Predicting Triage Waiting Time in Maternity Emergency Care by means of Data Mining

Sónia Pereira[1], Luís Torres[1], Filipe Portela[1,2], Manuel F. Santos[1], José Machado[1], António Abelha[1]

[1]Algoritmi Centre, University of Minho, Portugal [2] ESEIG, Porto Polytechnic
b7004@dps.uminho.pt; luistorres1792@gmail.com; {cfp, mfs}@dsi.uminho.pt; {jmac, abelha}@di.uminho.pt

Abstract. Healthcare organizations often benefit from information technologies as well as embedded decision support systems, which improve the quality of services and help preventing complications and adverse events. In Centro Materno Infantil do Norte (CMIN), the maternal and perinatal care unit of Centro Hospitalar of Oporto (CHP), an intelligent pre-triage system is implemented, aiming to prioritize patients in need of gynaecology and obstetrics care in two classes: urgent and consultation. The system is designed to evade emergency problems such as incorrect triage outcomes and extensive triage waiting times. The current study intends to improve the triage system, and therefore, optimize the patient workflow through the emergency room, by predicting the triage waiting time comprised between the patient triage and their medical admission. For this purpose, data mining (DM) techniques are induced in selected information provided by the information technologies implemented in CMIN. The DM models achieved accuracy values of approximately 94% with a five range target distribution, which not only allow obtaining confident prediction models, but also identify the variables that stand as direct inducers to the triage waiting times.

1 Introduction

Healthcare professionals are increasingly turning to clinical decision support systems (CDSS) as well as health information technologies that provide hospitals with patient-specific assessments and recommendations to aid clinical decision making [1] [2]. In the emergency department (ED), triage CDSSs are used in order to optimize the patient workflow, improving the quality of care and reduce the risks associated to prolonged waiting lists [3] [4]. The most common triage systems are the classification system with five levels of severity, such as the Emergency Severity Index (ESI) and the Manchester Triage System (MTS), efficient in general emergency units [5].

In the maternity emergency care, these triage systems became inadequate for their lack of flexibility addressing the specific patients that attend the gynaecology and

obstetrics (GO) care services, such as pregnant women at different gestation stages and conditions [5] [6]. Accordingly, the emergency care of GO in CMIN resorts to a pre-triage system, developed in CHP, specified to categorize women in emergency (URG) and consultation (ARGO) classes. The system integrates the intelligent CDSS implemented in CMIN and allows increase patient's safety [5].

The current study aims to predict the patients' waiting time – the time between the patient's triage and their clinical admission, through the induction of data mining (DM) models. The research is based on real data provided by the information systems used in CMIN to collect and store the patients' clinical records. After inducing DM techniques in several data scenarios, the case study achieved useful knowledge to support the maternity emergency room, since the best DM models reached accuracy values of approximately 94%, concerning a five range target dataset. The prediction of triage waiting times helps the emergency service identify the clinical and environmental features leading to longer waiting times. It assists the professionals prioritizing patients and operations, as well as avoiding medical errors, overcrowding and patient elopement.

This article includes five sections in addition to the introduction. The second section presents the context and related work, and the study description follows in section three. Section four tracks the data mining process following the Cross Industry Standard Process for Data Mining (CRISP-DM) phases. Formerly, section five contains a discussion about the obtained results, while the last section includes the conclusions and future directions to the accomplished work.

2 Background and Related Work

2.1 Context

In maternity care, a variety of gynaecology and obstetrics (GO) conditions are presented, since labour assessment issues to antepartum fetus threatening symptoms [6]. An adequate triage system has to be flexible and dedicated to GO guidelines, in order to provide patients with the proper care.

This study focus on the prediction of the patient's waiting time between the triage and the medical admission, which can be essential to the improvement of the GO emergency room patient flow and satisfaction.

2.2 Pre-triage System for Gynaecology and Obstetrics Care in CMIN

In CMIN, the maternity emergency room has to provide patients with proper GO services such as evaluation of labour, fetal examination and obstetric nurture. In this context, a specific emergency pre-triage system is implemented in CMIN since 2010 and establishes clinical priorities according to the severity of the patients' clinical condition. The triage is based on a set of predefined queries in form if rules of a decision tree. The pre-triage system is inserted in the intelligent decision support system (IDSS) developed and currently deployed in CHP that supports this process,

indicating the triage result (urgent or consultation) [5]. The IDSS is an interactive and adaptable system, which uses artificial intelligence techniques and decision models to answer a question. The presence of the IDSS offers a better understanding of the patient's real state [6].

Currently, the pre-triage system helps increasing patient safety for women in need of immediate care and reducing high-risk care in low-risk patients, maximizing the use of resources [7].

