Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 113 (2017) 571–576

Procedia Computer Science

www.elsevier.com/locate/procedia

International Workshop on Healthcare Interoperability and Pervasive Intelligent Systems (HiPIS 2017)

Predicting the need of Neonatal Resuscitation using Data Mining

Ana Morais[a], Hugo Peixoto[b], Cecília Coimbra[b], António Abelha[b] and José Machado[b]*

[a]*University of Minho, Campus Gualtar, Braga 4710, Portugal*
[b]*Algoritmi Research Center, University of Minho, Campus Gualtar, Braga 4710, Portugal*

Abstract

It is estimated that approximately 10% of newborns require some kind of assistance for breathing at birth. Aiming to prevent neonatal mortality, the goal behind this paper is to predict the need for neonatal resuscitation given some health conditions of both the newborn and the mother, and also the characteristics of the pregnancy and the delivery using Data Mining (DM) models induced with classification techniques. During the DM process, the CRISP-DM Methodology was followed and the WEKA software tool was used to induce the DM models. For some models, it was possible to achieve sensitivity results higher than 90% and specificity and accuracy results superior to 98%, which were considered to be satisfactory.

Peer-review under responsibility of the Conference Program Chairs.

Keywords: Data Mining; Decision Support Systems; CRISP-DM; Classification; Neonatal Resucitation

1. Introduction

Nowadays, information systems are essential for organizations since they provide useful information for decision-making processes by storing, processing and analysing large amounts of data[1]. DM is considered as the set of methods and techniques for exploring and analyzing large datasets in an automatic form with the aim of finding unknown or hidden rules, associations or patterns. The DM techniques can be classified as descriptive, which includes clustering techniques, or predictive, which includes classification and regression techniques. This paper is focused on classification techniques, in which the model or classifier predicts categorical labels[2].

* Corresponding author. Tel.: +351 253 604 430; fax: +351 253 604 471.
E-mail address: jmac@di.uminho.pt

1877-0509
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2017.08.287

Due to the large amount and the complex nature of data generated by transactions in healthcare environments, the interest of healthcare organizations on using DM has been increasing since it can greatly benefit all parties involved. Namely, the analysis provided by DM allows healthcare institutions to improve operating efficiency while maintaining a high level of quality of care. However, DM's applications in healthcare have limitations such as the often difficult accessibility to data, which can be due to the existence of raw inputs in different settings and systems. Thus, the data needs to be collected and integrated before it can be used in DM[3].

Decision Support Systems (DSS) can be defined as the class of computer-based information systems that support decision making activities[4]. In order to provide quick and reliable decision support, DSS require the automated analysis of data to find tendencies and extract knowledge, which can be performed by DM techniques[5]. When applied to healthcare, DSS are called Clinical DSS and offer support to clinicians at the various stages of the care process. They can also take over some routine tasks by warning clinicians of potential problems or providing suggestions for them to consider and provide clinicians or even patients with knowledge and person-specific information that can be intelligently filtered and presented at appropriate times. In addition, other advantages of the usage of Clinical DSS in medical practice include ensuring accurate and timely diagnoses for preventing diseases, lowering operating costs, improving efficiency and reducing patient inconvenience[6].

2. Background and Related Work

Pereira et al. (2015) proved the viability of using DM models to predict the most appropriate type of delivery considering the pregnancy characteristics of patients[7]. Namely, real data from the perinatal and maternal care unit of Centro Hospitalar of Oporto (CHP) was used and four different classification techniques were implemented: Decision Trees (DT), Generalized Linear Models (GLM), Support Vector Machines (SVM) and Naïve Bayes (NB). The best induced model acquired satisfactory results by achieving sensitivity values around 90% and was afterwards included in the Business Intelligence platform already employed in CHP[7].

Portela et. al (2015) tested the usage of DM classification techniques, namely GLM, SVM, DT and NB, to predict the probability of a patient to have a blood pressure critical event in the following hours by combining a set of patient data extracted in real-time from CHP. The achieved results demonstrated to be quite promising, with sensitivity values around 95% and the best induced model was afterward included in the INTCare ensemble engine, representing an important step in helping the prevention of possible cases of hypertension or hypotension[8].

Neonatal resuscitation is defined as the resuscitation of newborns with birth asphyxia and its goal is to re-establish adequate spontaneous respiration and cardiac output and to prevent neonatal mortality[9].It is estimated that approximately 10% of newborns require some degree of assistance to start breathing at birth. The risk factors for these occurrences include malpresentation of the fetus, gestational age lower than 35 weeks, low birth weight infants, increased or decreased maternal age, multiple birth pregnancy, among others[10]. An accurate evaluation of the risk factors can lead to the anticipation of resuscitation need which, in turn, allows adequate preparation of the necessary equipment and staff to perform neonatal resuscitation. In fact, how quickly and successfully the resuscitation is performed can be truly decisive for the infant's health, namely for avoiding hypoxic damage on the organs or even brain damage[11]. Thus, having a decision-support system that can accurately predict the need for resuscitation would be of great interest for both patients and health professionals. Indeed, obstetricians could know beforehand whether the newborn needs resuscitation and could perform all the necessary procedures right after the newborn's birth, improving the efficiency of the provided care and reducing medical errors.

