Preventing Patient Cardiac Arrhythmias by Using Data Mining Techniques

Filipe Portela – *University of Minho*, Manuel Filipe Santos – *University of Minho*, Álvaro Silva – *Centro Hospitalar do Porto*, Fernando Rua – *Centro Hospitalar do Porto*, António Abelha – *University of Minho*, José Machado – *University of Minho*

***Abstract*— Cardiac Arrhythmia (CA) is very dangerous and can significantly undermine patient condition. New tools are fundamental to forecast and to prevent possible critical situations. In order to help clinicians acting proactively, predictive data mining real-time models were induced using online-learning. As input variables were considered those acquired at the patient admission and complementary variables (vital signs, laboratory results, therapeutics) hourly collected. The results are very motivating; sensitivity near to 95% was obtained when using Support Vector Machines. The approach explored in this work reveals to be an interesting contribution to the healthcare in terms of predicting CA and a good direction to be further explored.**

I. Introduction

Nowadays, hospitals have a set of mechanisms to collect patient data, however in many cases these data are not used in the decision process. Typically the data are analyzed in the moment when it is collected and then they are stored in databases, being not considered for future actions. Having these data a lot of hidden potential it is notorious a rapidly increasing of the development of decision support systems and the use of Data Mining (DM).

DM allows the introduction of new knowledge / patterns essential for the decision making process. In this area DM can be used to predict future events e.g. clinic, management or financial. When applied in a real context, DM can bring many benefits to the healthcare, being the main benefit the patient condition improvement. One of the main patient problems verified in the society / hospitals is the occurrence of Cardiac Arrhythmia (CA). CA is identified in a patient when the heart activity is irregular or it is faster or slower than the normal [1].

When a patient has a Heart Rate (HR) value out of the normal range for a certain period of time or presents a too high or a too low value, can be told that the patient had a HR Critical Event (CE). At the present time, to the medical staff it is very difficult to anticipate a CA, because there are a lot of variables influencing it. At same time, they cannot analyze in a reduced period of time all the patient variables collected during the last minutes / hours. Motivated by this limitation an exploratory study was conducted in order to develop models for predicting the probability of a patient be a victim of CA during the next hours.

The work was performed in a real environment using data collected in real-time in the Intensive Care Unit of Centro Hospitalar do Porto and having an inpatient time under the first 120 hours (5 days), because after this length of stay extra complications may appear (e.g. infections, bacteria, others dysfunctions).

The dataset includes a set of patient data: data acquired during the admission process (admission from, admission type, age, risk factors and other) and some others collected or calculated in real-time (Heart Rate, SOFA, Critical Events and other). Four DM techniques were considered: Generalized Linear Models (GLM), Support Vector Machine (SVM), Decision Trees (DT) and Naïve Byes (NB). To assess the results a confusion matrix has been considered for each model to calculate three metrics: sensitivity, specificity and accuracy.

This paper is divided in five sections. The first one introduces the work, the second presents the background focused in Cardiac Arrhythmia, Intelligent Decision Support Systems and Data Mining. The third section presents the work developed. Fourth section discusses the results achieved and the last section presents some final remarks and points some future work.

II. BACKGROUND

A. Cardiac Arrhythmia

Cardiac Arrhythmia (CA) is defined as an abnormal Heart Rhythms - important changes in the normal heart activity. A heartbeat that is too fast is called tachycardia (over 100 beats per minute) and a heartbeat that is too slow is called bradycardia (less than 60 beats per minute) [1]. Using this definitions and the values associated to them it was defined the concept of Critical Events (CE) [2, 3]. CE is composed by a collection of critical values, i.e., values out of the normal range.

This work makes use of two different concepts: critical values and CE. Critical values are values that are out of a normal range during an undefined period of time. CE is a label to identify that a variable had critical values for more than the admissible time span (type a), as defined in Table I.

Research supported by FCT – Fundação para a Ciência e Tecnologia.

Filipe Portela (corresponding author, e-mail: cfp@dsi.uminho.pt.), Manuel Filipe Santos (mfs@dsi.uminho.pt), António Abelha (abelha@di.uminho.pt) and José Machado (jmac@di.uminho.pt) are from Algoritmi Centre, University of Minho, Portugal.

