

# Real-Time models to predict the use of vasopressors in monitored patients

André Braga<sup>1</sup>, Filipe Portela<sup>1,2\*</sup>, Manuel Filipe Santos<sup>1</sup>, António Abelha<sup>1</sup>, José Machado<sup>1</sup>,  
Álvaro Silva<sup>3</sup> and Fernando Rua<sup>3</sup>

<sup>1</sup>Algoritmi Research Centre, University of Minho, Portugal

<sup>2</sup>ESEIG, Porto Polytechnic, Vila do Conde, Portugal

<sup>3</sup>Intensive Care Unit, Centro Hospitalar do Porto, Portugal

**Abstract.** The needs of reducing human error has been growing in every field of study, and medicine is one of those. Through the implementation of technologies is possible to help in the decision making process of clinics, therefore to reduce the difficulties that are typically faced. This study focuses on easing some of those difficulties by presenting real-time data mining models capable of predicting if a monitored patient, typically admitted in intensive care, will need to take vasopressors. Data Mining models were induced using clinical variables such as vital signs, laboratory analysis, among others. The best model presented a sensitivity of 94.94%. With this model it is possible reducing the misuse of vasopressors acting as prevention. At same time it is offered a better care to patients by anticipating their treatment with vasopressors.

**Keywords.** Vasopressors, INTCare, Intensive Medicine, Real-Time, Data Mining, Vital Signs, Laboratory Results.

## 1 Introduction

There is an ongoing effort to implement Information Technologies (IT) in medical facilities, since they can ease various developed activities. In Intensive Medicine (IM) there are numerous devices and technologies helping intensivists to develop their job more precisely, in order to take care of critically ill patients.

The use of Data Mining (DM) is one of such technologies becoming more common in Intensive Care Units (ICU) where the patient is continuously monitored and it is possible collecting data in real-time. Objectively, it seeks to use data produced by many devices and transform it into new knowledge helping the decision making process.

In ICUs the use of vasopressors is very common in order to improve patient condition, however sometimes the therapy is not applied in the correct time. In order to provide a better patient care this study was conducted. The aim of this study is to use DM in order to predict if a patient will need to take a vasopressor or not. If the prediction is verified, the intent is to alert the intensivists. So that they can act before that necessity materializes itself. Otherwise these models are useful to avoid wrong prescriptions.

Data Mining models were induced using real data provided by Hospital Santo António, Centro Hospitalar do Porto, Porto, Portugal. The data were collected from vital signs monitors presented in the ICU, laboratory analysis and Electronic Health Record (EHR).

Analysing the developed DM models, very good results were achieved. The best model reached a sensitivity of 94.94%.

Apart from the introduction, this article is composed by other four chapters. Background is the second chapter where it presents information relative to related work. The third chapter - Study Description - enunciates the methods and tools used in the development of the study and it is presented through the CRISP-DM phases. In the fourth chapter is presented and discussed the results of the study. Lastly, in the fifth chapter is drawn the conclusion regarding to the study.

## 2 Background

Information Systems and Technologies are very important nowadays and allowed progress and success beyond imagined in many diverse areas. The area of Intensive Medicine (IM) belongs to the medical sciences and it seeks to help critically ill patients [1]. Through prevention, diagnosis and treatment, intensivists try to recover the patient's to prior state of health [2]. These activities take place in Intensive Care Units (ICU) which are facilities specifically designed for these types of patients, where the existence of life-support devices that monitor patients' vital signs is abundant. Along with the use of drugs these represent a strong ally to help patients [3].

The use of vasopressors is on focus here. This is a type of drug normally used to increase blood pressure (BP) in patients where BP is minimal and patient's life is at risk [4]. Adrenaline, Noradrenaline and Dopamine are three developed vasopressors that mimic the effects of neurotransmitter substances of the sympathetic nervous system. They act as agonists in  $\alpha_1$  and  $\alpha_2$  adrenergic receptors being responsible for vasoconstriction, which in turn increases blood pressure [5]. The use of vasopressors need to be very carefully planned when a patient is with arrhythmias [6, 7].

