dealing with the multiple criteria decision analysis and ASA score determination by using the above mentioned method while the ASA score is widely used by all DSA's during their pre-anesthetic examinations.

Since in medicine it is preferred to use well-know methods that allow to explain a choice [9], we advocate in this paper the use of an outranking model, based on MR-Sort, an ELECTRE method, and the use of an additive value function model.

2.2 Decision Support Systems for Anesthesia

(author?) [8] proposed a probabilistic model for evaluating the surgical mortality. The objective of this work is to predict the patient's mortality after a non-cardiac surgery, in order to diminish 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 of high or intermediate risk and whether or not the surgery is urgent. A huge database composed of 298,772 patients has been gathered from different medical plants 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 [11]. This system assesses three classifications techniques based on supervised learning. The assessment has been done by using WEKA software for the three following classification techniques: classification and regression trees systems, neural networks and Bayesian naïve classification. The database used contained 362 patients evaluated on 37 descriptors.

In [16], an automatic system classifying patients in different anesthetic risk levels has been developed. A modified version of the method has been presented in [10]. One of the descriptors used to predict the patient's risk level is the ASA score of the patient.

In [6], a new model was developed in view of predicting the operative risk for the patients. The objective of this work was to predict the mortality and the morbidity of patients on basis of the ASA classification. A database composed of 1936 patients built on the input of two hospitals has been used to predict the operative risks by means of a machine learning algorithm, called logistic regression.

Several machine learning algorithms have been used in [12] in view of predicting patients' ASA score. In the field of anesthesia, the ASA score is widely used in pre-anesthetic examinations. It is used to assess patients health status before a surgical operation. Each patient is evaluated on a scale from 1 to 6, reflecting his health status. The scale is composed 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.

Several supervised machine learning algorithms have been assessed in [12] in view of predicting patients ASA scores. Results have shown that Support Vector Machine (SVM) algorithm gives the best results.

In [13], a computer aided diagnosis system aiming to help doctors in pre-anesthesia examination has been suggested. Five supervised machine learning techniques have been used: SVM, Radial Basis Function (RBF), C4.5 decision tree classifier, K-nearest neighbor (KNN) and Multilayer Perceptron (MLP). The last algorithm used in this work is the Majority Voting which consists of assigning the patient to the category in which it has been assigned to by a majority of the 5 others machine learning algorithms. A database containing 898 patients evaluated on multiple attributes has been used to evaluate the classifiers. This paper first treat the determination of the ASA score of each patient. The second concern consisted of determining whether or not the patients are accepted for surgery. The third one is the selection of the type of anesthetic method either general or local. The last one consists of determining whether patient's tracheal intubation is easy or hard. For each algorithm and each concern, a cross validation was performed. The validation was done by splitting randomly the dataset in a learning and test set.

In the present paper, we are interested in the determination of ASA score and patient acceptance or refusal for surgery. We remind the main results of [13] regarding these two cases. The attributes taken into account in order to learn a model predicting the ASA score are given in Table 1. The table contains the domain of each attribute and whether if it should be maximized and/or minimized. For the determination of patient acceptance or refusal for surgery, 3 attributes, given in Table 2, are taken into account.

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.

Table 1: List of attributes taken into account in the prediction of the ASA score in [13]

Table 3 shows the average classification accuracy of the test set for the prediction of the ASA score and acceptation 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. It restores 93.59% of the

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

Table 2: List of attributes taken into account in the prediction of acceptance or refusal of a patient for surgery score in [13]

assignments for ASA score prediction and 94.07% of the assignments for the 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

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/Refusal for surgery

3 Method

3.1 Majority Rule Sorting model

In order to classify patients for pre-anesthetic examination and accepting or refusing 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, in a category selected among a set of pre-defined and ordered categories. The method has been characterized axiomatically in [1, 2]. We consider in this paper the MR-Sort model without veto. A complete description of the model can be found in [14, 17].

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 modelled 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, 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 in C^1 or in C^2, the model compares the object to the profile b^1 delimiting the two categories. The object is assigned in C^2 if it is considered at least equal or better than the profile b^1. Otherwise it is assigned in category C^1. To be considered equal or better than b^1, an object should have at least equal or better performances than 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 or greater than a threshold λ. If the majority is reached, a is assigned in C^2, otherwise it is assigned in C^1. Formally, we express the assignment rule as follows:

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

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

.

As an example, consider that the model given on Figure 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: their weights are equal on the 5 criteria. The majority threshold is set to 70. With this setting, a patient is assigned to category ASA 1 if he is at least equal or better than the profile delimiting the category on 4 criteria. As an example, patient 1 represented on Figure 1 is not assigned in category ASA 1 because he has worse performance than the profile delimiting the class ASA 1 from the lower ones on 3 criteria. Indeed he is diabetic, he has hypertension and his blood glucose level is higher than 0.92 cm Hg. On the contrary, patient 2 is assigned in ASA 1 because his performances are at least equal or better than the one of the profile on 4 criteria. Patient 2 is not diabetic, he doesn't suffer from hypertension, his blood glucose level is not higher than 0.92 cm Hg and his systolic blood pressure is lower than 15 cm Hg. This coalition of 4 criteria represent a sufficient majority to assign patient 2 in class ASA 1 since his profile corresponds to 80% of the weights.

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

1. a is equal or better than b^{h-1};
2. a is not equal or better than b^h.

Formally, the assignment rule writes:

$$a \in C^h \iff \sum_{j:a_j \geq b_j^{h-1}} w_j \geq \lambda \text{ and } \sum_{j:a_j \geq b_j^h} w_j < \lambda \tag{1}$$

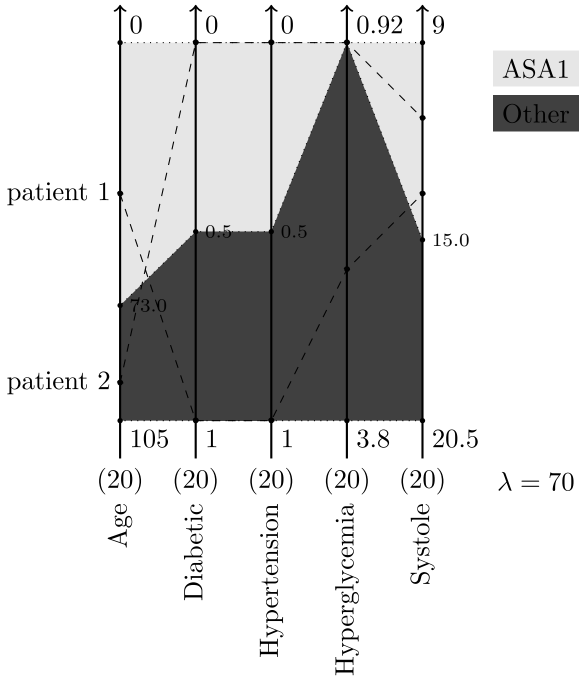


Figure 1: Criteria and categories of a fictive MR-Sort model used to determined patient ASA score.

