

Available online at www.sciencedirect.com**ScienceDirect**

Procedia Computer Science 64 (2015) 601 – 609

Procedia
Computer Science

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

Predicting Type of Delivery by Identification of Obstetric Risk Factors through Data Mining

Sónia Pereira, Filipe Portela*, Manuel Filipe Santos, José Machado, António Abelha

Algoritmi Research Center, University of Minho, Portugal

Abstract

In Maternity Care, a quick decision has to be made about the most suitable delivery type for the current patient. Guidelines are followed by physicians to support that decision; however, those practice recommendations are limited and underused. In the last years, caesarean delivery has been pursued in over 28% of pregnancies, and other operative techniques regarding specific problems have also been excessively employed. This study identifies obstetric and pregnancy factors that can be used to predict the most appropriate delivery technique, through the induction of data mining models using real data gathered in the perinatal and maternal care unit of Centro Hospitalar of Oporto (CHP). Predicting the type of birth envisions high-quality services, increased safety and effectiveness of specific practices to help guide maternity care decisions and facilitate optimal outcomes in mother and child. In this work was possible to acquire good results, achieving sensitivity and specificity values of 90.11% and 80.05%, respectively, providing the CHP with a model capable of correctly identify caesarean sections and vaginal deliveries.

© 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: Data Mining; Type of delivery; Interoperability; Real data; Obstetrics Care; Maternity Care; Pregnant.

1. Introduction

Several techniques are used to perform child delivery in Maternity Care. Commonly, the delivery procedures are

* Corresponding author. Tel.: +351-253 510 319.
E-mail address: cfp@dsi.uminho.pt

addressed as vaginal deliveries and caesarean sections. Vacuum extraction and obstetric forceps are operative procedures used during complicated vaginal deliveries [1]. A caesarean section is performed in risk pregnancies when natural vaginal delivery and operative procedures are not appurtenant. The decision as to whether or not a particular birth requires assistance and the choice and timing of any intervention must involve consideration of the patient conditions and the risks of the potential techniques, as well as the urgency of the need to expedite the birth process [2] [3]. Predicting the most accurate delivery mode in advance would allow identifying which pregnant women truly need operative procedures or caesarean section, reducing the proportion of unnecessary assisted procedures used liberally with marginal medical benefit [4]. Furthermore, by supporting the physician in decision-making process, the healthcare service would be able to avoid malpractice and negligence, leading to the quality improvement in maternity care and providing a better care to the pregnant.

The purpose of this study is to induce Data Mining (DM) classification models in real-time able to predict the type of delivery more compatible with the pregnancy characteristics of each patient. The required information is provided by the information systems and technologies used in the perinatal and maternity care unit (CMIN) of Centro Hospitalar of Oporto (CHP). In CMIN, Centro Materno Infantil do Norte, clinical information, patient data and their admission form are produced and record through SAPE, a decision support system focused on nursing practices, and EHR, the Electronic Health Record system implemented by Archive and Diffusion of Medical Information (AIDA). Thus, through this information it becomes possible to extrapolate useful knowledge to support medical decision-making in context of child delivery. This knowledge is achieved by means of Data Mining (DM) techniques, enabling predictive models based on evidence. Several DM models were explored, acquiring specificity and sensitivity values of approximately 80.05% and 90.11%, respectively.

This article includes six sections. After the introduction, the second section presents the concepts and related work. The methodologies, materials and methods used in this work are described in section three. Section four describes all steps of knowledge discovery following Cross Industry Standard Process for Data Mining. The results are discussed and a set of considerations are made in section five. The conclusions and future work compose the last section.