The interest in study the use of vasopressors comes from the realization of INTCare [8], a research project being developed in Centro Hospitalar do Porto (CHP). This project implemented an information system in the hospital where the acquisition process was modified. The data acquisition changed from a manual, on paper way to an automatic, electronic and real-time process [9, 10]. The interactions within the system are done through intelligent agents. These agents act autonomously and belong to four different subsystems: Data Acquisition, Knowledge Management, Inference and Interface [11]. Due to an iterative process, it is now a Pervasive Intelligent Decision Support System (PIDSS) which using Data Mining (DM) supports the decision making process in ICU. This PIDSS is able to predict patient's outcome [12, 13], organ failure [14], readmission [15, 16], discharge and length of stay [17, 18], among others [19]. Data Mining can be define as a process of looking for patterns in great amounts of data, with the intent of describing the data or use it to predict future events. It is the conversion of data into useful information [20].

Being DM an integral part of the INTCare system [21] this study seeks to improve on a prior study [22] which had the objective of developing data mining models capable of predicting the use of vasopressors in monitored patients of intensive care units. DM models used variables from vital signs monitors, laboratory analysis and Electronic Health Record (EHR) to make predictions upon the future necessity of vasopressor

intake. The variables were selected based on possible causes for various health conditions that require vasopressors as treatment. In total six different scenarios combining the variables and four different algorithms – Naïve Bayes, Support Vector Machines, Decision Tree and Generalized Linear Model – were induced. In this first phase only raw data without classes were used. The results achieved were very satisfactory having the best model reached the 90.72% in the sensitivity metric.

Despite the results being very good there was still the intent to try and improve upon those results. This time some changes were made to the dataset and to the percentage of test/training of the models, in order to see how the results would cope with that. While before the vital signs and laboratory analysis variables used raw values, this time the values were aggregated into classes, with clinical meaning.

This work represents an ongoing effort of trying to innovate and augment the conditions offered by intensive care facilities of CHP, so that the treatment given to patients is always improving and the intensivists can make more precise decisions.

### **3 Study Description**

#### **3.1 Method and Tools**

This study was developed according to the guidelines provided by the Cross Industry Standard Process for Data Mining (CRISP-DM), a methodology often used in solving data mining related problems. This methodology is composed by six phases, which are Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment. The possibility of moving backwards and forward is one of the characteristics that makes CRISP-DM a very versatile methodology.

Oracle SQL Developer was the tool used to perform the Extract Transforming and Loading Process (ETL) and induce data mining models. It is an Integrated Development Environment (IDE) used for development and management of Oracle Databases. The modeling part of CRISP-DM in which data mining models were developed was done recurring to the Oracle Data Miner extension available in Oracle SQL Developer.

#### **3.2 Business and Data Understanding**

Based on patient's clinical data, the goal of this study is to improve the results of previous developed data mining models to predict if a patient will need vasopressors or not. This way it is possible to provide patients a higher quality of care, by allowing intensivists to make more precise decisions. Also this study differentiates from the previous study because it seeks to aggregate the various variables in classes (using clinical knowledge) and realize if that will increase the quality of the results.

CHP is the provider of the data being used in this study, which concerns vital signs, laboratory analysis and Electronic Health Record (EHR) of patients admitted to the Intensive Care Unit (ICU). The data being used ranges from January 6<sup>th</sup>, 2015 to May 18<sup>th</sup>, 2015 and holds 1259 rows of data concerning 56 distinct patients.

In total seventeen variables were used: SPO2, ECG\_HR, ART\_SYS, TEMP\_T1, PH, Erythrocytes, Potassium, Glucose, Leucocytes, Lactate, PCO2, Hemoglobin, Age, Sex, Provenience, Type\_Hospitalization and Hospitalization\_Surgery. The target variable is VSPGeral which represents if the patient did take a vasopressor or not. Below in Table 1, it is possible to overview the variables used, their type and source.

**Table 1.** Variables Overview

Variable	Type	Source
Saturation of Oxygen (SPO2)	String	Vital Signs
Heart Rate (ECG_HR)	String	Vital Signs
Blood Pressure (ART_SYS)	String	Vital Signs
Temperature (TEMP_T1)	String	Vital Signs
Leucocytes and Erythrocytes	String	Lab Analysis
Potassium and Potential Hydrogen (PH)	String	Lab Analysis
Glucose and Hemoglobin	String	Lab Analysis
Pressure of Carbone Dioxide (PCO2) and Lactate	String	Lab Analysis
Dopamine, Noradrenaline and Adrenaline	String	Lab Analysis
Age	Number	EHR
Sex	String	EHR
Provenience	String	EHR
Type_Hospitalization	String	EHR
Hospitalization_Surgery	String	EHR
VSPGeral	String	-

### 3.3 Data Preparation

In this phase the Extract, Transform and Loading (ETL) process was executed. Since great part of the issues concerning the data used in the previous study, the only iterations done this time was the transformation of the values from string type to number type through the use of a function and the aggregation of the variables' results into classes with clinical meaning. The reason for this change is centered on the fact that a normal distribution could hold results in the wrong class. Therefore the results were divided in three classes: Critical Low, Normal and Critical High. Critical Low represents values below the reference values, Normal represents value within the reference values and Critical High represents values above the reference values.