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 explicitly is not easy as it involves a lot of values to determine. One often prefers to give 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 models parameters on basis of assignment examples. Mixed integer programs are proposed in [3, 14] 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 [17], a metaheuristic has been presented in order to learn all the parameters of an MR-Sort model on basis of large set of assignment examples and their vector of performances. The input of the algorithm is a set of assignment examples. The output is a MR-Sort model that tends to be compatible with as much 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 metaheuristic elaborated in [17] to learn the parameters of MR-Sort models predicting the ASA score of a patient and whether or not he is accepted for surgery. To address the two cases, we reuse the dataset of [13] which is composed of 898 patients. Table 4 gives the repartition 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.

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 4: ASA data set: number of patients per ASA score

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

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

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

As using a MR-Sort model requires attributes which are monotone, it implies that some attributes of Table 1 have to be modified in order to have a monotonic scale for each attribute. Indeed, attributes "Heart rate" and "Blood glucose level" are not monotone. The preference for 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 minutes (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, the attributes are split in four sub-attributes: "Bradycardia", "Tachycardia", "Hyperglycemia" and "Hyperglycemia". Table 6 lists the four criteria and whether it should

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 6: Attributes split in two in order to determine of the ASA score with a MR-Sort model

be maximized or minimized.

Attributes used to determine the acceptance or refusal of a patient for surgery (Table 2) are all monotone. 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 algorithm, we perform a cross validation on the dataset and compare the results to the one obtained in [13]. 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 metaheuristic 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 class an alternative chosen at random from the lower class higher than another alternative chosen at random from an upper class [18].

First, we assess the ability of the MR-Sort metaheuristic to return a model that is compatible with the highest number of examples. The results are given in Table 7. 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. Indeed, the classification accuracy is improved with almost 4% with MR-Sort compared to the Majority Voting algorithm used in [13]. We also note that the value of the area under the curve is high which means that the model can efficiently discriminate alternatives from different classes.

		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

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

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 metaheuristic. We observe that the model can restore 92% of the examples. Compared to the Majority Voting algorithm used in [13], it is about 2% less efficient. Regarding the area under the curve, we note that the algorithm is less efficient than for the prediction of the ASA score.

		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

Table 8: Prediction of Patient Acceptance/Refusal for Surgery using three attributes for different sizes (30%, 50%, 70%) of the data set

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 as learning set for the MR-Sort metaheuristic in order to learn models restoring as much examples as possible. We select one of the models learned with the metaheuristic and describe it.

4.2.1 Reduction of the number of attributes

To simplify the model as much as possible, we identify attributes having the less influence in the model. Therefore, we initialize 100 instances of the metaheuristic. Each instance is initialized at random with a population of 20 MR-Sort models. The metaheuristic is configured to run 20 times the heuristic improving the profiles and 20 times to run the heuristic adjusting the profiles. For each instance of the metaheuristic, we keep the model restoring the highest number of assignments. After running 100 instances of the metaheuristic, we obtain a list of 100 models. We observe that several criteria are not used in the models found by the metaheuristic. The histogram given in Figure 2 shows the number of time each criterion has been 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. It shows that some criteria do not add a lot of information for the determination of the ASA score.

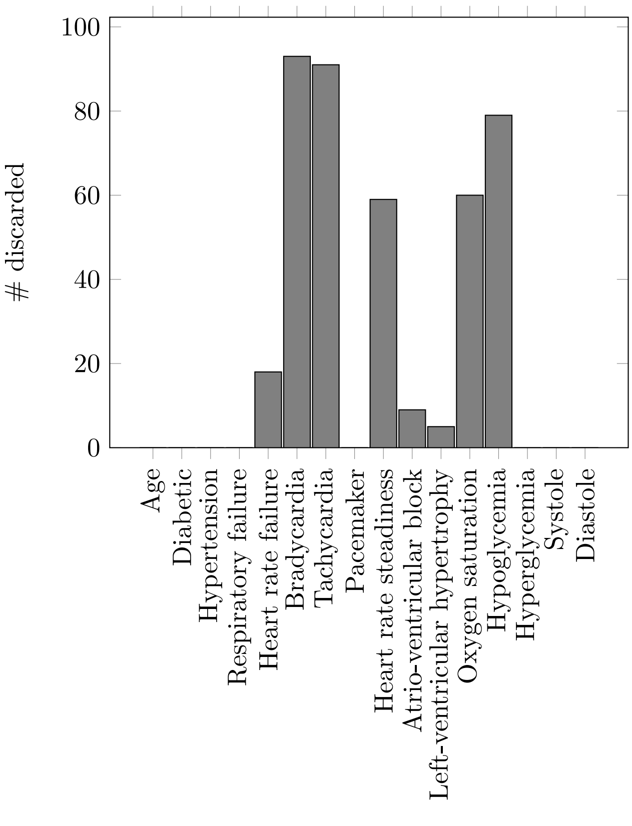


Figure 2: Number of time each criterion has been discarded among the 100 best models corresponding to the 100 instances of the metaheuristic