2. Background and Related Work

2.1. Types of Birth Delivery

Conceptually, a vaginal birth refers to the natural method of birth that requires no assistance. In turn, when the pregnancy conditions require additional support, operative vaginal delivery techniques, forceps and vacuum, are demanded. They are suitable for cord prolapse, exhaustion and certain neurological and heart conditions of the mother [1] [5]. Caesarean section is the surgical alternative to the latter. There is conflicting evidence for and against the use of these procedures [2]. Caesarean section has been associated with endometritis, the need for transfusion and pneumonia in addition to placental problems [6] [7]. The risk of maternal injury and severe lacerations, as well as facial nerve palsies is presented in forceps delivery. Moreover, vacuum delivery can lead to cephalohematomas and retinal or intracranial hemorrhages in the neonate [8] [9]. Serious complications are rare for both methods but can lead to neonatal death, thus, those procedures should only be performed if there is an appropriate indication [2] [10]. Physicians often make decisions based on the marginal expected benefit of the intervention, making the process uncertain, leading unnecessarily to the mentioned complications in the mother and child's health [3].

Some studies to identify which pregnancy characteristics can identify the correct method of delivery have been conducted. Lee, K. and Gay, C. conducted a study in 2004 [11] testing the hypothesis that fatigue and sleep disturbance in late pregnancy are associated with the type of delivery and labor duration, concluding that women with sleep deprivation are more likely to have caesarean deliveries. Identifying the pregnancy attributes that influence the type of delivery can also allow avoiding operative deliveries by providing health measures to control the risk factors.

2.2. CMIN and the Interoperability of its Systems

As mentioned in the previous section, this study is based on real data acquired from the maternal and perinatal

care unit of CHP, named CMIN, through its information systems; Support Nursing Practice System (SAPE) and Electronic Health Registration (EHR).

CMIN stand as one of the four constituent hospitals of CHP, being prepared to provide neonatology, obstetrics and gynecology services and accompany the women and child conditions since the early pregnancy until the first stages of child growth [12]. The SAPE resulted from the Information Systems in Nursing and records clinical episodes associated with each patient as an alternative to the traditional way of information on paper [13]. EHR is a system employed by hospitals to store and retrieve detailed patient information, as the admission form, helping monitoring, improving and reporting data on health care quality and safety [14] [15]. SAPE and EHR are able to provide the desired information since they are connected by the Agency for Integration, Archive and Diffusion of Medical Information (AIDA) that allows the interoperability of the hospital existing systems. AIDA is a platform based on the use of intelligent agents. This multi-agent system enables the standardization of clinical systems and overcomes the medical and administrative complexity of the different sources of information from the hospital [16].

2.3. Knowledge Discovery and Data Mining Models in Healthcare

The Knowledge Discovery from Databases (KDD) process is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [17] [18]. The KDD process is a set of five steps, beginning with the selection of the dataset. The second step includes cleaning and processing the data, making it consistent. Then, follows the transformation step, where the data is transformed according to the study goal. In the fourth step the data mining models (DM) are induced, and finally, remains the interpretation and evaluation of the patterns attained [17]. While the KDD process refers to the whole process of discovering useful knowledge in data, DM refers to the application of algorithms to extract data models [19].

Data mining is becoming increasingly popular and essential to healthcare organizations, benefiting different health services, through many applications since identifying effective treatments and best practices until providing quality healthcare to the patients [20] [21]. Data mining techniques remain a straightforward tool, finding useful patterns in complex and voluminous data transforming it into useful information for decision making [17] [21].

In obstetrics and maternal care there have been many studies that apply data mining techniques to identify solutions for health services limitations. Some of those studies included early prediction of later preeclampsia, a pregnancy-specific syndrome that causes substantial maternal and fetal morbidity [22], and predicting risk pregnancy in women performing voluntary interruption of pregnancy (VIP) [23].

To predict the type of birth through pregnancy characteristics, DM classification algorithms are persuaded. The classification algorithms process a training set containing a group of attributes in order to discover relationships between the attributes that would make it possible to predict the outcome [24].