3. Methodologies, Material and Methods

The data presented in this work was extracted from Electronic Health Records (EHR) and admission records from the obstetrics service of a Portuguese hospital. This data comprises the year of 2016 and has information about 3163 newborns, along with information about their mothers and the respective delivery episodes. During the Data Mining Process, the Cross Industry Standard Process for Data Mining (CRISP-DM) Methodology was followed, which is a hierarchical process model that divides the process of data mining into six phases: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation* and *Deployment*[12]. This methodology was followed due to its advantages such as increasing the success of DM projects and allowing the implementation of DM models in real environments[13].

3.1. Business Understanding

The business goal of the work presented in this paper is the prediction of the need for neonatal resuscitation in a newborn baby, considering the baby's characteristics and also given the type of delivery, the characteristics of the pregnancy and the health conditions of the mother. This prediction must be highly sensitive and accurate since it can be decisive for the life of the newborn. In addition, predicting beforehand that a newborn will need resuscitation can help obstetricians manage their time and efforts better and therefore deliver more effective care to newborns.

3.2. Data Understanding

Each data instance consists of a set of 12 variables: type of delivery, baby's weight, height and cephalic perimeter, presentation of fetus, gestational age in weeks, risk classification of the pregnancy, age and BMI of mother, twins (whether it is a twin pregnancy or not), Robson classification and resuscitation.

The target variable *Resuscitation* represents whether the newborn needed resuscitation and has two possible values: *yes* or *no*. Figure 1 shows the data distribution of this variable on the used dataset and as it can be observed, 11.1% of the registered newborns needed neonatal resuscitation.

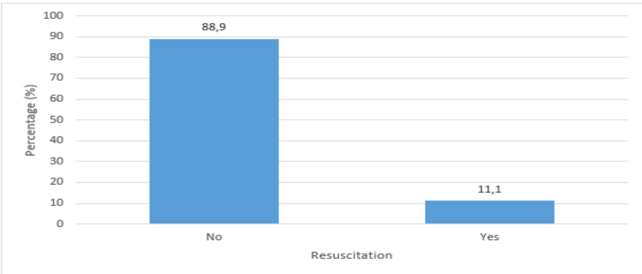

Fig. 1. Data distribution of the target variable *Resuscitation*

3.3. Data Preparation

This phase of the DM process involved the selection and preparation of the data to be used by the DM models. Firstly, it was necessary to eliminate all data with null or noise values to ensure that there was no incomplete or inconsistent information. At the end of this cleaning process, there were 2171 records left on the dataset. There was also the need to convert all the instances of some variables to the same units: all the BMI values were converted to kg/m^2, all the newborn's weight values were converted to grams (g) and the newborn's height and cephalic perimeter values were all both converted to centimeters (cm). To improve the performance of the DM models, all the data, except for the target variable, was converted to the numerical type. After that, the data was normalized in order to have its values between 0 and 1. To evaluate the effect of oversampling the data on the performance of the DM models, an additional dataset with 4342 instances was created by duplicating all the instances from the initial dataset.

3.4. Modeling

This phase consisted of inducing the Data Mining Models (DMM) in *Weka* using the prepared data. Since Classification was the chosen Approach (A), there were 5 different DM techniques (DMT) that were used: *Logistic Regression* (LR), *Naïve Bayes* (NB), *k-Nearest Neighbors* (kNN), *Random Forest* (RF) and *Support Vector Machines* (SVM). These algorithms were used with the standard configurations in Weka.

For each DM technique, two sampling methods (SM) were tested: *Holdout sampling*, with 66% of the data used for training and the remaining amount for testing, and *Cross Validation*, using 10 folds and where all data is used for testing. In addition, there were two data approaches (DA) tested: *Without Oversampling* or *With Oversampling* of all instances. There was only one target variable, which was the *Resuscitation* variable, and the considered scenarios, in order to evaluate which attributes were the most relevant to predict the need for performing neonatal resuscitation, were: S1: {All attributes}; S2: {Weight, Cephalic Perimeter, Height, Gestational Age, Type of Delivery, Risk Classification, Robson Classification, Presentation of Fetus, Twin}; S3: {Weight, Cephalic Perimeter, Height, Gestational Age, Risk Classification, BMI of Mother, Age of Mother} and S4: {Weight, Cephalic Perimeter, Height, Gestational Age}.