In Table it is presented the variables and values' range for each of the classes.

**Table 2.** Variables and values' range per class

	Variable	Units	Critic Low	Normal	Critic High
	Temperature (TEMP)	°C	34-36	36-38	38-45
Vital Signs	Blood Pressure (ART SYS)	mmHg	0-90	90-180	180-500
	Heart Rate (ECG_HR)	BPM	0-60	60-120	120-250
	SPO2	%	0-90	90-100	100-500
Laboratory Analysis	Hemoglobin	g/dL	<12,00	12,0 - 18,0	>18,00
	PH		<7,35	7,35 - 7,45	>7,45

Leukocytes	10 <sup>3</sup> /μL	<4,00	4,00 - 11,00	>11,00
Potassium	mmol/L	<3,50	3,50 - 5,30	>5,50
Lactate	mmol/L	<0,50	0,50 - 2,20	>2,20
Glucose	mg/dl	<70,00	70,00-105,00	>105,00
PCO2	mmHg	<32,00	32,00 - 45,00	>45,00
Erythrocytes	10 <sup>6</sup> /μL	<4,10	4,10 - 5,50	>5,50

To change each attribute's field to a class an update to the attributes table was made. This operation changed the field of the attribute to a class name according to the value in the attribute's field. To calculate these classes an algorithm to track patient condition [23] able to calculate critical events [24] was adopted. This method was also repeated for every single one of the variables belonging to vital signs and laboratory analysis. As example it is the updated of Leucocytes attribute:

```
UPDATE PATIENT_DMCLASSES
SET LEUCOCYTES = CASE
WHEN LEUCOCYTES < 4.0 then "Critical Low"
WHEN LEUCOCYTES >= 4.0 AND LEUCOCITOS <=11.0 then
"Normal"
WHEN LEUCOCYTES > 11.0 then "Critical High"
END.
```

The creation of the target variable VSPGeral resulted from the combination of the three vasopressor variables: Dopamine, Adrenaline and Noradrenaline. When a row had at least one of these three variables with value "1", the VSPGeral value for that row was 1. Otherwise the patient did not receive a vasopressor and VSPGeral value was 0.

### 3.4 Modeling

Data mining models were built according to the objective of the study in this phase. Since the objective is to predict a discrete result, i.e., if a patient will need to take a vasopressor or not, the data mining function to be used is Classification.

The structure of the data mining models was maintained the same as in the previous study, resulting in a distribution of the target variable (VSPGeral) with 37.81% of patients having taken vasopressors while 62.19% of patients did not take any vasopressor. The variables were aggregated into groups in order to understand how they could influence the results of the data mining models. In Table 3 it is possible to see which group each variables belong.

**Table 3.** Variables aggregated in groups

Group	Variables
Vital Signs (VS)	SPO2, ECG_HR, ART_SYS, TEMP_T1
Lab Analysis (LA)	PH, Erythrocytes, Potassium, Glucose, Leucocytes, Lactate, PCO2, Hemoglobin
Patient Admission (PA)	Provenience, Type_Admission, Admission_Surgery
Case Mix (CM)	Age, Sex

After the formation of the groups, six new scenarios were modeled. At first the same distribution was used for test and training, respectively, 40% and 60%, but afterwards Cross Validation (all data for training and test) was used for test in order to see the effects on the results. Table 4 displays the modeled scenarios and the groups that compose them.

**Table 4.** Scenarios and their Group of Variables

Scenarios	Groups
S1	VS + LA + PA + CM
S2	VS + LA + PA
S3	VS + LA + CM
S4	LA + CM
S5	VS + CM
S6	PA + CM

Each one of the scenarios were executed using four data mining algorithms: Support Vector Machine (SVM), Decision Tree (DT), Naïve Bayes (NB) and Generalized Linear Model (GLM), which are all the algorithms available in the software being used.

