



Conference on ENTERprise Information Systems / International Conference on Project  
MANagement / Conference on Health and Social Care Information Systems and Technologies,  
CENTERIS / ProjMAN / HCist 2015 October 7-9, 2015

## Intelligent Decision Support to predict patient Barotrauma risk in Intensive Care Units

Sérgio Oliveira<sup>a</sup>, Filipe Portela<sup>a\*</sup>, Manuel Filipe Santos<sup>a</sup>, José Machado<sup>a</sup>, António  
Abelha<sup>a</sup>, Álvaro Silva<sup>b</sup>, Fernando Rua<sup>b</sup>

<sup>a</sup>Algoritmi Research Center, University of Minho, Portugal

<sup>b</sup>Intensive Care Unit, Centro Hospitalar do Porto, Porto, Portugal

---

### Abstract

The occurrence of Barotrauma is identified as a major concern for health professionals, since it can be fatal for patients. In order to support the decision process and to predict the risk of occurring barotrauma Data Mining models were induced. Based on this principle, the present study addresses the Data Mining process aiming to provide hourly probability of a patient has Barotrauma. The process of discovering implicit knowledge in data collected from Intensive Care Units patients was achieved through the standard process Cross Industry Standard Process for Data Mining. With the goal of making predictions according to the classification approach they several DM techniques were selected: Decision Trees, Naive Bayes and Support Vector Machine. The study was focused on identifying the validity and viability to predict a composite variable. To predict the Barotrauma two classes were created: "risk" and "no risk". Such target come from combining two variables: Plateau Pressure and  $PCO_2$ . The best models presented a sensitivity between 96.19% and 100%. In terms of accuracy the values varied between 87.5% and 100%. This study and the achieved results demonstrated the feasibility of predicting the risk of a patient having Barotrauma by presenting the probability associated.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of SciKA - Association for Promotion and Dissemination of Scientific Knowledge

**Keywords:** INTCare; Barotrauma; Intensive Care; Data Mining; Mechanical Ventilation; Decision Support; Probability; Patient-centered.

---

---

\* Corresponding author. Tel.: +351-253 510 319.  
E-mail address: [cfp@dsi.uminho.pt](mailto:cfp@dsi.uminho.pt)

## 1. Introduction

It is commonly recognized that the data provided by organization can be a source of implicit knowledge. Since technological advances have provided mechanisms to acquire data from several processes / business areas. The organizations have been able to extract knowledge using their database tools in order to provide it with better working practices or even assist in restructuring processes [1]. Hospitals also follow the principle that their database can be important knowledge source in several areas. Forecasting or diagnosing possible complications of their patients are examples of what hospitals can achieve from their databases. The use of data mining process (DM) has been adopted in hospitals [2] to predict patients conditions. Health professionals are able to identify more effective treatments, and more appropriate practices providing to the patient a better healthcare with a focus on their health problems. The DM provides not only the methodology but also the technology which is able to transform datasets into useful knowledge capable of assisting decision-making process [3]. Taking into account the potential presented by the DM process, this study follows the respective process, in order to identify the feasibility of conducting hourly forecasts based on a target. The target (binary variable) is set as a combination of two variables values: Plateau Pressure and  $PCO_2$ . Being very difficult to have totally sure if a patient had or not barotrauma, the risk of a patient have barotrauma was identified based in the values collected. This new variable helped to identify if a patient had "risk" or not "no risk" of having Barotrauma. All the work was performed using real data provided by Centro Hospitalar do Porto. The achieved results (sensitivity upper than 96% and accuracy upper than 85%) help to hourly predict the probability of a patient has barotrauma. This results helps to early detect a possible barotrauma occurrence and prevent their occurrence in a patient. By focusing the work in the patient it is possible provide best care in the patient best interest.

This paper consists of five sections. The first section addresses the problem presented. In the following chapter are addressed some aspects related to the issue studied, as also the concepts and the technology used. Then in chapter three it is presented the work made following CRISP-DM methodology. The fourth chapter presents and discusses the most important aspects identified throughout the development process component of practice. Finally, the next chapter presents the most relevant conclusions.