3. Study Description

This study was conducted by following the KDD process introduced in section 2.3 and inducing data mining models (DM) through classification algorithms. The DM methodology employed was the Cross Industry Standard Process for Data Mining (CRISP-DM), a non-rigid sequence of six phases, which allow the implementation of a DM model to be used in a real environment, helping to support business decision, increasing the success of DM projects [25]. These steps are [26]: Business Understanding; Data Understanding; Data Preparation; Modelling; Evaluation and Deployment. To induce the DM models, four different techniques were implemented: Decision Trees (DT), Generalized Linear Models (GLM), Support Vector Machine (SVM) and Naïve Bayes (NB), for their feasibility in predicting the most likely outcome in this kind of study. This study considers pregnant patients' data collected from 4236 clinical episodes occurred in the maternal and perinatal care unit (CMIN) of Centro Hospitalar of Oporto (CHP), comprising a period between 2012-07-01 and 2015-01-31, a total of 1120 days.

4. Data Mining Process

To pursue the knowledge discovery, the Data Mining process followed the CRISP-DM methodology, mentioned in the previous section. Following, all phases of that process are described.

4.1. Business Understanding

Essentially, the business goal of this study involves the prediction of the most appropriate delivery mode to be performed in a patient about to give birth, given her obstetric conditions. The sensitivity required in the matter and the complications linked with some delivery procedures were the incentive to conduct this study.

Thereafter, the Data Mining (DM) goal is to develop accurate and sensitive models able to predict the type of child delivery built on data from CMIN clinic cases. These models have to show satisfactory statistical metrics as sensitivity and specificity, in order to support the physicians' decision-making process efficiently.

4.2. Data Understanding

From SAPE and EHR records, the pregnancy characteristics that could impact the type of delivery were identified and integrated in the study dataset. Thus, each data instance consists of a set of 26 variables: age, allergies, programmed (whether or not a delivery is programmed), motive, vigilance, gestation (singular or multiple pregnancy), grav170 (normal delivery or unexpected events), weight, height, body mass index (BMI), blood pressure (HBP and LBP), number of gestation weeks, marital status, blood type, a group of medical exams; - cardiotocography (CTG), RFC, streptococcus, rhesus - and specific end pregnancy characteristics - fetus weight, dilatation, consistency, extinction, position, Bishop score and Hodge plan.

Statistical measures related to the numerical variables are represented in table 1. For a better understanding of the dataset, table 2 shows the percentage of occurrences associated with the classes of some used variables.

Table 1. Statistical measures of the numerical variables used in the dataset.

Variables	Minimum	Maximum	Average	Standard Deviation
<i>Age</i>	15	48	30.34	5.74
<i>Weight (kg)</i>	37	154	78.15	13.75
<i>Height (m)</i>	1.16	1.86	1.63	6.05
<i>BMI (kg/m²)</i>	14.33	51.36	29.48	4.78
<i>LBP (mm Hg)</i>	39	116	78.20	8.29
<i>HBP (mm Hg)</i>	91	200	122.17	10.11
<i>Gestation Weeks</i>	22	41	38.20	2.25
<i>Fetus weight (kg)</i>	0.56	4.48	3.05	0.57

Table 2. Percentage of occurrence of some variables' values.

Variable	Class	Percentage
<i>Programmed</i>	True	15.76%
<i>Vigilance</i>	Not monitored	5.95%
	Insufficient monitored	1.32%
	Monitored	92.73%
<i>Gestation</i>	Singular	96.70%
	Multiple	3.30%
<i>Grav170</i>	Normal	79.97%
	Intercurrences	20.03%
<i>Streptococcus</i>	Negative	78.55%
	Positive	12.81%
	Not realized	8.64%
<i>Cardiotocography</i>	Pathologic	0.25%
	Suspect	1.78%
	Normal	97.97%
<i>Rhesus</i>	Positive	78.09%
<i>RFC</i>	Positive	78.39%
<i>Consistency</i>	Soft	37.26%
	Medium	56.23%
	Hard	6.51%
<i>Dilatation</i>	0	12.10%
	1-2	38.69%
	3-4	41.54%

	≥ 5	7.68%
<i>Extinction</i>	0-30	32.49%
	40-50	21.20%
	60-70	28.11%
	80-10	18.20%
<i>Position</i>	Inferior	11.29%
	Intermediate	57.80%
	Posterior	30.91%

The target variable *Type of Delivery* represents the four delivery methods; normal, caesarean, forceps and vacuum extraction. Figure 1 shows the data distribution in terms of the target classes, as can be observed more than 70% of the delivery are normal or caesarean, being the values approximated.