Álvaro Silva (alvaromoreirasilva@clix.pt) and Fernando Rua (fernandorua.sci@chporto.min-saude.pt) are from Serviço Cuidados Intensivos, Centro Hospitalar do Porto, Hospital Santo António, Porto, Portugal.

CE can also be verified when a critical value is so out of range that it is considered serious regardless of the duration of that observation (type b). For example, a CE happens whenever the patient's heart rate stays above 120 bmp or below 60 bpm for more than 1 hour. Also, a CE may happen every time the Heart Rate drops below 30 bpm or it is upper 180 bpm.

TABLE I. THE PROTOCOL FOR THE OUT OF RANGE PHYSIOLOGIC MEASUREMENTS (ADAPTED FROM [2])

Definition	**Heart Rate (bpm)**
Normal range	60 to 120
Critical event [a]	>= 1h
Critical event [b]	<30 >180

a Defined when continuously out of range.

b Defined anytime.

Arrhythmias increases the patient risk of developing conditions such Stroke and Heart failure (cardiovascular organ failure).

Stroke is defined by the National Stroke Association (NSA) [4] as: a stroke or "brain attack" occurs when a blood clot blocks an artery (a blood vessel that carries blood from the heart to the body) or a blood vessel (a tube through which the blood moves through the body) breaks, interrupting blood flow to an area of the brain. When either of these things happens, brain cells begin to die and brain damage occurs. According to NSA a stroke is considered the fourth leading cause of death in America and a leading cause of adult disability.

Using as reference the definition of NIH Heart, Lung and Blood Institute, Heart Failure is "a condition in which the heart can't pump enough blood to meet the body's needs. In some cases, the heart can't fill with enough blood. In other cases, the heart can't pump blood to the rest of the body with enough force. Some people have both problems. The term "heart failure" doesn't mean that your heart has stopped or is about to stop working. However, heart failure is a serious condition that requires medical care."

Analyzing the study provided by the American Heart Association [5], the number of patients which evidences the presence of stroke are meaningful, being upper in patient older than 60 years (6-7%) and most common in patient with more than 80 years (14%). In terms of Heart Failure, in USA approximately 11% of the deaths were due this failure. This type of symptoms leads to single or multiple organ failures. Normally this type of patients is cared in the Intensive Care Units (ICU).

In ICU is used knowledge of a specific area of medicine namely, Intensive Medicine. The ICUs are composed by a specific type of patients, normally in weak conditions (risk-life) and with multiple organ failures. The cardiovascular system is one of the organ systems that can more contribute to the outcome of critical ill patients [6].

B. Intelligent Decision Support Systems

Intelligent Decision Support Systems (IDSS) are designed to support the decision making process. Typically this type of systems use intelligent agents (to perform their tasks automatically) [7] and artificial intelligence (e.g. make predictions using Data Mining Techniques) to achieve their goals. The decision support process is divided in five phases [8]: Intelligence, Design, Choice, Implementation and Monitoring.

In the group of IDSS specialized to ICU arises INTCare. INTCare is a Pervasive IDSS that changed the way of how patient data are collected and then disseminated [9, 10]. This system collects automatically and in real-time a set of parameters (vital signs, ventilation, laboratory results and therapeutics) making them available anywhere and anytime. Combining these parameters with others obtained from the patient admission phase and from nursing records, is possible to induce Data Mining models in order to predict clinical events.

INTCare is divided in two platforms, one for monitoring patient data and another for presenting new knowledge obtained by means of Data Mining. Actually, it predicts in real-time and using online-learning some patient features and clinical conditions [11-14]: outcome, organ failure, sepsis level, readmission and length of stay. The approach followed in this paper is an extension of the work developed until now, focused in a very particular disease: CA.

C. CRISP-DM

Cross Industry Standard Process for Data Mining (CRISP-DM) provides a structured and interactive approach for planning a data mining project.

As can be observed in figure 1 this approach is divided in six phases [15] [16]: Business Understanding (understanding the project objectives and requirements), Data Understanding (data collection and statistics), Data Preparation (transforming phase) , Modeling (modeling techniques are selected and applied) , Evaluation (assess the models results) and Deployment (deploy the model, e.g. embedding it into an IDSS).