## 2. Background

### 2.1. Barotrauma and related variables

Barotrauma terminology is used to identify patients who have complications arising from mechanical ventilation. This type of patients are essentially admitted in Intensive Care Units (ICU). The occurrence of Barotrauma is also associated to respiratory diseases. It is most common in patients who have severe pulmonary diseases, especially in patients with Acute Respiratory Distress Syndrome (ARDS) or with chronic lung disease (e.g. pneumonia obstructive pulmonary) [4]. Although mechanical ventilation is able of supporting and assist the survival of patients, it also can damage patient lungs. Barotrauma results from a high pressures of lung volumes. Its occurrence is often responsible for morbidity and mortality. In patients with acute respiratory distress syndrome (ARDS) and incidence of Pneumothorax or Barotrauma, the mortality varies between 0% and 76% [5]. Several researchers argue that there is a relationship between PEEP and the occurrence of Barotrauma, however there are other researchers who do not defend this point [4]. In this case the intensivist knowledge should be considered as the most important. The occurrence of Barotrauma may vary with the severity of lung disease, however the mechanical ventilation pattern also could explain this occurrence. This happens especially when the Plateau Pressure is higher than 35  $cmH_2O$  [5]. Plateau pressure or static elastic recoil pressure of the respiratory system is useful for carrying out the elastic recoil of the respiratory system of patients with respiratory failure. Plateau pressure provides relevant diagnostic information and it is important to maintain its value  $\leq 30$   $cmH_2O$  for lung protection of patients [6].

In a study [7] was found that peak values of inspiratory pressure, positive pressure levels at the end of expiration, respiratory rate and tidal volume were significantly higher in patients who developed Barotrauma. In the opposite are the patients who did not develop Barotrauma and had lower values of the respective variables. In fact high values reflect the high incidence of Barotrauma in patients with ARDS [7].

## 2.2. Related Work

Due to the importance of detecting Barotrauma in the patients, being able to make probability predictions, is the natural next step for the research. There is already a study where it was demonstrated the feasibility of predicting correlated variables using only variables collected from the ventilation system [8]. In this previous study Barotrauma was predicted by analyzing if a patient had a Plateau pressure higher or equal to 30 cmH<sub>2</sub>O. The study demonstrated the viability of predicting Barotrauma using Plateau pressure variable. Taking as reference the results of the forecasts made for the two classes created < 30 cmH<sub>2</sub>O and >= 30 cmH<sub>2</sub>O, the accuracy was between 95.52% and 98.71%. Besides the good results was also identified a set of variables that it is correlated with the generated classes: Dynamic Compliance, Means Airways Pressure and Peak Pressure. Based on the results achieved, the set of variables correlated with the target, the approach used in the early study revealed the need of continuing the study in order to create a better target able to identify the feasibility of conducting hourly forecasts to Barotrauma risk. The study performed and the results achieved gave origin to the present work. This new work added a new variable (PCO<sub>2</sub>) to the target and uses more variables (e.g. laboratory results) as input. These new variables were identified by the intensivists.

## 2.3. INTCare

This study was developed under the research project called INTCare. INTCare gave origin to an Intelligent Decision Support System (IDSS) [9]. The second phase of the project is focused in predicting ventilation and respiratory diseases as is barotrauma. The respective system is in operation at ICU of Hospital Santo António – Centro Hospitalar do Porto (CHP). INTCare is able to monitoring the conditions in which the patient is and is also able to make predictions related to the patient's condition, patient's organ failures, readmissions, length of stay and suggests procedures of therapies and treatments.

## 2.4. Data Mining

Data Mining (DM) is presented as a process that uses artificial intelligence techniques, the statistics and mathematics to extract information and useful knowledge from large volumes of data. The discovery of patterns in the data may be in the form of business rules, affinities, correlations or terms of prediction models [10].

In this study the application of DM process was achieved by the statistical system R. R is a programming language oriented to solve problems in statistical environments [11]. To the development of the practical component of the study several packages geared to the DM process were used. The e1071 library [12] was used to completed the implementation of the techniques used: Support Vector Machine (SVM), Decision Trees (DT) and Naive Bayes (NB). The DM models evaluation was carried out using the library rminer [13].

## 3. Knowledge Discovering Process

Cross Industry Standard Process for Data Mining (CRISP-DM) [14] methodology assists the process of knowledge discovery, becoming easier to develop solutions that best suits the goals. As such, the need for a methodology to study the inherent senses is obvious and CRISP-DM methodology was followed to carry out the present study.