The combination of scenarios and algorithms results in a total of 48 models (6 scenarios \* 4 algorithms \* 2 validation techniques). The following expression can represent a general model:

$$M_n = A_x + S_i + T_z + V_t$$

In the expression  $M_n$  is the model with a classification approach ( $A_x$ ), a scenario ( $S_i$ ), an algorithm ( $T_z$ ) and a validation technique ( $V_t$ ):

In Table 5 is presented the settings used for each one of the data mining algorithms.

**Table 5.** Data Mining Algorithm's Settings

Support Vector Machine	
Kernel Function	Gaussian
Tolerance Value	0,001
Active Learning	Yes
Decision Tree	
Homogeneity Metric	Gini
Maximum Depth	7
Minimum Records in a Node	10
Minimum Percent of Records in a Node	0,05
Minimum Records for a Split	20
Minimum Percent of Records for a Split	0,1
Naive Bayes	
Pairwise and Singleton Threshold	0
Generalized Linear Model	
Reference Class name	1
Missing Value Treatment	Mean Mode
Ridge Regression	Disable
Approximate Computation	Disable

### 3.5 Evaluation

When the evaluation phases of CRISP-DM is reached, the results given by each one of the data mining model, each scenario should be analyzed in order to conclude which scenario has the best results. In this study the results were analyzed based on a confusion matrix and Receiver Operating Characteristic (ROC) curves. The analysis of the confusion matrix resulted on the production of results based on three different metrics: sensitivity, specificity and accuracy.

Being the main goal the prediction of the necessity to take a vasopressor, the main metric selected was sensitivity, because it focuses in predicting true positive values, i.e. patients that will have to take a vasopressor. These analyzes were done over two attempts: the first attempt (CL1) differentiates from the previous study because it has vital signs and analysis' values aggregated in classes. The second attempt (CL2) is equal to CL1, except instead of using a distribution of 40/60 (%) for test/training. CL2 used Cross Validation for test and training. The intent is to understand how the clinical classes and the percentage of test/training of data can affect the final results.

## 4 Discussion

As in the previous study, the focus was to have very sensitive models. In this case the objective was to see if the results got better results based on the aggregation of the variables' values into classes and later by evaluating using CV. The general results were better not only for sensitivity, but also for specificity and accuracy. It was used the same threshold: 85% for sensitivity, 70% for accuracy and 60% for specificity.

The first attempt using the aggregation in classes and a distribution of test/training of 40/60 (%) resulted in better results than in the previous study without classes.

In Table 6 are presented the scenarios (Sc), algorithms (Al) and values of the best three models for metric, belonging to each attempt.

**Table 6.** Top 3 models per metric

Previous Study			CL1			CL2		
Sensitivity			Sensitivity			Sensitivity		
Sc	Al	Value	Sc	Al	Value	Sc	Al	Value
S1	SVM	90.72%	S6	SVM	94.18%	S6	SVM	94.94%
S1	NB	89.22%	S1	DT	93.18%	S3	SVM	94.44%
S3	NB	88.68%	S3/S4	DT	92.89%	S1	SVM	94.01%
Specificity			Specificity			Specificity		
Sc	Al	Value	Sc	Al	Value	Sc	Al	Value
S3	SVM	82.72%	S1	SVM	85.14%	S1	SVM	87.57%
S4	SVM	76.00%	S3	SVM	79.17%	S4	SVM	85.51%
S1	NB	70.44%	S2	SVM	75.94%	S3	SVM	83.49%
Accuracy			Accuracy			Accuracy		
Sc	Al	Value	Sc	Al	Value	Sc	Al	Value
S3	SVM	86.02%	S1	SVM	89.61%	S1	SVM	91.50%
S4	SVM	82.84%	S3	SVM	87.29%	S4	SVM	90.23%
S1	SVM	81.14%	S6	SVM	84.75%	S3	SVM	89.91%

Through the observation of table 6 we can infer that the use of classes in CL1 resulted in an increase of the metrics values, when compared to the Previous Study in which were used the variable values, instead of classes. On the other hand, CL2 which used cross validation (CV) also proved to increase results for all measures when to compare to the first attempt, and even more so when compared to the previous study.

In parallel with the models of the previous study, all of the metrics increased. In CL1, sensitivity increased in average 3.88%, specificity 3.69% and accuracy 3.87%.