To simplify the model, we apply a leave-one-out procedure to remove

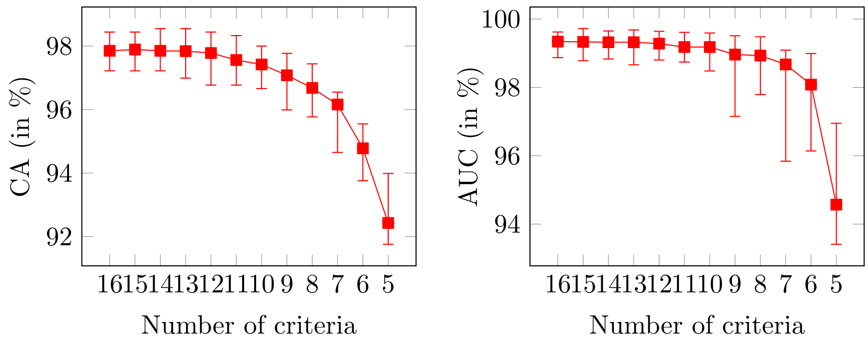


Figure 3: Evolution of the classification accuracy (CA) and area under the curve (AUC) when using 16 to 5 attributes

some attribute of the model. It consists in repeating the experiment by removing one attribute at a time from the data set. For a data set composed of 16 attributes, it means to repeat the experiment 16 times, each time with another subset of 15 attributes. After applying the leave-one-out procedure, we compute the average classification accuracy of the models for the different subsets. Finally, we remove the attribute decreasing the less the classification accuracy. The same procedure is repeated for 15 criteria and so on. We take care of keeping attributes that are considered important by doctors for the determination of the ASA score. As an example, "Oxygen saturation" is not often used in the models but we choose to keep this criterion because doctors consider that it is an important parameter in the determination of the ASA score.

Plot given on Figure 3 shows the evolution of the average classification accuracy (CA) and area under the curve (AUC) when using 16 to 5 attributes in the model.

We observe that the classification accuracy and area under the curve slightly decrease when attributes are removed. The difference 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 percents of the AUC.

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 metaheuristic instances provide several models having similar classification accuracy and area under the curve. Some of these models are given in appendix B.

Among the 100 best MR-Sort models obtained, we keep the one represented in Figure 4. This model can restore the ASA score of 96,21% of the patients of the learning set, with an AUC equal to 98.48%. The confusion matrix is given in Table 9.

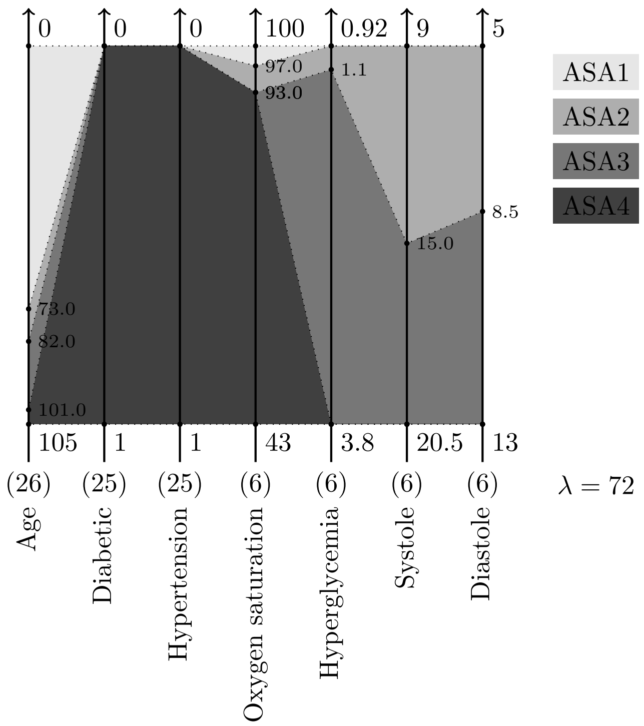


Figure 4: MR-Sort model for the prediction of the ASA score. Values in parenthesis below the axis are the weights of the model.

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

Table 9: Prediction of patient ASA score: confusion matrix.

Using MR-Sort, patient performances are compared to the profiles delimiting the categories in ascending order, i.e. the comparison begins with the profile delimiting the worst categories. To be assigned in 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 Figure 4, a patient is as good as the profile, if his performances are at least equal to those of the profile on each criterion of one of these 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 above ASA 4 if his performances are as good as the performances of the profile delimiting the category ASA 3 from ASA 4. In other words, a patient has always a score greater than 4 if he satisfies the three following conditions:

1. he is 105 years old or younger;
2. he is not diabetic;
3. he doesn't suffer from hypertension.

The ASA score of a patient is also greater than 4 if he satisfies two of these conditions in conjunction with a level of oxygen saturation equal or above 93%. On the contrary, a patient who does not satisfy two of the three conditions listed above is always assigned in category ASA 4. A patient is assigned to a category greater than 3 if he satisfies the three following conditions:

1. he is 82 years old or younger;
2. he is not diabetic;
3. he doesn't suffer from hypertension.

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

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

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

1. he is 73 years old or younger;
2. he is not diabetic;
3. he doesn't suffer from hypertension.

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

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

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

The acceptance or refusal of a patient for surgery is made on basis of its performance 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 metaheuristic hundred times to obtain a set of models. By using the 898 patients of the database as learning set, we obtain an average classification accuracy equal to 0.9254 and an AUC equal to 0.7537. Among the 100 best models given by the metaheuristic instances, a large majority of them are identical. We show one of these models in Figure 5.

Following this model, a patient is accepted for surgery if he fulfills the two following conditions:

1. his ASA score is greater than 4;
2. he hasn't been subject to a cerebro-vascular accident or myocardial infarction.

The confusion matrix is given in Table 10. We note that the method is optimistic. All patients that should be accepted for surgery are correctly classified by the model. However, the model accepts a large majority of patients that should be refused for surgery. In the determination of patient acceptance or refusal for surgery, the use of a MR-Sort model seems to be less interesting than for the prediction of the ASA score.