### 3.1. Business Understanding

This first phase of the methodology was mainly focused on understanding the project goals and requirements. The aim of this work is to provide useful and timely information for health-care workers to make decisions supported by relevant information. In the previous study [8] the Plateau pressure was used as a relevant factor for predict the occurrence of Barotrauma. After a meeting with the physicians it was identified another variable that can be related with the presence of barotrauma. The present study used a new variable PCO<sub>2</sub>, to create the target variable. At same time this study is totally different from the previous because the data frame used (contents and

format) is not the same. Now the goal is to predict the probability by hour. In this context it was defined the data mining goal: identify the feasibility of predict barotrauma probability using as target Plateau Pressure and  $PCO_2$ .

The data used for the study are from the ICU of CHP. Data Mining models were induced by a set of variables collected by ventilators, laboratory and Electronic Health Record. The target was defined according to the values of the Plateau and  $PCO_2$  pressure. The class "risk" corresponds to Plateau pressure  $\geq 30$   $cmH_2O$  and the  $PCO_2 > 45$  mmHg. The target is equal to "no risk" when the Plateau pressure value is  $< 30$   $cmH_2O$  and the  $PCO_2 \leq 45$  mmHg.

It is important to highlight that the main goal is not to predict if a patient will have or not barotrauma, but the probability [0-100%] of it occurring. It happens because in the dataset this fact is not obvious (there is not a single variable to easily identify barotrauma). For that reason the target is a combination of two variables that represents if the patient is or not in risk. Finally being the goal predicting the risk, both variables also were used as input in order to improve the final results when combined with the other patient variables. Case the purpose of the study was to predict if the patient will have barotrauma (yes or no), these variables are discarded from the input dataset.

### 3.2. Data Understanding

The implementation of data mining models is dependent from a dataset. As such, the sample used to make this work covers the period from 01.09.2014 to 31.10.2014, about two months were used. This sample was composed by 3362 records. The records were organized by hour and patient, i.e., for each hour a new data line was created:

- CDYN – (F\_1): Dynamic compliance in mL/  $cmH_2O$ ;
- CSTAT – (F\_2): Static compliance from inspiratory pause measured in mL/  $cmH_2O$ ;
- FIO2 – (F\_3): Fraction of inspired oxygen (%);
- Flow – (F\_4): Peak flow setting in liters per minute;
- RR – (F\_5): Respiratory rate setting in berths per minute;
- PEEP – (F\_6): Positive End-Expiratory Pressure in  $cmH_2O$ ;
- PMVA – (F\_7): Mean airway pressure in  $cmH_2O$ ;
- Plateau Pressure – (F\_8): End inspiratory in  $cmH_2O$ ;
- Peak Pressure – (F\_9): Maximum circuit pressure in  $cmH_2O$ ;
- RDYN – (F\_10): Dynamic resistance in  $cmH_2O$ ;
- RSTAT – (F\_11): Static resistance from inspiratory pause measured in  $cmH_2O/L/s$ ;
- Volume EXP – (F\_12): Exhaled tidal volume in liters;
- Volume INS – (F\_13): Tidal volume settings in liters;
- Volume Minute – (F\_14): Exhaled minute volume liters;
- Age – (F\_15): Patient Age;
- Gender – (F\_16): The sex of the patient;
- $PCO_2$  – (F\_17):  $CO_2$  partial pressure;
- Hours ventilation – (F\_18): The number of hours the patient was ventilated until the model was induced;
- Days of hospitalization – (F\_19): Number of days the patient was hospitalized until the model was induced.

Table 1 presents a statistical analysis for each field collected from the ventilators. For each one of the numeric values it was analyzed the minimum (MIN) and maximum (MAX) value collected, their average (MEAN), standard deviation (stDev) and coefficient of variation (CFV).

Table 1. Distribution of variables.