On the other hand, CL2 which also used classes but used CV for test and training showed even better results than the first attempt. This time the increase relative to the previous study was much more substantial, having the sensitivity increased in average 4.92%, specificity 9.14% and accuracy 7.21%.

The best model is now belonging to CL2, where scenario 6 and algorithm SVM surpassed the both best models from Previous Study and CL1 with 94.94%

Concerning the ROC curves, it is possible to observe in Fig. 1 and Fig. 2 that the use of classes increased marginally the area under the curve of the most sensitive models belonging to CL1 and CL2. While in the previous study the ROC curve of the most sensitive model had a value of 83.06%, now CL1 has a value of 84.75% and CL2 has a value of 83.80%.

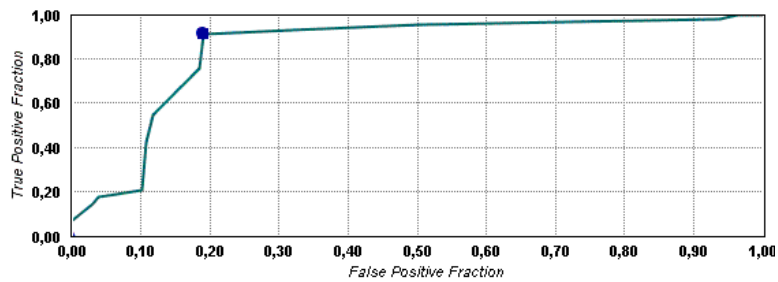


Fig. 1. CL1 – S6/SVM – ROC curve

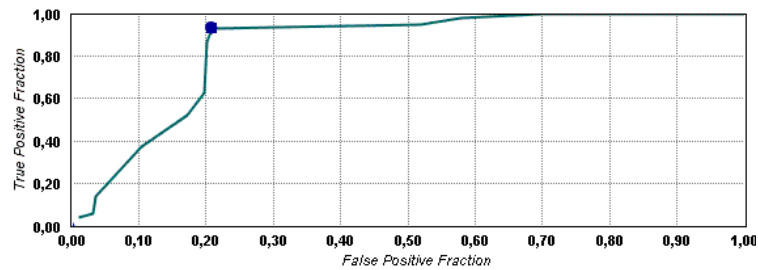


Fig. 2. CL2 – S6/SVM – ROC curve

## 5 Conclusion

The results obtained in this study represent the success of the objectives defined - the improvement of the results attained in the previous study. Prediction results increased on all metrics, but more importantly on sensitivity, which it is the metric that predicts



if the patient will need to take a vasopressor. This improvement will increase the strength that previous models already had. At first it benefits the patients by avoiding the side effects that vasopressors might have. In second this models seeking to improve their health condition. In fact intensivists will have less stressful and uncertain decisions and the hospital will reducing costs and waste of resources by avoiding wrong therapeutics associated to the use of vasopressors.

Overall the correct implementation of data mining models that can predict a relevant aspect concerning patient's health is always very important and welcome.