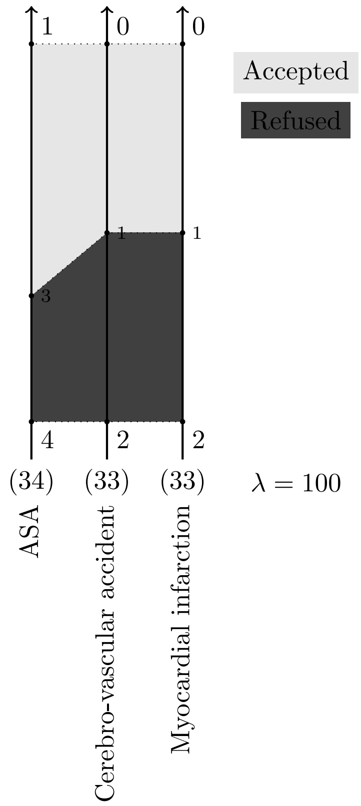


Figure 5: MR-Sort model for the prediction of the patient acceptance or refusal for surgery. Values in parenthesis below the axis are the weights of the model.

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

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

5 Discussion

Experimentations of the previous section show that MR-Sort is efficient for the prediction of the ASA score compared to other machine learning algorithms. Moreover the MR-Sort model is by nature explainable to doctors unlike the machine learning ones. However, for the prediction of acceptance or refusal of a patient, we observe that MR-Sort gives less good results than the machine learning algorithms used in [12].

In order to improve the efficiency of MR-Sort for the prediction of patient acceptance or refusal for surgery, we consider replacing the attribute ASA by the list of 16 attributes used to determine the ASA score of a patient. The results of the experiments are given in Table 11. We observe that the CA and AUC are largely improved by replacing the ASA criterion by the 16 criteria used to determine it. The model gives a CA that is up to 3 percents better when the learning set grows. But the main difference lies in the value of the AUC which increases from 75% with 3 criteria to 91% with 16 criteria. Compared to the best results obtained with the machine learning algorithms in [12], we observe a gain of more than one percent with MR-Sort.

It demonstrates that using more criteria helps to improve the quality of the model. Using the sole ASA attribute results in a descriptive loss which results in worse performances.

		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

Table 11: Prediction of Patient Acceptance/Refusal for Surgery using 18 attributes for different sizes (30%, 50%, 70%) of the data set

6 Conclusion

The development of the medical computer-aided diagnosis system is becoming today a very motivating research field. Indeed, numerous researchers working in the field of artificial intelligence are trying to suggest interpretable intelligent automatic systems ready to help doctors in their routine clinical work.

The objective of this paper is to suggest a Multiple Criteria Decision Analysis system based on the use of the MR-Sort method. This system, which consists in the detection of the ASA score to decide whether the patient is accepted or refused for surgery, is designed mainly for doctors specialized in anesthesia to ease a great part of their pre-anesthetic examination.

The results obtained with MR-Sort showed an improvement of the classification accuracy compared to machine learning algorithms. The experiments showed that the accuracy was improved by more than 5 percents for the determination of the ASA score. For the prediction of patient acceptance or refusal for surgery, we observed that MR-Sort can restore models having a better accuracy. Moreover the MR-Sort model is easier to interpret. We showed that the model can be described by a set of simple rules.

We conclude that the proposed methods can be practically used in pre-anesthetic examinations for helping doctors specialized in anesthesia to assess the risks for the patients. In our future works, we intend to enrich the database by adding patients that are not often represented in it (e.g. newborn children, patients with a pacemaker, etc.).

A Description of the dataset

The ASA score of a patient is determined on basis of multiple attributes which are all monotone:

1. Age of the patient, integer, to minimize. The younger the patient is, the better it is.

2. Diabetes, binary, to minimize. A value of '1' indicates that the patient is subject to diabetes. A value of '0' indicates the contrary. Patients with diabetes are less preferred to patients having no diabetes.

3. Hypertension, binary, to minimize. A value of '1' indicates that the patient is subject to hypertension. A value of '0' indicates the contrary. Patients with hypertension are less preferred to patients having no hypertension.

4. Respiratory failure, binary, to minimize. A value of '1' indicates that the patient is subject to respiratory failure. A value of '0' indicates the contrary. Patients with respiratory failure are less preferred to patient having no respiratory failure.

5. Hearth failure, binary, to minimize. A value of '1' indicates that the patient is subject to hearth failure. A value of '0' indicates the contrary. Patients with hearth failure are less preferred to patients having no hearth failure.

6. Hearth rate, integer, to minimize. The value corresponds to the average hearth rate of the patient in beats per minute (bpm). The lower the hearth rate of the patient, the best it is.

7. Hearth rate steadiness, binary, to maximize. A value of '1' indicates that the patient has a constant hearth rate. A value of '0' indicates the contrary. Patients having a constant hearth rate are preferred to patient having an unstable hearth rate.

8. Pacemaker, binary, to minimize. A value of '1' indicates that the patient wears a pacemaker. A value of '0' indicates the contrary. Patients having a pacemaker are less preferred to patients having no pacemaker.

9. Atrioventricular block, binary, to minimize. A value of '1' indicates that the patient has an impairment of the conduction between the atria and ventricles of the heart. Patients having an impairment are less preferred to patients having no impairment.

10. Left ventricular hypertrophy, binary, to minimize. A value of '1' indicates that the patient is subject to ventricular hypertrophy. A value of

'0' indicates the contrary. Patients having a ventricular hypertrophy are less preferred to patients having no ventricular hypertrophy.

11. Oxygen saturation, integer, to maximize. The value corresponds to the percentage of hemoglobin binding sites in the bloodstream occupied by oxygen. Patients having a high level of oxygen saturation are preferred to patients having a low level of oxygen saturation.

12. Blood glucose level, float, to minimize. The value corresponds to the level of glucose, in grams per liter (g/l), in the blood of the patient. Patients having a high level of glucose are often subject to diabetes and are less preferred to patients having a low level of glucose in their blood.