	MAX	MIN	MEAN	stDEV	CFV		MAX	MIN	MEAN	stDEV	CFV
F_1	200	0	39.60	27.12	68.49	F_10	100	0	14.36	7.53	52.43
F_2	126	0	23.31	24.76	106.25	F_11	51	0	9.67	9.46	97.73
F_3	100	22	55.59	14.52	26.12	F_12	1.98	0	0.53	0.16	31.86
F_4	80	0	36.45	18.84	51.70	F_13	0.71	0	0.39	0.19	49.36
F_5	29	0	11.90	8.06	67.72	F_14	23.1	0	9.57	2.41	25.23
F_6	10	3	5.65	1.50	26.57	F_19	43	1	10.87	9.36	86.07
F_7	34	0	11.96	2.82	23.64	F_18	72	0	32.04	21.32	66.54
F_8	85	0	22.29	7.61	34.13	F_16	88	33	64.21	13.57	21.14

F_9	85	0	29.38	9.51	32.39	F_17	62.70	15.60	39.34	6.00	15.26
-----	----	---	-------	------	-------	------	-------	-------	-------	------	-------

The sample is also composed by two qualitative variables: gender (F\_16) and the target, due to its properties could not be performed the distributions shown in Table 1. About 98.36% of the records have the target with the value of "no risk" and 1.64% with a value of "risk". In order to maintain the dataset integrity and although the number of risk cases be very lower it was not applied any sampling method (e.g. oversampling). In medicine this technique normally has a strong effect in the data significance. Approximately 61.45% of the cases correspond to male patients and 33.65% of cases are owned by female patients.

### 3.3. Data Preparation

Once the data mining approach of this study is the classification for forecasting, there was the need to make some changes in the data. The most significant change was made according to clinical knowledge to determine the rules that would give rise to the target. It was necessary to create a variable that could only have two different values, and this would have to be a qualitative variable. Their values may be "risk" or "no risk", where "risk" represents an instance in which the patient has higher values of Plateau Pressure and  $PCO_2$ . When the value of the target is "no risk" corresponds to an instance where the patient does have normal values of Plateau Pressure and  $PCO_2$ . At this stage performing tasks were performed. Based on the clinical knowledge, some records with values outside of the default values were eliminated in particular bad values monitored by the ventilator.

### 3.4. Modelling

At this stage it was important to select the Data Mining techniques that will best suit rating approach. The techniques selected to solve this classification problem were: Support Vector Machine (SVM), Decision Tree (DT) and Naive Bayes (NB). The choice of these techniques was based on two characteristics: interpretability and efficiency. The SVM reaches the second characteristic, but the DTs and NBs meet the two characteristics. One important aspect in modelling stage is the implementation of test mechanisms. The models was induced using the sampling method 10 Folds Cross Validation. Literature refers goods results of using of the 10 Folds Cross Validation [15].

The selected techniques underwent tuning function. The respective function is integrated in the library e1071. This generic function tunes hyperparameters of statistical methods using a grid search over supplied parameter ranges [12]. Two Kernels for SVM technique were used: Linear Kernel and Radial-Basic Function. The use of Kernels implies different parameterization, because each parameter has hyperparameters that differ from those used Kernels. Both for the Linear Kernel and Radial Basic Function Kernel, a range of values was defined for the parameter  $C$ . Their values vary a power of  $2^{(1..4)} = [2, \dots, 16]$ , where  $C > 0$ . The possible values for this range have some parameter introduces some flexibility in the separation of categories in order to control the trade-off between training errors and the stiffness of the edges [16] the hyperparameter Gamma ( $\gamma$ ) is also defined according to a second output  $2^{(-1,0,1)} = [0.5, 1, 2]$ , wherein the gamma curve determines the decision boundary of [17]. The variable  $\gamma$  was only configured in Radial Basic Function Kernel.

The DT technique was implemented according to the CART algorithm. Two rules of decomposition were used, Information Gain (IG) and the Gini Index (GI). In the implementation of NBs, it was not performed parameterization, however it was necessary to use the tuning function to identify the sampling method. The developed models can be represented by the following expression:

$$M_n = \{A_a; F_b; TV_c; TDM_D; SM_e\}$$

The model  $M_n$  belongs to an approach (A) and it is composed by Fields (F), a type of variable (TV) a DM technique (TDM) and a sampling method (SM):