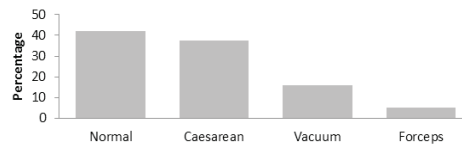


Fig. 1. Distribution of the target variable *Type of Delivery*.

4.3. Data Preparation

After selecting the data exposed in section 4.2, a pre-processing phase was performed, deleting all data with null or noise values, leaving 4236 records to be used by the DM models. Some normalization was also required since some of the registered variables have some inconsistent values. For instance, the patient's weight is recorded in a free writing field, enduring no unit of measure. To ensure the DM performance, all the weight values were transformed to kilograms, using the point to separate the decimal fraction. The patient age was acquired through the transformation of initial data; the birth and clinical episode dates. This phase can be performed multiple times, without order restrictions. In this case study, after the initial DM models induction, it was necessary to transform the dataset once more, in order to improve the results. One of these changes was applying oversampling to the data, a technique used to balance all the exits of the target variable, by replicating the instances of the less used delivery modes, forceps and vacuum, until the percentage of occurrences of each birth type was similar. In view of the evolving DM process, it was also required to study the dataset with different displays of the target variable, in order to find improved outcomes. Table 3 shows all the attempted approaches.

Table 3. Different organizations of the target variable

<i>Target / Approach</i>	<i>Number of Target Exits</i>	<i>Target Classes</i>
1 (initial)	4	{Normal}, {Caesarean}, {Forceps}, {Vacuum}
2 (normal versus instrumented)	2	{Normal}, {Caesarean, Forceps, Vacuum}
3 (most frequent versus less used)	2	{Normal, Caesarean}, {Forceps, Vacuum}
4 (caesarean versus vaginal)	2	{Caesarean}, {Normal, Forceps, Vacuum}

4.4. Modeling

The data mining models (DMM) were induced from the processed and transformed data, using the four DM techniques (DMT); GLM, SVM, DT and NB. For each DM algorithm, two sampling methods (SM) were applied: Holdout sampling, where 30% of the data is used for testing, and Cross Validation, with all data for testing. The DM models were also induced using the dataset with unbalanced target exits and with oversampled data, making two different data approaches. Finally, different variables' combinations were performed to identify which pregnancy characteristics stand for obstetric risk factors in predicting the type of delivery. This variables combination allowed the development of scenarios through a clinical study, conducted in the maternity unit. The considered scenarios are:

S1: {All variables}

S2: {Age, Height, Weight, BMI, Consistency, Dilatation, Extinction, Gestation, Motive, CTG, Bishop Score, Hodge Plan, Position, RFC, Rhesus, Streptococcus, Gestation Weeks, LBP, HBP}

S3: {Consistency, Dilatation, Extinction, Gestation, CTG, Bishop Score, Hodge Plan, Position, RFC, Rhesus, Streptococcus, Gestation Weeks, LBP, HBP}

S4: {Grav170, Consistency, Dilatation, Extinction, Gestation, Bishop Score, Hodge Plan, Position, Fetus Weight, Gestation Weeks}

S5: {Age, Grav170, Blood Type, Height, Weight, BMI, CTG, RFC, Rhesus, Streptococcus, Gestation, Motive, Programmed, Vigilance, LBP, HBP}

Hereupon, each DMM can be described by equation 1.

$$DMM_n = A_f \times S_i \times DMT_y \times SM_c \times DA_b \times TG_t \quad (1)$$

The model DMM_n belongs to the approach (A) and is composed by a scenario (S), a sampling method (SM), a data approach (DA), a data mining technique (DMT) and a target (TG):