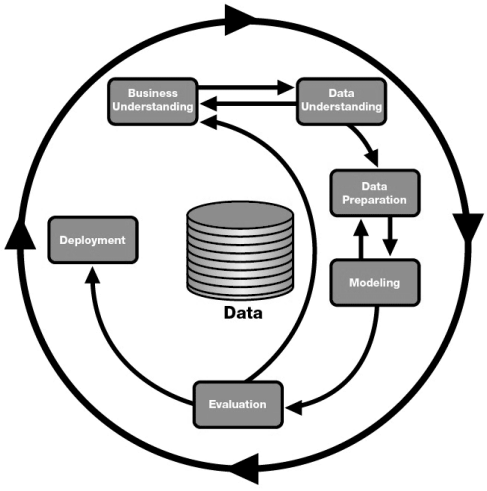

Figure 1. CRISP – DM approach

III. DATA MINING APPROACH

The Data Mining approach was conducted by CRISP-DM methodology.

A. *Business Understanding*

As previously referred, the cardiovascular system is a patient vital system. Being it responsible to a high number of deads, it is very important to a physician to determine in advance if a patient is or not in risk of having a CA. This concern can be translated into a DM question:

How Data Mining can help clinicians in predicting the CA occurrence probability associated to a patient?

B. *Data Understanding*

In order to induce DM models capable of answering to the problem it was used data provided from four data sources:

Vital Signs – {Blood Pressure and Heart Rate}

Electronic Health Records – {Admission variables and Age}

Therapeutic Plan – {Vasopressores}

Laboratory Results – {Bilirubin, Creatinine, Po2/Fio2 and Platelets}

Table II presents the class distribution (percentage of cases) of the age and the non-numeric variables. These variables were not submitted to the transformation phase, because the value associated do not requires extra calculations. It is a simple match between the values collected and the DM attributes classes.

TABLE II. NON-NUMERIC VARIABLE DISTRIBUTION

ID	Variable	Min	Max	Class	% Cases
Age	Age	18	46	1	16,94%
		47	65	2	35,09%
		66	75	3	20,51%
		76	130	4	27,47%
Admission Type	Urgent	-	-	U	80,23%
	Programmed	-	-	P	19,77%
Admission From	Chirurgic	-	-	1	46,80%
	Observation	-	-	2	0,14%
	Emergency	-	-	3	18,53%
	Nursing Room	-	-	4	15,28%
	Other ICU	-	-	5	2,21%
	Other Hospital	-	-	6	1,68%
	Other	-	-	7	15,37%
Insufficiencies Cardiac	Yes			1	92,77%
	No			0	7,23%

Figure 2 represents the target (critical / CA) distribution. The occurrence of a CA are represented by the value 1, and by 0 the patient that had not a CA. Analyzing the figure it is possible observe that 23% of the records has a CA associated.

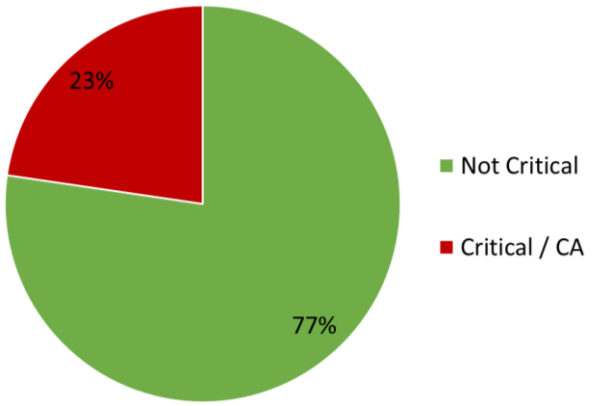

Figure 2. Output variable dsitribution

C. *Data Preparation*

In the data preparation phase, all the variables were validated. It was verified if there is not null values, or values out of the normal range (noise values). These tasks were performed by an Intelligent Agent. Then, after all the data be validated, another agent was responsible to make a mapping between the data collected and the DM variables. In this case, the following tasks were performed:

- To verify and group the admission type and admission from;
- To classify the heart rate values as critic or not;
- To determine the patient SOFA value for all the organic systems;
- To create a new variable which identifies if the patient has risk factors;
- To determine the last seven Heart Rate values (HRLV) collected;
- To calculate the accumulated critical events (ACE);

The Case Mix variables were provided by the Electronic Health Record (EHR) of the patient. For these models were used the variables: age, admission type, admission from, insufficiencies cardiac and risk patient (combination of patient admission variables). All these variables were obtained at patient admission and they were automatically transformed according to the DM attributes.