- $A_a = \{\text{Classification}_1\}$
- $F_b = \{F_{1_1}, F_{2_2}, F_{3_3}, F_{4_4}, F_{5_5}, F_{6_6}, F_{7_7}, F_{8_8}, F_{9_9}, F_{10_{10}}, F_{11_{11}}, F_{12_{12}}, F_{13_{13}}, F_{14_{14}}, F_{15_{15}}, F_{16_{16}}, F_{17_{17}}, F_{18_{18}}, F_{19_{19}}, F_{20_{20}}\}$

- $TV_c = \{Qualitative\ variables\ nominal_1, Quantitative\ variables\ continuous_2\}$
- $TDM_d = \{SVM\_L_1, SVM\_R_2, DT\_IG_3, DT\_GI_4, NB_5\}$
- $SM_e = \{10 - Folds\ Cross\ Validation_1\}$

With descriptive notation is possible to represent an example of DM model:

$$M_1 = \{A_1; F_{1\{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19\}}; TV_1; TDM_2; SM_1\}$$

The  $M_1$  is a model that follows the classification approach, input attributes have two types of qualitative and quantitative variables, the DM technique used was the SVM with Radial-Basic Function Kernel and the sampling method used was 10-Folds Cross Validation.

### 3.5. Evaluation

The evaluation of the models was performed according to three rating metrics: accuracy, specificity and sensitivity. The separation method of datasets for training and testing was carried out by 10-Folds Cross Validation. In the implementation of the respective dividing procedures, ten executions were performed for each one of them. A hundred tests experiences was accomplished in each model. Table 2 has the best models ( $BM_m$ ) generated by each DM techniques. The valuation metrics used to conduct the evaluation of the models were: accuracy (ACU), sensitivity (SEN) and specificity (SPE).

Table 2. Valuation models.

Model	Fields	TDM	ACU	SPE	SEN	Model	Fields	TDM	ACU	SPE	SEN
$BM_1$	$F_{TV_1\{1,2,3,4,5,6,7,8\}}\{9,10,11,12,13,14,15,17,18,19\}$ + $F_{TV_2\{16\}}$	$TDM_3$	100%	100%	100%	$BMP_1$	$F_{TV_1\{1,2,3,4,5,6,7,8\}}\{9,10,11,12,13,14,15,17,18\}$ + $F_{TV_2\{16\}}$	$TDM_3$	100%	100%	100%
$BM_2$	$F_{TV_1\{1,2,3,4,5,6,7,8\}}\{9,10,11,12,13,14,15,17,18\}$ + $F_{TV_2\{16\}}$	$TDM_2$	99.76%	99.82%	96.18%	$BMP_2$	$F_{TV_1\{1,2,3,4,5,6,7,8\}}\{9,10,11,12,13,14,15,17,18\}$ + $F_{TV_2\{16\}}$	$TDM_1$	87.5%	86.49%	100%
$BM_3$	$F_{TV_1\{1,2,3,4,5,6,7,8\}}\{9,10,11,12,13,14,15,17,18,19\}$ + $F_{TV_2\{16\}}$	$TDM_5$	96.19%	99.94%	29.61%	$BMP_3$	$F_{TV_1\{1,2,3,4,5,6,7,8\}}\{9,10,11,12,13,14,15,17,18\}$ + $F_{TV_2\{16\}}$	$TDM_5$	95%	94.12%	100%

In the present study the negative value corresponds to "risk and it is the event that indicate great potential of a patient has Barotrauma. The DM technique which presented better results was DT, where the results were 100% for all the used metrics. This happens due to the fact that the target is a composite variable and DT create a rule using only those variables. The DT can then identify the limits of the input variables that cause the target. The SVM and NB also had quite satisfactory results as the accuracy and specificity. However the application of NB results presented low sensitivity in  $BM_3$ . It is then shown that in this case the NB cannot predict the negative values. Using  $BM_3$  is possible concluded that the model is not able to properly predict patients who have Barotrauma indicator as "risk".

After talking with health professionals, it became evident that there was great concern in performing individual forecasts by patient. So some forecast models for several patients not present in the initial dataset were created.

Table 2 also show the best prediction models for a chosen patient  $BMP_p$ . This patient had 40 records. These records correspond to the number of hours that the patient was with mechanical ventilation. From the 40 records 8 of them were related with the risk of the patient has Barotrauma.