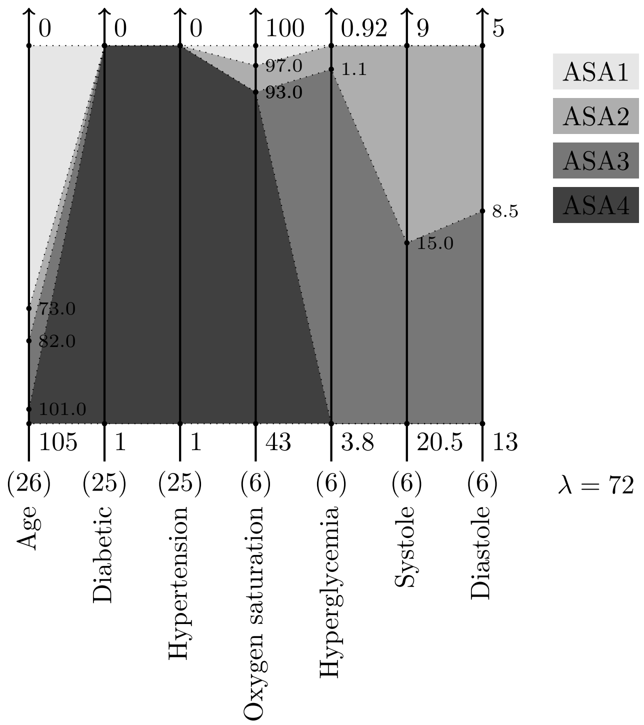
13. Systolic blood pressure, float, to minimize. The value indicates the systolic blood pressure, in centimeters of mercury. Patients having a high level of systolic blood pressure are often subject to hypertension and are therefore less preferred to patients having a low level of systolic blood pressure.

14. Diastolic blood pressure, float, to minimize. The value indicates the diastolic blood pressure, in centimeters of mercury. Patients having a high level of diastolic blood pressure are often subject to hypertension and are therefore less preferred to patients having a low level of diastolic blood pressure.

B MR-Sort models

B.1 MR-Sort model #1

B.1.1 Model parameters



B.1.2 Performances

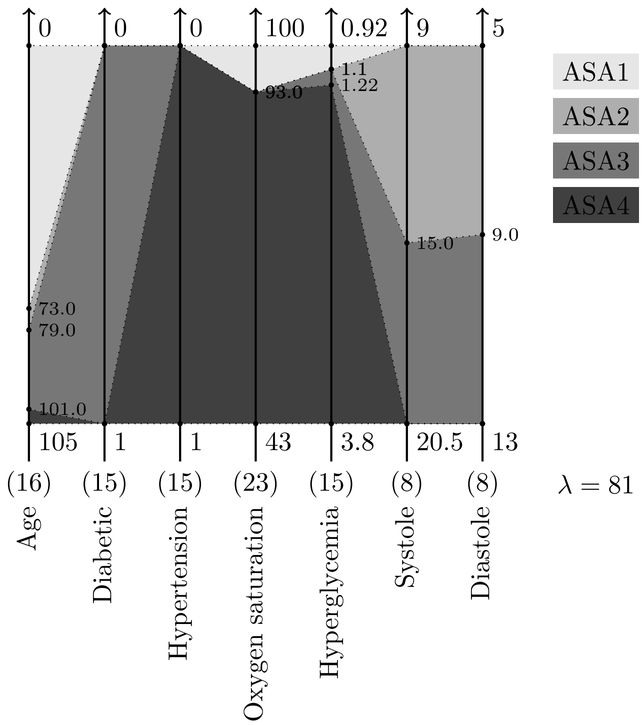
Classification accuracy: 0.9621
Area Under the Curve: 0.9843

B.1.3 Minimal winning coalitions

```
[Age, Diabetic, Oxygen saturation, Hyperglycemia, Systole, Diastole]
[Diabetic, Hypertension, Oxygen saturation, Hyperglycemia, Systole, Diastole]
[Age, Diabetic, Hypertension]
[Age, Hypertension, Oxygen saturation, Hyperglycemia, Systole, Diastole]
```

B.2 MR-Sort model #2

B.2.1 Model parameters



B.2.2 Performances

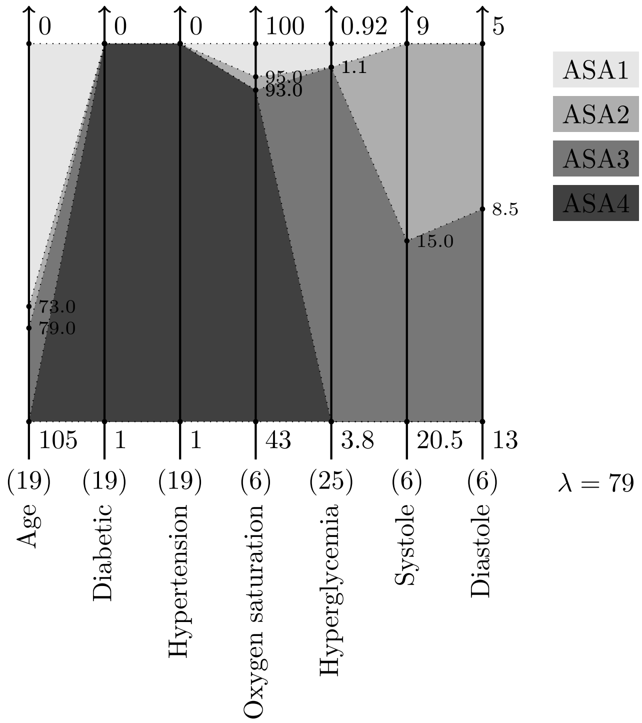
Classification accuracy: 0.9610
Area Under the Curve: 0.9847

B.2.3 Minimal winning coalitions

```
[Age, Diabetic, Hypertension, Oxygen saturation, Systole, Diastole]
[Age, Diabetic, Hypertension, Oxygen saturation, Hyperglycemia]
[Age, Diabetic, Oxygen saturation, Hyperglycemia, Systole, Diastole]
[Diabetic, Hypertension, Oxygen saturation, Hyperglycemia, Systole, Diastole]
[Age, Hypertension, Oxygen saturation, Hyperglycemia, Systole, Diastole]
```