The Sequential Organ Failure Assessment (SOFA) is used in ICU to score the degree of dysfunction/failure of the six organic systems (cardiovascular, respiratory, renal, liver, hematological and neurological) [17]. The value 0 represents a normal functioning of the organic system. In the case of the SOFA score be between 1 and 4 the patient has one or more organ failures. The attribute SOFA used in DM only considers two values 0 (SOFA = 0) or 1 (SOFA > 0).

The Accumulated Critical Events (ACE) includes three physiological variables: Blood Pressure (BP), Saturation of Oxygen (SPO2) and Heart Rate (HR). ACE are calculated by counting the number of critical events verified by hour since the patient admission.

The ratios were introduced by using ACE group. These ratios allow to determine a relation between the number of ACE verified and the maximum of ACE occurred by hour (R1) and a correspondence between the number of ACE verified and the maximum number of events verified in the past, until the hour in analysis (R2). Both the values are grouped by category and by patient.

After performing the pre-processing tasks, a set of new variables were introduced. These variables and their distribution are presented in the table III. For example, the SOFA_CARDIO attribute is 1 when at least one of the conditions defined is verified (e.g. Blood pressure (BP) >=0 and BP <=70) or the use of any vasopressor > 0). In the case of Risk Patient, the attribute is 1 when one of the symptoms or conditions defined is verified. These conditions were verified in 19.85% of the cases.

TABLE III. TRANSFORMED VARIABLES DSITRIBUTION

Attribute	Variable	Min	Max	Value	% Cases
SOFA_CARDIO	BP (mean)	0	70	1	74,07%
	Dopamine	0.0	-	1	
	Dobutamine	0.0	-	1	
	Epi / Norepi	0.0	-	1	
SOFA_RENAL	Creatinine	1.2	-	1	61,61%
SOFA_RESPIRAT	Po2/Fio2	0	400	1	20,57%
SOFA_HEPATIC	Bilirubin	1.2	-	1	44,44%
SOFA_HEMAT	Platelets	0	150	1	19,28%
Risk Patient	CVA	-	-	1	19,85%
	Alcoholism	-	-	1	
	Pacemaker	-	-	1	
	Addicted	-	-	1	
	Corticoids	-	-	1	
	Transplanted	-	-	1	
	Vasoactive Drug	-	-	1	

Table IV presents a static analysis (average, maximum, minimum, mode and standard deviation (stdev)) of the numeric variables: ACE, Ratios and Heart Rate (HRLV1-HRLV7).

TABLE IV. TRANSFORMED VARIABLES DSITRIBUTION

Attribute	Average	Max	Min	Mode	Stdev
ACE_HR	0,90	16,00	0,00	0,00	1,95
ACE_HR_MAX	0,09	1,00	0,00	0,00	0,18
ACE_HR_TIME	0,01	0,50	0,00	0,00	0,03
ACE_BP	0,68	44,00	0,00	0,00	1,98
ACE_BP_ MAX	0,14	4,40	0,00	0,00	0,31
ACE_BP_TIME	0,01	0,43	0,00	0,00	0,03
ACE_O2	1,73	60,00	0,00	0,00	4,04
ACE_O2_MAX	0,06	1,00	0,00	0,00	0,12
ACE_O2_TIME	0,04	1,00	0,00	0,00	0,07
ACE_TOT	3,32	74,00	0,00	0,00	5,83
ACE_TOT_MAX	0,09	1,00	0,00	0,00	0,14
ACE_TOT_TIME	0,06	1,00	0,00	0,00	0,10
HRLV1	81	250	0	75	21
HRLV2	81	250	0	75	21
HRLV3	81	250	0	75	21
HRLV4	81	250	0	75	21
HRLV5	81	250	0	75	22
HRLV6	81	250	0	75	22
HRLV7	81	250	0	75	22

Finally, in order to prepare all the DM scenarios, the values were organized by hour in turn of the following variables:

Hour: The hour associated to the values collected, all the models use this variable;