The results of the presented models were very satisfactory. In general all models were excellent to predict the class "risk". In both cases the sensitivity was equal to 100%, however only  $BMP_1$  were able to predict both classes.

### 3.5. Deployment

Since this is the last phase of the methodology is intended the implementation of the models generated in a decision support system. As such, data acquisition and transformation and models inducing and assessment are running in real-time. However the models only were included in INTCare system if they are an asset to support medical decision.

## 4. Discussion

Several prediction models were induced and the results were quite satisfactory. The best models presented results with accuracy higher than 96%. Also the displayed values corresponding to sensitivity and specificity were satisfactory. A relevant fact which show that the results are totally dependent of the test dataset is the case of NB. In this case the best model did not presented a satisfactory sensitivity (29.61%) but when it was induced using a different patient it was the best model (100%). This means that every time a new patient arrives it is important to run the models again because probably there is a better model to predict their probability.

In order to help to choose the best model a threshold that the models need to achieve before be chosen was defined: Accuracy and specificity  $\geq 85\%$  and sensitivity  $\geq 90\%$ . Additionally and as was proved DT presented always 100%. Being our goal to predict the probability of the patient be in risk, using DT the probability will be always 0% or 100%. In this case this technique is not useful for our study therefore it was discarded.

After execute the best model the probability of a patient be in risk of having a barotrauma will be higher and totally sure upper than 50%. For example if the prediction is 0, the probability of the patient be in risk is from 0 to 50% in the case of the prediction be 1, the probability of the patient be in risk is situated between 51% and 100%.

Table 3 shows the resulting forecast to the best model chosen:  $BMP_3$ .

Table 3. Confusion Matrix.

Target	Predictive	
	"no risk"	"risk"
"no risk"	32	2
"risk"	0	6

Confusion matrix (table 3) can provide certain useful information. The number of false negatives is 2/32 and most important is the number of risk patient classified correctly: 100%. It can be argued that the model is able to predict all the risk cases. Although the DT not be used in the final assessment, it proven be very useful. It was possible to identify the variables with the greatest influence (F\_17, F\_8, F\_13 and F\_9) to the risk class. The Peak Pressure (F\_9) once again demonstrates its importance in the implementation of forecasting models related to ventilation data.

The following graph (Fig. 1.) is presenting the probability of the patient used in the evaluation phase be in risk of occurring Barotrauma using  $BMP_3$ . In this figure were considered only the risk cases (six cases).

The blue line corresponds to the predict point "risk", the red line represents the probability of occurring "risk" and the black line represents the probabilities of occurring "no risk". There is only one case where the probability of occurrence of Barotrauma is low – the third hour. For the remaining hours this patient is in risk and has a high probability of having Barotrauma (probability very close to 100%).

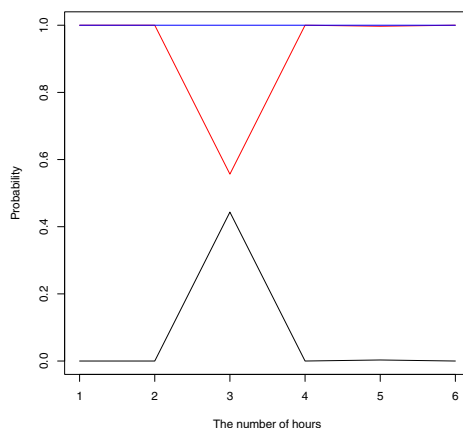


Fig. 1. Hour probability to occur "risk".

## 5. Conclusion

After assessing the results presented by DM models was possible to conclude that it is feasible to carry out hourly forecasts for each patient representing if the patient is in risk or not of suffering Barotrauma. The achieved results by the prediction models following the classification approach were quite satisfactory as well for a particular patient.

The threshold created allows a better understanding about which is the best model. Being the main goal to predict the risk (probability), the models should be sensitive, however and in order to avoid the high number of false negatives the other two metrics also were considered. These models are viable and are able to predict the probability of patients who are at risk of having Barotrauma. The achieved results gave a very good confidence to conclude this INTCare goal.

In the future it will be performed the last phase of this work (predicting barotrauma). The data will be processed in real-time and the models will be induced automatically. Additionally the inclusion of other variables that can be calculated for the next hours (e.g. accumulated critical events and ratios) will be considered in order to be possible predict barotrauma probability for the next 24 hours.