Each DMM can be described as belonging to an approach (A), being composed by a scenario (S), a data mining technique (DMT), a sampling method (SM), a data approach (DA) and a target (T): $DMM_n = \{A_f, S_i, DMT_y, SM_c, DA_b, TG_j\}$. For this work the DMM was described by $A_f = \{Classification\}$; $S_i = \{S1, S2, S3, S4\}$; $DMT_y = \{LR, NB, KNN, RF, SVM\}$; $SM_c = \{Holdout\ Sampling, Cross\ Validation\}$; $DA_b = \{Without\ Oversampling, With\ Oversampling\}$; $TG_j = \{Resuscitation\}$. In total, 80 models were induced using DMM = {1 Approach, 5 Scenarios, 5 DM Techniques, 2 Sampling Methods, 2 Data Approaches, 1 Target}.

3.5. Evaluation

The performance of each DMM was assessed through its confusion matrix (CMX), which presents the number of True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN). With these results, it is possible to calculate sensitivity, specificity and accuracy in order to evaluate the algorithms performance. Tables 1, 2, 3 present the models that, for each DM technique, achieved the best sensitivity, specificity and accuracy results, respectively.

Table 1. DM models with the best sensitivity results for each DM technique

DM Technique	Scenario	Sampling Method	Data Approach	Sensitivity
LB	S2	Holdout Sampling	Without Oversampling	25.926%
NB	S3	Holdout Sampling	Without Oversampling	45.679%
KNN	S1, S3	Cross Validation	With Oversampling	90.871%
RF	S2	Cross Validation	With Oversampling	90.456%
SVM	-	-	-	0%

Table 2. DM models with the best specificity results for each DM technique

DM Technique	Scenario	Sampling Method	Data Approach	Specificity
LB	S4	Holdout Sampling	With Oversampling	99.715%
NB	S3, S4	Holdout Sampling	Without Oversampling	95.738%
KNN	S3	Cross Validation	With Oversampling	99.430%
RF	S3	Cross Validation	With Oversampling	100%
SVM	-	-	-	100%

Table 3. DM models with the best accuracy results for each DM technique

DM Technique	Scenario	Sampling Method	Data Approach	Accuracy
LB	S2, S4	Holdout Sampling	Without Oversampling	91.192%
NB	S3	Holdout Sampling	Without Oversampling	90.244%
KNN	S2	Cross Validation	With Oversampling	98.480%
RF	S1	Cross Validation	With Oversampling	98.895%
SVM	-	-	-	89.024%

As it can be observed, overall the best achieved sensitivity, specificity and accuracy values were 90.871%, 100% and 98.895%, respectively. In order to choose the most suitable model, a threshold was established and the models were ranked according to their sensitivity results. The defined threshold was sensitivity >90% and specificity and accuracy >85%. Table 4 presents the best three models that achieved the threshold by their ranking order.

Table 4. DM models with the best accuracy results for each DM technique

DM Technique	Scenario	Sampling Method	Data Approach	Sensitivity	Specificity	Accuracy
KNN	S3	Cross Validation	With Oversampling	90.871%	99.430%	98.480%
KNN	S1	Cross Validation	With Oversampling	90.871%	99.378%	98.434%
RF	S3	Cross Validation	With Oversampling	90.456%	99.845%	98.802%

It is observed that all the best three models used Cross Validation as the sampling method and used the data with oversampling. It can also be seen that there was a tie on the sensitivity values between the first and the second ranked models. Despite this tie, the first ranked model was, from these two, the one with the highest accuracy value. Thus, it is possible to claim that the most suitable model, from all the 80 induced models, is DMM = {Classification, S3, KNN, Cross Validation, With Oversampling, TG1}.