B.3 MR-Sort model #3

B.3.1 Model parameters



B.3.2 Performances

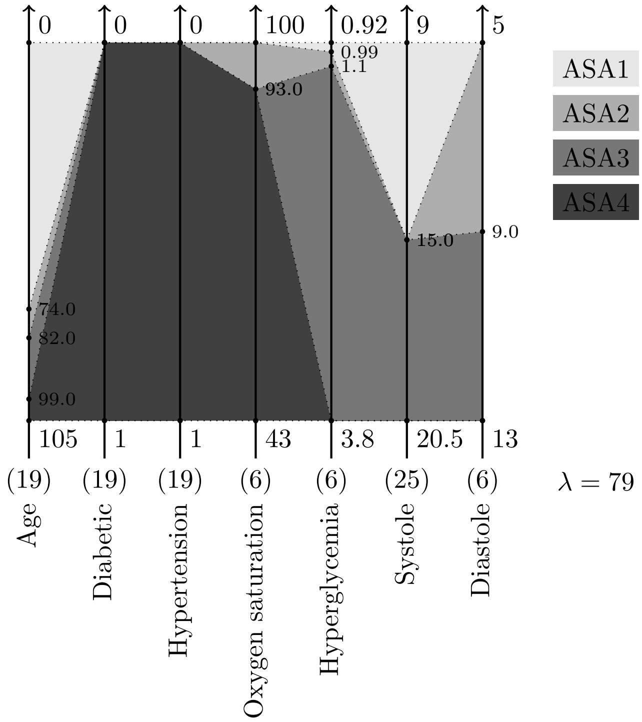
Classification accuracy: 0.9621
Area Under the Curve: 0.9870

B.3.3 Minimal winning coalitions

```
[Age, Diabetic, Hypertension, Hyperglycemia]
[Age, Diabetic, Oxygen saturation, Hyperglycemia, Systole, Diastole]
[Diabetic, Hypertension, Oxygen saturation, Hyperglycemia, Systole, Diastole]
[Age, Hypertension, Oxygen saturation, Hyperglycemia, Systole, Diastole]
```

B.4 MR-Sort model #4

B.4.1 Model parameters



B.4.2 Performances

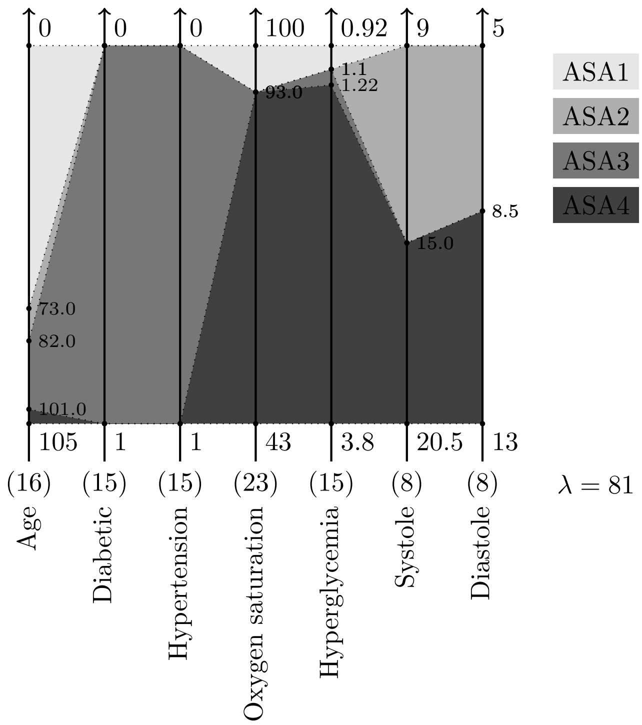
Classification accuracy: 0.9644
Area Under the Curve: 0.9882

B.4.3 Minimal winning coalitions

```
[Age, Diabetic, Hypertension, Systole]
[Age, Diabetic, Oxygen saturation, Hyperglycemia, Systole, Diastole]
[Diabetic, Hypertension, Oxygen saturation, Hyperglycemia, Systole, Diastole]
[Age, Hypertension, Oxygen saturation, Hyperglycemia, Systole, Diastole]
```

B.5 MR-Sort model #5

B.5.1 Model parameters



B.5.2 Performances

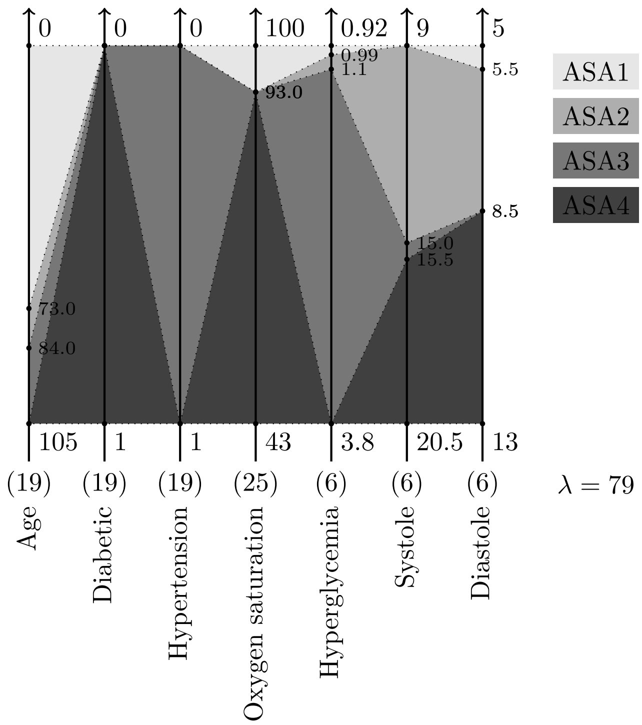
Classification accuracy: 0.9610
Area Under the Curve: 0.9878

B.5.3 Minimal winning coalitions

```
[Age, Diabetic, Hypertension, Oxygen saturation, Systole, Diastole]
[Age, Diabetic, Hypertension, Oxygen saturation, Hyperglycemia]
[Age, Diabetic, Oxygen saturation, Hyperglycemia, Systole, Diastole]
[Diabetic, Hypertension, Oxygen saturation, Hyperglycemia, Systole, Diastole]
[Age, Hypertension, Oxygen saturation, Hyperglycemia, Systole, Diastole]
```