$A_f = \{Classification\}$

$S_i = \{Scenario 1 \dots Scenario 5\}$

$DMT_y = \{SVM, GLM, DT, NB\}$

$SM_c = \{Holdout Sampling, Cross Validation\}$

$DA_b = \{With Oversampling, Without Oversampling\}$

$TG_t = \{initial, normal versus instrumented, most frequent versus less used, caesarean versus vaginal\}$

Thus, a total of 80 models were induced for the initial target distribution. The same models were used for the three other target approaches (described in table 3), when its application pertinent (for instance, for target 2, oversampling was not required). For all the four targets they were induced 280 models.

$DMM = \{5 \text{ Scenarios}, 4 \text{ Techniques}, 2 \text{ Sampling Methods}, 2 \text{ Data Approaches}, 4 \text{ Target}\}$

All models were induced using the Oracle Data Miner with the standard configurations concerning the DM classification algorithms.

4.5. Evaluation

In order to evaluate the induced models, the statistic metrics described in equations 2, 3 and 4 were applied. These metrics are estimated through the results provided by the confusion matrix (CMX) of each model, provided by the data miner:

$$Sensitivity = TP / (TP + FN) \quad (2)$$

$$Specificity = TN / (TN + FP) \quad (3)$$

$$Accuracy = TP / (TP + FT + TN + FN) \quad (4)$$

CMX obtains four types of results: the number of True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN). In the first target approach there is one target exit for each delivery mode (four exits). Since the target is not binary, it only makes sense to consider accuracy to evaluate the models. Table 4 shows the best results in view of this approach.

Table 4. Best accuracy results achieved in the first target approach.

Scenario	Model	Approach	Sampling	Accuracy
3	DT	Without oversampling	30%	0.5081
3	DT	Without oversampling	All	0.5057
4	DT	Without oversampling	All	0.5068

From a clinical point of view, these results are redundant, requiring the recurring to the preparation phase to redistribute the target variable in binary combinations, so that improved outcomes were reached. The best results attained by the induced models for each target approach declared in the preparation phase are presented in table 5.

Table 5. Best sensitivity, specificity and accuracy results achieved with the three aided target approaches.

Target Approach	Scenario	Model	Approach	Sampling	Accuracy	Sensitivity	Specificity
2	2	DT	Without oversampling	30%	0.7200	0.6660	0.7599
3	1	GLM	Without oversampling	All	0.5857	0.8904	0.2914
	2	GLM	Without oversampling	All	0.5899	0.9110	0.3031
	2	SVM	Without oversampling	All	0.6206	0.8557	0.2792
4	2	DT	Without oversampling	All	0.8391	0.8828	0.8005
	2	GLM	Without oversampling	All	0.7741	0.8427	0.6755
	1	NB	Without oversampling	All	0.7469	0.8429	0.6298

The best achieved accuracy and specificity, 83.91% and 80.05% respectively, were accomplished by the fourth target distribution (0 being caesarean section and 1 vaginal delivery) with scenario 2 induced by DT technique, without oversampling and all data for testing. The best sensitivity (91.11%) was also achieved by scenario 2, without oversampling and using all data for testing, using the GML technique, with the target separating the normal deliveries and the ones that need assistance.

4.6. Deployment

After the evaluation process, the best induced models and the achieved knowledge are reported to the maternity care unit, in order to assist physicians in the decision-making process regarding the delivery mode of the pregnant patients. In this case, these DMM will be included in the Business Intelligence platform [27], already deployed in CMIN, assisting the physician to accurately choose the type of delivery to which a new case belongs.