### Acknowledgments

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

### References

1. Kaur, M., Pawar, M., Kohli, J.K., Mishra, S.: Critical events in intensive care unit. *Indian Journal of Critical Care Medicine: Peer-reviewed, Official Publication of Indian Society of Critical Care Medicine* 12, 28 (2008)
2. Silva, Á.J.B.M.d.: Modelos de inteligência artificial na análise da monitorização de eventos clínicos adversos, disfunção/falência de órgãos e prognóstico do doente crítico. (2007)
3. Ramon, J., Fierens, D., Güiza, F., Meyfroidt, G., Blockeel, H., Bruynooghe, M., Van Den Berghe, G.: Mining data from intensive care patients. *Advanced Engineering Informatics* 21, 243-256 (2007)
4. Portela, F., Filipe Santos, M., Abelha, A., Machado, J., Rua Martins, F., Silva, Á.: Real-time Decision Support using Data Mining to predict Blood Pressure Critical Events in Intensive Medicine Patients. 9456, (2015).
5. Elliott, J.: Alpha-adrenoceptors in equine digital veins: Evidence for the presence of both alpha-1 and alpha-2-receptors mediating vasoconstriction. *Journal of veterinary pharmacology and therapeutics* 20, 308-317 (1997)
6. Greenberg, H.B.: Cardiac Arrhythmias: Their Mechanisms, Diagnosis, and Management. *JAMA* 246, 169-169 (1981)
7. Filipe Portela, M.S., Jose Machado, Antonio Abelha, Álvaro Silva, Fernando Rua: Preventing Patient Cardiac Arrhythmias by Using Data Mining Techniques. In: 2014 IEEE Conference on Biomedical Engineering and Sciences. (2014)
8. Portela, F., Santos, M.F., Machado, J., Abelha, A., Silva, Á., Rua, F.: Pervasive and intelligent decision support in Intensive Medicine—the complete picture. *Information Technology in Bio-and Medical Informatics*, pp. 87-102. Springer (2014)
9. Portela, C.F., Santos, M.F., Silva, Á., Machado, J., Abelha, A.: Enabling a Pervasive Approach for Intelligent Decision Support in Critical Health Care. In: *ENTERprise Information Systems*, vol. 221, pp. 233-243. Springer Berlin Heidelberg (2011)
10. Portela, F., Santos, M.F., Silva, A., Machado, J., Abelha, A., Rua, F.: Data mining for real-time intelligent decision support system in intensive care medicine. (2013)
11. Santos, M.F., Portela, F., Vilas-Boas, M.: Intcare: multi-agent approach for real-time intelligent decision support in intensive medicine. (2011)
12. Santos, M.F., Mathew, W., Portela, C.F.: Grid Data Mining for Outcome Prediction in Intensive Care Medicine. In: *Enterprise Information Systems*, vol. 221, pp. 244-253 (2011)

13. Boas, M.V., Santos, M.F., Portela, F., Silva, A., Rua, F.: Hourly prediction of organ failure and outcome in real time in Intensive Care Medicine (2010)
14. Portela, F., Santos, M.F., Machado, J., Abelha, A., Silva, Á.: Pervasive and Intelligent Decision Support in Critical Health Care Using Ensembles. *Information Technology in Bio- and Medical Informatics*, pp. 1-16. Springer Berlin Heidelberg (2013)
15. Pedro Braga, Portela, F., Santos, M.F.: Data Mining Models to Predict Patient's Readmission in Intensive Care Units. In: ICAART - International Conference on Agents and Artificial Intelligence. (2014)
16. Rui Veloso, F.P., Manuel Filipe Santos, Álvaro Silva, Fernando Rua, António Abelha, José Machado: Categorize readmitted patients in Intensive Medicine by means of Clustering Data Mining. *International Journal of E-Health and Medical Communications (IJEHMC)* (2015) (accepted for publication)
17. Veloso, R., Portela, F., Santos, M., Machado, J.M.F., Abelha, A., Silva, Á., Rua, F.: Real-time data mining models for predicting length of stay in intensive care units. (2014)
18. Filipe Portela, R.V., Sérgio Oliveira, Manuel Filipe Santos, António Abelha, José Machado, Álvaro Silva and Fernando Rua: Predict hourly patient discharge probability in Intensive Care Units using Data Mining. *ScienceAsia Journal (ICCSCM 2014)* (2014)
19. Oliveira, S., Portela, F., Santos, M.F.: Pervasive Universal Gateway for Medical Devices. *Recent Advances in Electrical Engineering and Education Technologies (SCI 2014)*, pp. 205-210 (2014)
20. Hardin, J.M., Chhieng, D.C.: Data Mining and Clinical Decision Support Systems. In: Berner, E.S., FACMI, FHIMSS (eds.) *Clinical Decision Support Systems*. Springer, New York (2007)
21. Filipe Portela, Filipe Pinto, Santos, M.F.: Data Mining Predictive Models For Pervasive Intelligent Decision Support In Intensive Care Medicine. In: INSTICC (ed.) *KMIS 2012*, Barcelona (2012)
22. Braga, A., Portela, F., Filipe Santos, M., Machado, J., Abelha, A., Silva, Á., Rua, F.: Data Mining to predict the use of Vasopressors in Intensive Medicine Patients. *Jurnal Teknolog. Penerbit UTM Press* (2016). (accepted for publication)
23. Filipe Portela, S.O., Manuel Filipe Santos, António Abelha, José Machado: A Real-Time Intelligent System for tracking patient condition. In: Springer (ed.) *LNCS - Ambient Intelligence for Health*, vol. 9456, (2015)
24. Portela, F., Gago, P., Santos, M.F., Machado, J., Abelha, A., Silva, Á., Rua, F.: Pervasive real-time intelligent system for tracking critical events in intensive care patients. (2013)