## Acknowledgements

This work has been supported by FCT - Fundação para a Ciência e Tecnologia within the Project Scope UID/CEC/00319/2013 and PTDC/EEI-SII/1302/2012 (INTCare II)

## References

- [1] M. Santos and C. Azevedo, *Data Mining Descoberta de conhecimento em base de dados*, 1ª Edição. FCA, 2005.
- [2] M. Santos, M. Boa, F. Portela, Á. Silva, and F. Rua, "Real-time prediction of organ failure and outcome in intensive medicine," presented at the Information Systems and Technologies (CISTI), 2010 5th Iberian Conference on, 2010.
- [3] H. Koh and G. Tan, "Data mining applications in healthcare," *J. Healthc. Inf. Manag. JHIM*, vol. 19, no. 2, pp. 64–72, 2005.
- [4] A. Anzueto, F. Frutos-Vivar, A. Esteban, I. Alía, L. Brochard, T. Stewart, S. Benito, M. J. Tobin, J. Elizalde, F. Palizas, C. M. David, J. Pimentel, M. González, L. Soto, G. D'Empaire, and P. Pelosi, "Incidence, risk factors and outcome of barotrauma in mechanically ventilated patients," *Intensive Care Med.*, vol. 30, no. 4, pp. 612–619, Apr. 2004.
- [5] M. Boussarsar, G. Thierry, S. Jaber, F. Roudot-Thoraval, F. Lemaire, and L. Brochard, "Relationship between ventilatory settings and barotrauma in the acute respiratory distress syndrome," *Intensive Care Med.*, vol. 28, no. 4, pp. 406–413, Apr. 2002.
- [6] N. Al-Rawas, M. J. Banner, N. R. Euliano, C. G. Tams, J. Brown, A. D. Martin, and A. Gabrielli, "Expiratory time constant for determinations of plateau pressure, respiratory system compliance, and total resistance," *Crit. Care Lond. Engl.*, vol. 17, no. 1, p. R23, 2013.
- [7] R. B. Gammon, M. S. Shin, and S. E. Buchalter, "Pulmonary barotrauma in mechanical ventilation. Patterns and risk factors," *Chest*, vol. 102, no. 2, pp. 568–572, Aug. 1992.



- [8] S. Oliveira, F. Portela, M. F. Santos, J. Machado, A. Abelha, Á. Silva, and F. Rua, “Predicting Plateau Pressure in Intensive Medicine for Ventilated Patients,” in *New Contributions in Information Systems and Technologies*, A. Rocha, A. M. Correia, S. Costanzo, and L. P. Reis, Eds. Springer International Publishing, 2015, pp. 179–188.
- [9] F. Portela, M. F. Santos, J. Machado, A. Abelha, Á. Silva, and F. Rua, “Pervasive and Intelligent Decision Support in Intensive Medicine – The Complete Picture,” in *Information Technology in Bio- and Medical Informatics*, M. Bursa, S. Khuri, and M. E. Renda, Eds. Springer International Publishing, 2014, pp. 87–102.
- [10] E. Turban, R. Sharda, and D. Delen, *Decision Support and Business Intelligence Systems*, 9ª Edição. Prentice Hall, 2011.
- [11] L. Torgo, *Data Mining with R: Learning with Case Studies*. CRC Press - Taylor & Francis Group, 2011.
- [12] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, “Misc Functions of the Department of Statistics (e1071).” 2012.
- [13] P. Cortez, “Simpler use of data mining methods (e.g. NN and SVM) in classification and regression.” 2013.
- [14] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, “CRISP-DM 1.0 Step-by-step data mining guide.” SPSS inc, 2000.
- [15] I. Witten, E. Frank, and M. Hall, *Data Mining Practical Machine Learning Tools and Techniques*, 3ª Edição. Morgan Kaufmann, 2011.
- [16] M. Kantardzic, *Data Mining Concepts, Models, Methods, and Algorithms*, 2ª Edição. Wiley - IEEE Press, 2011.
- [17] A. Ben-Hur and J. Weston, “A User’s Guide to Support Vector Machines,” in *Data Mining Techniques for the Life Sciences*, O. Carugo and F. Eisenhaber, Eds. Humana Press, 2010.