Case Mix (CM) – Age, admission type, admission from, Risk Patient, Insufficiencies Cardiac;

SOFA – Cardiovascular, Respiratory, Renal, Hepatic, Hematologic;

Accumulated Critical Events (ACE) – ACE of Blood Pressure (BP), ACE of Oxygen Saturation (SPO2), ACE of Heart Rate (HR) and Total ACE;

Ratios 1 (**R1)** – ACE of BP / max number of ACE of BP, ACE of SPO2 / max number of ACE of SPO2, ACE of HR / max number of ACE of HR;

Ratios 2 (R2) – ACE of BP / elapsed time of stay, ACE of SPO2 / elapsed time of stay, ACE of HR / elapsed time of stay, Total of ACE / elapsed time of stay;

Ratios (R) – Union of the two sets of ratios (R1 and R2).

Heart Rate Last Values (HRLV): The last seven values collected in an hour, being one value for each column.

The data used in the models correspond to patients admitted into ICU of CHP from 2012.02.01 to 2014.05.28 (848 days), considering 407 patients and 310160 records.

D. Modelling

Finished the data preparation phase, a set of Data Mining models (DMM) were induced using four DM techniques (DMT): GLM, SVM, DT and NB. The developed models used 70% of the data for training and 30% for testing (holdout sampling).

Scenarios 1 to 7 were manually configured. Scenario 8 (S8) was automatically configured by using variables selected by the DM engine (based in heuristic rules). A total of 32 models were induced.

DMM = {*8 Scenarios, 4 Techniques, 1 Target*}

Where the scenarios are:

S1: {CM, ACE, R, SOFA}
S2: {CM, SOFA}
S3: {CM, ACE, R}
S4: {CM, ACE, R, SOFA, HRLV}
S5: {CM, SOFA, HRLV}
S6: {CM, ACE, R, HRLV}
S7: {CM, HRLV}
S8: {Automatic}

Techniques:

T1: Generalized Linear Models (GLM),
T2: Support Vector Machine (SVM),
T3: Decision Trees (DT),
T4: Naïve Byes (NB)

Target:

1: Cardiac Arrhythmia

Each one of the models were induced automatically and in real-time, using streamed data and online-learning. The data mining engine used the data prepared and stored in the DM input table. DMM can be represented by the following tuple:

DMM = <Δ, α, DMT, Hour, Age, AdmissionFrom, AdmissionType, Risk, ace_bp, ace_bp_time, ace_bp_max, ace_hr, ace_hr_time, ace_hr_max, ace_spo2, ace_spo2_time, ace_spo2_max, total_ace, total_ace_time, total_ace_max, SOFA_Respiratory, SOFA_Cardiovascular, SOFA_Hepatic, SOFA_Renal, SOFA_Hematologic, HRLV1, HRLV2, HRLV3, HRLV4,HRLV5, HRLV6, HRLV7 >

Where,

Δ is the DM rules,
α is the DM model configuration,
DMT is the DM technique,
Hour ... HRLV7 are the variables used by each model

For example if the Model 1 is composed by S1 and GLM can be represented as:

DMM1 = <Δ, α, GLM, Hour, Age, AdmissionFrom, AdmissionType, Risk, ace_bp, ace_bp_time, ace_bp_max, ace_hr, ace_hr_time, ace_hr_max, ace_spo2, ace_spo2_time, ace_spo2_max, total_ace, total_ace_time, total_ace_max,, SOFA_Respiratory, SOFA_Cardiovascular, SOFA_Hepatic, SOFA_Renal, SOFA_Hematologic>

E. Evaluation

In order to evaluate the models it was designed a confusion matrix for each one of the models. The confusion matrix (CMX) allows to determine the number of True Positives (TP) (predicted 1 and real 1), False Positives (FP) (1, 0), True Negatives (TN) (0, 0) and False Negatives (FN) (0, 1). Using the CMX it was possible calculate some measures. To this study it was used: accuracy, sensitivity and specificity.

Sensitivity = TP / (TP +FN)

Specificity: TN / (TN+FP)

Accuracy: TP/ (TP+FP+TN+FN)

Table V presents the results of the best scenarios for each measure distributed by technique. For example, for all the techniques, the best specificity value was achieved by the same scenario (S3). The best sensitivity (95.93%) was achieved by the scenario1 using SVM.