B.6 MR-Sort model #6

B.6.1 Model parameters



B.6.2 Performances

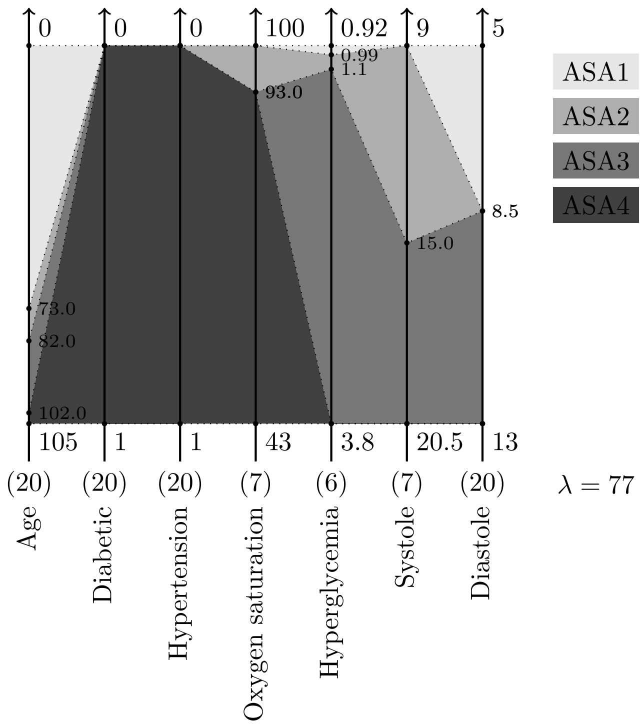
Classification accuracy: 0.9610
Area Under the Curve: 0.9855

B.6.3 Minimal winning coalitions

```
[Age, Diabetic, Hypertension, Oxygen saturation]
[Age, Hypertension, Oxygen saturation, Hyperglycemia, Systole, Diastole]
[Age, Diabetic, Oxygen saturation, Hyperglycemia, Systole, Diastole]
[Diabetic, Hypertension, Oxygen saturation, Hyperglycemia, Systole, Diastole]
```

B.7 MR-Sort model #7

B.7.1 Model parameters



B.7.2 Performances

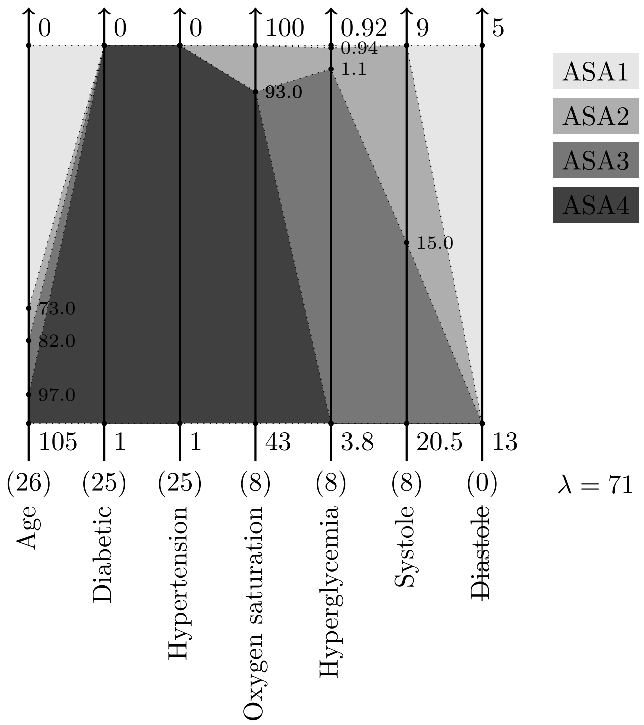
Classification accuracy: 0.9621
Area Under the Curve: 0.9884

B.7.3 Minimal winning coalitions

```
[Diabetic, Hypertension, Oxygen saturation, Hyperglycemia, Systole, Diastole]
[Age, Diabetic, Oxygen saturation, Hyperglycemia, Systole, Diastole]
[Age, Diabetic, Hypertension, Diastole]
[Age, Diabetic, Hypertension, Oxygen saturation, Hyperglycemia, Systole]
[Age, Hypertension, Oxygen saturation, Hyperglycemia, Systole, Diastole]
```

B.8 MR-Sort model #8

B.8.1 Model parameters



B.8.2 Performances

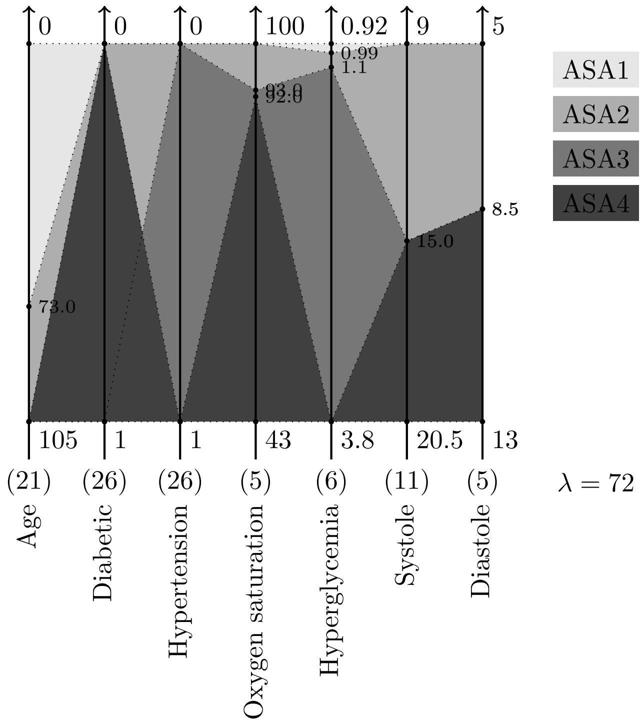
Classification accuracy: 0.9477
Area Under the Curve: 0.9791

B.8.3 Minimal winning coalitions

```
[Age, Hypertension, Oxygen saturation, Hyperglycemia, Systole]
[Diabetic, Hypertension, Oxygen saturation, Hyperglycemia, Systole]
[Age, Diabetic, Hypertension]
[Age, Diabetic, Oxygen saturation, Hyperglycemia, Systole]
```