4. Discussion

The analysis of the obtained results allowed to conclude that using oversampling made it possible to obtain sensitivity values higher than 90% for some models, which was not possible without oversampling. Models which used *Cross Validation* as the sampling method achieved better results when compared to those which used *Holdout Sampling*. The reason behind this is that the *Cross Validation* method uses all the data for training, while *Holdout Sampling* only uses a certain percentage, and algorithms learn more effectively when they use more data for training. This difference and the influence of oversampling were noticeable on the results due to the relatively low number of instances on the used dataset. It could be observed, for the models that achieved the best sensitivity value that the choice between the first and the third scenario did not have an impact on the sensitivity result, which might mean that the risk factors of the mother can, indeed, be determinant for the need of neonatal resuscitation. It was also verified that the algorithms that had the best results were KNN and RF and that the technique with the worst results was SVM, with sensitivity values of 0% on all models, which means that this technique is not suitable at all for the used dataset. The NB technique did not have a great performance as well, which might be due to the assumption this technique makes about the independence of the variables. This assumption might have not worked well on the used dataset due to the high correlation that exists between the attributes. All the induced models achieved specificity results higher than 89%. Since newborns that did not need resuscitation represent the majority of newborns registered on the used dataset, it was relatively easy for the algorithms to correctly classify these cases. Similarly, the accuracy results were higher than 85% on all the induced models. Thus, neither specificity nor accuracy were considered the most relevant criteria when choosing the most suitable model, but sensitivity was. The reason behind this consideration is the fact that, in this particular case, it is preferable to predict that a newborn needs resuscitation when that is not true than to predict that resuscitation is not needed when, in fact, it is. In addition, the relatively low number of newborns that needed resuscitation registered on the dataset made it difficult for algorithms to correctly identify TP. Therefore, and due to the critical nature of this problem, FN must be avoided at all costs and models that can best identify TP, and which consequently can achieve better sensitivity results, should be prioritized. In order to choose the best model, a threshold was defined to filter the models that could ensure results with enough quality to be used in a clinical environment. Therefore, the model with the highest sensitivity value was considered the most suitable, which was the model that considered the third scenario, used the KNN technique, the *Cross Validation* sampling method and used the data with oversampling.

5. Conclusions and Future Work

This work proved that, using real data from EHR and admission records from the obstetrics service, it is viable to use DM models to predict the need for neonatal resuscitation given some health conditions of both the newborn and the mother, and the characteristics of the pregnancy and the delivery. It was possible, for some DM models, to achieve sensitivity results higher than 90% and specificity and accuracy results higher than 98%, which are considered to be quite satisfactory. The best induced model, which achieved a sensitivity value of 90.871% and 98.480% of accuracy, used the KNN algorithm, the third scenario of attributes and the *Cross Validation* sampling method. However, this model used data with oversampling, which allows to conclude that future work should incorporate more instances, especially of newborns that needed neonatal resuscitation. The number of this kind of infants on the used dataset was not representative enough for DM techniques to successfully learn how to classify them well enough to be trusted for usage in real clinical practice. In this particular case, the number of false positives must be minimized as much as possible, so, even though the obtained results were satisfactory, before implementing a decision-support system based on one of the models induced during this work, there is still more future work to be performed, such as trying new DM approaches and performing more tests.

Acknowledgments

This work has been supported by Compete: POCI-01-0145-FEDER-007043 and FCT within the Project Scope UID/CEC/00319/2013.

References

1. Brandão, A., Pereira, E., Portela, F., Santos, M. F., Abelha, A., & Machado, J. (2014). Managing voluntary interruption of pregnancy using data mining. Procedia Technology, 16, 1297-1306.
2. Tufféry, S. (2011). Data mining and statistics for decision making.
3. Koh, H. C., & Tan, G. (2011). Data mining applications in healthcare. Journal of healthcare information management, 19(2), 65.
4. Power, D. J. (2002). Decision support systems: concepts and resources for managers. Greenwood Publishing Group.
5. Bâra, A., & Lungu, I. (2012). Improving Decision Support Systems with Data Mining Techniques. INTECH Open Access Publisher.
6. Berner, E. S. (2009). Clinical decision support systems: state of the art. AHRQ publication, 90069.
7. Pereira, S., Portela, F., Santos, M. F., Machado, J., & Abelha, A. (2015). Predicting Type of Delivery by Identification of Obstetric Risk Factors through Data Mining. Procedia Computer Science, 64, 601-609.
8. Portela, F., Santos, M. F., Machado, J., Abelha, A., Rua, F., & Silva, Á. (2015, December). Real-time decision support using data mining to predict blood pressure critical events in intensive medicine patients. In Ambient Intelligence for Health (pp. 77-90). Springer International Publishing.
9. Chadha, I. A. (2010). Neonatal resuscitation: Current issues. Indian journal of anaesthesia, 54(5), 428.
10. Kaye, D. (2003). Antenatal and intrapartum risk factors for birth asphyxia among emergency obstetric referrals in Mulago Hospital, Kampala, Uganda. East African medical journal, 80(3), 140-143.
11. Wyllie, J., Bruinenberg, J., Roehr, C. C., Ruediger, M., Trevisanuto, D., & Urlesberger, B. (2015). Resuscitation and support of transition of babies at birth. NOTFALL & RETTUNGSMEDIZIN, 18(8), 964-983.
12. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide.
13. Pereira, S., Portela, F., Santos, M. F., Machado, J., & Abelha, A. (2015, July). Clustering-based Approach for Categorizing Pregnant Women in Obstetrics and Maternity Care. In Proceedings of the Eighth International Conference on Computer Science & Software Engineering (pp. 98-101). ACM.