TABLE V. BEST MODEL FOR EACHE TECHNIQUE ANS MEASURE

Technique	Sensitivity		Specificity		Accuracy	
GLM	S3	0.8120	S3	0.8554	S3	0.8480
SVM	S1	0.9593	S3	0.6576	S7	0.6998
DT	S4	0.8652	S3	0.8251	S3	0.8277
NB	S3	0.8199	S3	0.8284	S6	0.8192

IV. DISCUSSION

Using the IDSS the physicians can choose which type of models they prefer. Sensitivity – very good models to predict the occurrence of CA, but can present a high number of false positives. Specificity – very good models to predict the probability of non-occurring a CA, but can present a high number of false negatives. Accuracy - good models to predict CA, however can fail in predicting an important number of positive or negative cases.

In general in this type of problems, the physicians prefer to act in order to prevent some clinical situations. Due this reason they normally prefer sensitive models. Although this measure presents false positives, the number of true positives is higher and in the most cases these models can foresee all the positive occurrences (sensitivity near of 100%).

Table VI presents the best model for each metric and the respective values of other metrics calculated. The most sensitive model (Scenario 1 and Technique 2 – S1T2) scored correctly in 95% (0.9558) of CA cases; however predicted 65% (0.6576) of positive cases, i.e., these cases were not verified in the reality.

The high number of false positives happens due to the fact of the physicians sometimes can anticipate a probable occurrence of CA and then act in time to prevent / gain more time to avoid the CA. In this cases the system predicts a CA but in the reality did not happens due to a quick and efficiently response by the physicians. The most sensitive model (S1T2) used the scenario 1 (CM, ACE, R, SOFA) and technique 2 (SVM).

The most accurate model has all the metrics upper than 80%, being induced by using GLM technique and scenario 3.

TABLE VI. BEST MODE BY EACHE MEASUE

Model	Accuracy	Sensitivity	Specificity
S1T2	0.4473	**0.9558**	0.3397
S3T1	**0.8489**	0.8120	0.8566
S3T1	0.8489	0.8120	**0.8566**

After discuss the results with the health professionals it was defined a threshold combining accuracy, sensitivity and specificity in order to have if possible a good model with a reduced number of false positives. To a model be select, it need to achieve the following threshold:

- Sensitivity >= 90%
- Accuracy >= 65%
- Specificity >= 60 %

Having as base the threshold defined, the best models which achieved the measure are presented in the table VII.

TABLE VII. BEST MODE BY THE THREASHOLD

Model	Accuracy	Sensitivity	Specificity
S8T2	0.6940	0.9083	0.6312
S5T2	0.6961	0.9074	0.6343
S6T2	0.6922	0.9042	0.6301

In Figure 3 it is possible to observe the receiver operating characteristic (ROC) curve for the model which achieved the threshold and presented the best sensitivity (S8T2). The curve is created by plotting the true positive rate (sensitivity) against the false positive rate (specificity) at various threshold settings. Analyzing the ROC it is possible observe the model behavior for multiple probability thresholds.

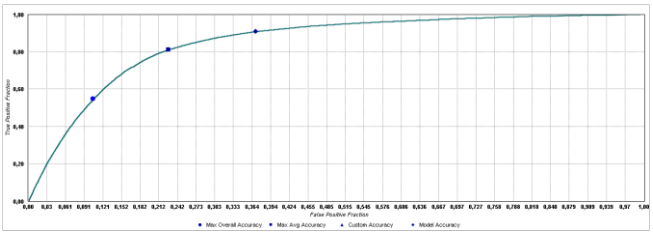

Figure 3. ROC for t he most sensitivity model

V. CONCLUSION AND FUTURE WORK

The approach presented in this work can be considered an important contribution for predicting Cardiac Arrhythmias. DM models shown to be very useful for supporting the decision making process and for preventing future critical events. Eight different scenarios were considered by combining different variable of the patient data. The most sensitive model achieved a result of 95.58% and the most accurate a result of 84.89%.

Analyzing the threshold defined, the best model is defined automatically by the engine using heuristic rules. In this case it is important to enforce the idea of this model be the best now, but in the future with more data the results can be different. Being these models part of the INTCare system, the DM engine is able to always ensure that it is chosen the best model.