B.9 MR-Sort model #9

B.9.1 Model parameters



B.9.2 Performances

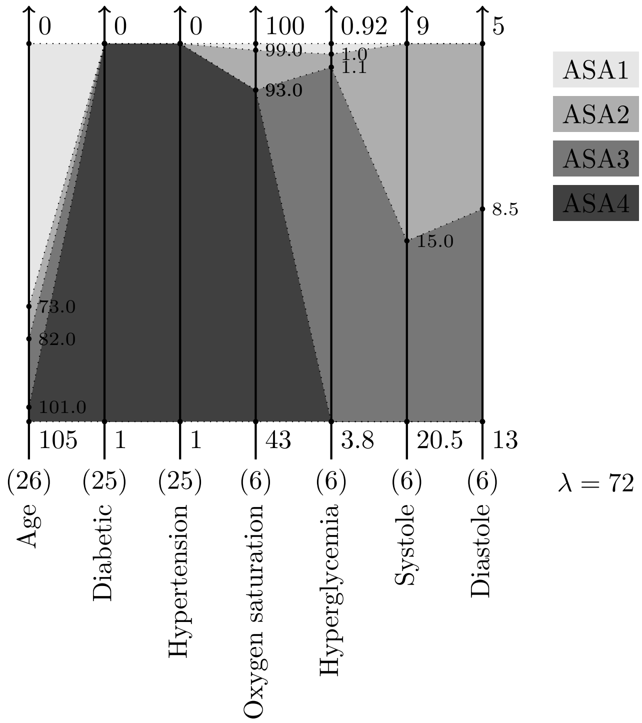
Classification accuracy: 0.9465
Area Under the Curve: 0.9671

B.9.3 Minimal winning coalitions

```
[Diabetic, Hypertension, Oxygen saturation, Hyperglycemia, Systole]
[Age, Diabetic, Hypertension]
[Diabetic, Hypertension, Oxygen saturation, Systole, Diastole]
[Age, Diabetic, Oxygen saturation, Hyperglycemia, Systole, Diastole]
[Diabetic, Hypertension, Hyperglycemia, Systole, Diastole]
[Age, Hypertension, Oxygen saturation, Hyperglycemia, Systole, Diastole]
```

B.10 MR-Sort model #10

B.10.1 Model parameters



B.10.2 Performances

Classification accuracy: 0.9621
Area Under the Curve: 0.9851

B.10.3 Minimal winning coalitions

```
[Age, Diabetic, Oxygen saturation, Hyperglycemia, Systole, Diastole]
[Diabetic, Hypertension, Oxygen saturation, Hyperglycemia, Systole, Diastole]
[Age, Diabetic, Hypertension]
[Age, Hypertension, Oxygen saturation, Hyperglycemia, Systole, Diastole]
```

References

[1] D. Bouyssou and T. Marchant. An axiomatic approach to noncompensatory sorting methods in MCDM, I: The case of two categories. *European Journal of Operational Research*, 178(1):217–245, 2007.

[2] D. Bouyssou and T. Marchant. An axiomatic approach to noncompensatory sorting methods in MCDM, II: More than two categories. *European Journal of Operational Research*, 178(1):246–276, 2007.

[3] O. Cailloux, P. Meyer, and V. Mousseau. Eliciting ELECTRE TRI category limits for a group of decision makers. *European Journal of Operational Research*, 223(1):133–140, 2012.

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

[5] R. L. Chatburn and F. P. Primiano. Decision analysis for large capital purchases: how to buy a ventilator. *Respiratory Care*, 46(10):1038–53, 2001.

[6] A. Donati, M. Ruzzi1, E. Adrario1, P. Pelaia1, F. Coluzzi, V. Gabbanelli, and P. Pietropaoli. A new and feasible model for predicting operative risk. *British Journal of Anaesthesia*, 93(3):393–399, 2004.

[7] J. R. Figueira, J. A. Dias, S. Matias, B. Roy, M.J. Carvalho, and C. E. Plancha. Electre Tri-C, a multiple criteria decision aiding sorting model applied to assisted reproduction. *International Journal of Medical Informatics*, 80(4):262–273, 2011.

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

[9] C. S. Goodman and R. Ahn. Methodological approaches of health technology assessment. *International Journal of Medical Informatics*, 56(1-3):97–105, 1999.

[10] J. Hussmann and R.C. Russel. *Memorix: Surgery.* Chapman & Hall, 1997.

[11] S. Karpagavalli, K. Jamuna, and M. Vijaya. Machine learning approach for pre-operative anaesthetic risk prediction. *International Journal of Recent Trends in Engineering*, 1(2), 2009.

[12] M. E. A. Lazouni, M. El Habib Daho, N. Settouti, M.A. Chikh, and S. Mahmoudi. Machine learning tool for automatic asa detection. In A. Amine, O. Aït Mohamed, and L. Bellatreche, editors, *Modeling Approaches and Algorithms for Advanced Computer Applications*, volume 488 of *Studies in Computational Intelligence*, pages 9–16. Springer, 2013.

[13] M.A. Lazouni, M.A. Chikh, and S. Mahmoudi. A new computer aided diagnosis system for pre-anesthesia consultation. *Journal of Medical Imaging and Health Informatics*, 3(4):471–479, 2013.

[14] A. Leroy, V. Mousseau, and M. Pirlot. Learning the parameters of a multiple criteria sorting method. In R. Brafman, F. Roberts, and A. Tsoukiàs, editors, *Algorithmic Decision Theory*, volume 6992 of *Lecture Notes in Artificial Intelligence*, pages 219–233. Springer, 2011.

[15] M. J. Liberatore and R. L. Nydick. The analytic hierarchy process in medical and health care decision making: A literature review. *European Journal of Operational Research*, 189(1):194–207, 2008.

[16] P. Lutz. The medical algorithms project, ch31. anaesthesiology, section: Preoperative patient classification and preparation, 2008.

[17] O. Sobrie, V. Mousseau, and M. Pirlot. Learning a majority rule model from large sets of assignment examples. In P. Perny, M. Pirlot, and A. Tsoukiás, editors, *Algorithmic Decision Theory*, Lecture Notes in Artificial Intelligence, pages 336–350, Paris, France, 2013. Springer.

[18] W. Waegeman, B. De Baets, and L. Boullart. Roc analysis in ordinal regression learning. *Pattern Recognition Letters*, 29(1):1–9, 2008.

[19] D. West, P. Mangiameli, R. Rampal, and V. West. Ensemble strategies for a medical diagnostic decision support system: A breast cancer diagnosis application. *European Journal of Operational Research*, 162(2):532–551, 2005.