5. Discussion

Analyzing the best results was possible to observe that the oversampling approach did not had a positive impact in the models induction. On the other hand, using Cross Validation as the sampling method improved the metric results. Concerning the different target distributions attempted in this study, it was visible that the fourth approach, which isolates the caesarean sections on a single exit, holds the best DM models with superior accuracy results. These results are inflicted by some variables, like dilatation and consistency that present very specific values in case of caesarean section, making its classification easier. In this approach, the specificity shows superior outcomes, showing its ability predicting the exit 0; caesarean sections. In contrast, if the most pertinent model is the one most sensitive to identify the normal deliveries, the best approach would be the third one, which shows lower accuracy and specificity, but excellent sensitivity to predict the exit 1, normal deliveries. The majority of the best models represented in table 5 show that the obstetric factors that mostly influence the type of delivery to be performed are the ones fulfilling scenario 2. In order to help choosing the best model, a threshold was considered, filtering the models with specificity, sensitivity and accuracy values higher than 80%, ensuring the marginal quality of results to be used in a clinical environment. Therefore, the best induced model is scenario 2, considering the DT algorithm, without oversampling and all data for testing, when targeting caesarean sections and vaginal deliveries.

Considering the study business aim, it was not possible to find a model to predict the most appropriate delivery mode, separately, but accomplishing a model that can successfully identify caesarean sections and vaginal deliveries is able to help the medical and administrative decision making concerning the maternity care services, allowing better monitoring and quality services. Being these work focused in the pregnant and their delivery, these results give to them some confidence in the clinical work, because the physicians has now extra knowledge to help them to know which type of deliveries are more adequate.

6. Conclusions and Future Work

Using real data from CMIN, it was possible to prove the viability of using DM models to predict which type of delivery should be pursued, through the pregnancy characteristics of the pregnant patients. Satisfactory results were achieved regarding statistic metrics, by inducing Decision Tree algorithm and using all data for testing through the characteristics composing scenario 2, achieving approximately 88% of sensitivity, 80% of specificity and 84% of accuracy, allowing the prediction of caesarean sections and vaginal deliveries. The best model induced and the results achieved will be included in the Business Intelligence platform [27], already employed in CMIN, supporting the physicians in decision-making and the healthcare unit in avoiding malpractice and negligence regarding child delivery, leading to quality improvements in maternity care. Future work will incorporate new variables in the predictive models and other types of data mining techniques will be applied. For instance, inducing Clustering techniques would allow creating groups relating specific pregnancy factors with types of delivery.

Acknowledgements

This work has been supported by FCT - Fundação para a Ciência e Tecnologia within the Project Scope UID/CEC/00319/2013.