Further work will include the introduction of new variables and classes for the numeric attributes, the use of ensembles and the improvement of a target metric to validate the models.

ACKNOWLEDGMENTS

This work has been supported by FCT – Fundação para a Ciência e Tecnologia in the scope of the projects: PEstOE / EEI / UI0752 / 2014 and PestOE / EEI / UI0319 / 2014.

The authors would like to thank FCT (Foundation of Science and Technology, Portugal) for the financial support through the contract PTDC/EEI-SII/1302/2012 (INTCare II).

REFERENCES

[1] H. B. Greenberg, "Cardiac Arrhythmias: Their Mechanisms, Diagnosis, and Management," *JAMA*, vol. 246, pp. 169-169, 1981.

[2] Á. Silva, P. Cortez, M. F. Santos, L. Gomes, and J. Neves, "Rating organ failure via adverse events using data mining in the intensive care unit," *Artificial Intelligence in Medicine*, vol. 43, pp. 179-193, 2008.

[3] F. Portela, P. Gago, M. F. Santos, J. Machado, A. Abelha, Á. Silva, *et al.*, "Pervasive real-time intelligent system for tracking critic events in intensive care patients," 2013.

[4] A. National Stroke, "Stroke 101 fact sheet," *Retrieved April*, vol. 1, 2010.

[5] A. S. Go, D. Mozaffarian, V. L. Roger, E. J. Benjamin, J. D. Berry, M. J. Blaha, *et al.*, "Heart disease and stroke statistics--2014 update: a report from the American Heart Association," *Circulation*, vol. 129, p. e28, 2014.

[6] V. D. Mayr, M. W. Dünser, V. Greil, S. Jochberger, G. Luckner, H. Ulmer, *et al.*, "Causes of death and determinants of outcome in criticly ill patients," *Critic Care*, vol. 10, p. R154, 2006.

[7] L. Cardoso, F. Marins, F. Portela, M. Santos, A. Abelha, and J. Machado, "The Next Generation of Interoperability Agents in Healthcare," *International journal of environmental research and public health*, vol. 11, pp. 5349-5371, 2014.

[8] E. Turban, J. E. Aronson, and T.-P. Liang, *Decision Support Systems and Intelligent Systems*, 7 ed.: Prentice Hall, 2005.

[9] F. Portela, M. F. Santos, Á. Silva, F. Rua, A. Abelha, and J. Machado, "Adoption of Pervasive Intelligent Information Systems in Intensive Medicine," *Procedia Technology*, vol. 9, pp. 1022-1032, 2013.

[10] F. Portela, M. F. Santos, and M. Vilas-Boas, "A Pervasive Approach to a Real-Time Intelligent Decision Support System in Intensive Medicine," in *Communications in Computer and Information Science*. vol. 0272, Springer, Ed., ed, 2012, p. 14.

[11] F. Portela, M. F. Santos, J. Machado, A. Abelha, and Á. Silva, "Pervasive and Intelligent Decision Support in Critic Health Care Using Ensembles," in *Information Technology in Bio-and Medical Informatics*, ed: Springer Berlin Heidelberg, 2013, pp. 1-16.

[12] Pedro Braga, F. Portela, and M. F. Santos, "Data Mining Models to Predict Patient's Readmission in Intensive Care Units," in *ICAART - International Conference on Agents and Artificial Intelligence*, Angers, France, 2014.

[13] João M. C. Gonçalves, Filipe Portela, Manuel F. Santos, Álvaro Silva, José Machado, and A. Abelha, "Predict Sepsis Level in Intensive Medicine - Data Mining Approach," in *Advances in Intelligent Systems and Computing*, Springer, Ed., ed: Springer, 2013.

[14] R. V. Filipe Portela, 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.

[15] R. Wirth and J. Hipp, "CRISP-DM: Towards a standard process model for data mining," in *4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 2000, pp. 29-39.

[16] G. Harper and S. D. Pickett, "Methods for mining HTS data," *Drug Discovery Today*, vol. 11, pp. 694-699, 2006.

[17] J. L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonca, H. Bruining, *et al.*, "The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure," *Intensive care medicine*, vol. 22, pp. 707-710, 1996.