References

- Demissie, K., Rhoads, G. G., Smulian, J. C., Balasubramanian, B. A., Gandhi, K., Joseph, K. S., and Kramer, M. Operative vaginal delivery and neonatal and infant adverse outcomes: population based retrospective analysis. *Bmj*, 329(7456), 24. (2004)
- Cargill, Y. M., MacKinnon, C. J., Arseneault, M. Y., Bartellas, E., Daniels, S., Gleason, T., and Wilson, A. K. Guidelines for operative vaginal birth. *Journal of obstetrics and gynaecology Canada: Journal d'obstetrique et gynecologie du Canada: JOGC*, 26(8), 747-761. (2004)
- Sakala, C., and Corry, M. Evidence-based maternity care: What it is and what it can achieve. *Milbank Memorial Fund*. (2008)
- Simpson, K. R., and Thorman, K. E. Obstetric “conveniences”: elective induction of labor, cesarean birth on demand, and other potentially unnecessary interventions. *The Journal of perinatal and neonatal nursing*, 19(2), 134-144. (2005)
- Cunningham, F. G., MacDonald, P. C., Gant, N. F., Leveno, K. J., and Gilstrap, L. C. Conduct of normal labor and delivery. *Williams obstetrics*, 1, 409-443. (2001)
- Burrows, L. J., Meyn, L. A., and Weber, A. M. Maternal morbidity associated with vaginal versus cesarean delivery. *Obstetrics and Gynecology*, 103(5, Part 1), 907-912. (2004)
- Hemminki, E., and Meriläinen, J. Long-term effects of cesarean sections: ectopic pregnancies and placental problems. *American journal of obstetrics and gynecology*, 174(5), 1569-1574. (1996)
- Johanson, R. B., and Menon, V. Vacuum extraction vs. forceps for assisted vaginal delivery. *The Cochrane Library*, (1). (2003)
- Towner, D., Castro, M. A., Eby-Wilkens, E., and Gilbert, W. M. Effect of mode of delivery in nulliparous women on neonatal intracranial injury. *New England journal of medicine*, 341(23), 1709-1714. (1999)
- Ali, U. A., and Norwitz, E. R. Vacuum-assisted vaginal delivery. *Reviews in Obstetrics and Gynecology*, 2(1), 5. (2009)
- Lee, K. A., and Gay, C. L. Sleep in late pregnancy predicts length of labor and type of delivery. *American Journal of Obstetrics and Gynecology*, 191(6), 2041-2046. (2004)
- Centro Hospitalar do Porto. Retrieved April 16, 2015, from <http://www.chporto.pt/index.php?und=ch>
- Brandão, A., Pereira, E., Portela, F., Santos, M. F., Abelha, A., Machado, J.: Managing voluntary interruption of pregnancy using data mining. *Procedia Technology*, 16, 1297-1306. (2014)
- Portela, F., Cabral, A., Abelha, A., Salazar, M., Quintas, C., Machado, J., Santos, M. F. Knowledge Acquisition Process for Intelligent Decision Support in Critical Health Care. *Healthcare Administration: Concepts, Methodologies, Tools, and Applications*, 270. (2014)
- Silow-Carroll, S., Edwards, J. N., Rodin, D. Using electronic health records to improve quality and efficiency: the experiences of leading hospitals. *Issue Brief (Commonw Fund)*, 17, 1-40. (2012)
- Abelha, A., Analide, C., Machado, J., Neves, J., Santos, M., and Novais, P. Ambient intelligence and simulation in health care virtual scenarios. In *Establishing the Foundation of Collaborative Networks* (pp. 461-468). Springer US. (2007)
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37. (1996)
- DeGruy, K. B. Healthcare applications of knowledge discovery in databases. *Journal of Healthcare Information Management*, 59-70. (2000)
- Maimon, O., and Rokach, L. Introduction to knowledge discovery in databases. In *DMKD Handbook* (pp. 1-17). Springer US. (2005)
- Koh, H. C., and Tan, G. Data mining applications in healthcare. *Journal of healthcare information management*, 19(2), 65. (2011)
- Obenshain, M. K. Application of data mining techniques to healthcare data. *Infection Control*, 25(08), 690-695. (2004)
- Kenny, L. C., Broadhurst, D. I., Dunn, W., Brown, M., North, R. A., McCowan, L., and Baker, P. N. Robust early pregnancy prediction of later preeclampsia using metabolomic biomarkers. *Hypertension*, 56(4), 741-749. (2010)
- Brandão, A., Pereira, E., Portela, F., Santos, M. F., Abelha, A., and Machado, J. Predicting the risk associated to pregnancy using data mining. *ICAART 2015 - 7th International Conference on Agents and Artificial Intelligence*. Portugal. SciTePress. (2015)

24. Mendes, R. R., de Voznika, F. B., Freitas, A. A., and Nievola, J. C. Discovering fuzzy classification rules with genetic programming and co-evolution. In *Principles of Data Mining and Knowledge Discovery* (pp. 314-325). Springer Berlin Heidelberg. (2001)
25. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. CRISP-DM Step-by-step data mining guide. (2000)
26. Gonçalves, J. M., Portela, F., Santos, M. F., Silva, Á., Machado, J., and Abelha, A. Predict Sepsis Level in Intensive Medicine–Data Mining Approach. In *Advances in Information Systems and Technologies* (pp. 201-211). Springer Berlin Heidelberg. (2013)
27. Eliana Pereira, Andreia Brandão, Filipe Portela, Manuel Filipe Santos, José Machado, António Abelha. Business Intelligence in Maternity Care. *IDEAS 2014 - International Database Engineering & Applications Symposium*. Porto, Portugal. ACM. (2014)