2.3 Interoperability, Archive and Diffusion of Medical Information

This study support data was gathered from distinct information systems (Support Nursing Practice System - SAPE and Electronic Health Record – HER) used at CMIN by the Agency for Integration, Archive and Diffusion of Medical Information (AIDA). The AIDA platform is built on a set of pro-active agents that ensure the standardization of clinical systems, overcoming the medical and administrative complexity of the diverse sources of data from the hospital [8]. By providing this interoperation between the hospital existing systems, this platform allows a suitable information management [9] [10].

The data managed by SAPE, the system supporting nursing practices, regards the records of clinical episodes associated with each patient, as an alternative to the traditional way of saving this information on paper. By the other hand, the EHR system handles storing and retrieving of detailed patient information, as the admission form, assisting monitor, report and improve data on health care quality and safety [11] [12].

2.4 Knowledge Discovery and Data Mining in Healthcare

The knowledge Discovery from Databases (KDD) is a five steps process that aims to identify new valid, meaningful and potentially useful information hidden in large data repositories [13] [14]. First of all, the dataset needs to be chosen, which corresponds to the first step. On the second stage, all the data is cleaned and processed, in order to become consistent. Thereafter, accordingly to the study goal, the data undergoes a transformation so it can be properly explored. The fourth step is the core of all the process, and corresponds to the Data Mining (DM), which is where the knowledge is retrieved from the data, through pattern discovery. Here the DM techniques used will depend on the nature of the problem, which can be segmentation, association, prediction and summarization. Finally, the last phase of the process is to interpret and evaluate the results obtained [15].

Healthcare organizations, which nowadays store most of their data in databases, can benefit a lot from the use of DM techniques. The possible applications go from the identification of effective treatments and best practices (better care to the patients), to anticipating patient's future behaviour and finding solutions concerning institution's management (better services and clinical decisions supported by evidence) [16] [17].

The Intensive Care Unit in CHP employ data mining models to predict patient outcome, readmissions, length of stay and organ failure in real-time, among others [18] [19] [20]. Furthermore, many DM studies have been conducted regarding

obstetrics and maternal care, in order to identify services limitations and possible solutions. Among them, DM classification algorithms were used to predict the type of birth using pregnancy characteristics and to predict events in the voluntary interruption of pregnancy [21] [22]. In these cases, a training set containing a group of attributes was provided, so the classification algorithms could discover relationships between them that would make it possible to predict the outcomes.

3   Methodologies, Materials and Methods

Following the KDD process described in section 2.3, the current study uses SAPE and EHR information to accomplish useful knowledge concerning the maternity emergency triage system. The DM phase applies the Cross Industry Standard Process for Data Mining (CRISP-DM), a sequence of defined six steps that allow to structure and guide the DM process [23]. The six stages are business understanding, data understanding, data preparation, modelling, evaluation and deployment, which support the development of DM models to be used in real environments [24].

A total number of 73330 admissions on CMIN's GO care emergency room are included in the study dataset, comprising a period between 2010-01-06 and 2015-06-25 (1850 days), regarding 31620 women patients.

The exploration of the dataset as well as the DM process was performed using the R language and the interface R Studio, for their comprehensiveness and availability. Different classification techniques were considered: Decision Trees (DT), Naïve Bayes (NB), Generalized Linear Models (GLM), Support Vector Machine (SVM) and Neural Networks (NN). The selection of the DM techniques was based on the interpretability of the models, the engine efficiency and their suitability regarding the dataset features.

4   Data Mining Process

The current section describes all the work developed through the DM process, following the KDD process, according to the CRISP-DM phases, having into account the methods and methodologies described previously.

4.1 Business Understanding

The main business goal of the study is to identify the triage features that provide information about the patient waiting time, and therefore, be able to predict it. The prediction of the triage waiting times in the emergency room of the CMIN's GO unit will contribute to the improvement of the triage process.

Thus, the DM aims to develop accurate models able to predict the triage waiting time from the environmental and personal attributes available in the collected data. Decent results can be used by the CMIN's IDSS, enhancing the quality of services among the patient satisfaction.

4.2 Data Understanding

The study dataset meets the data provided by the information systems SAPE and EHR, considering a set of attributes available at triage time. A total of 17 variables were considered, some of those environmental features: the day of the week, the part of the day, the month, the day of the month, the part of the month, the trimester, the hour, the season of the year, the identification number (ID) of the triage professional and their medical specialty, the number of triage professionals working (NTP) and the number of patients waiting in the room (NPW). Additionally, it also considers the age and the gestation weeks (in case of pregnancy) of the patient, as well as their triage result, triage module and motive of visit.

In order to better understand these attributes and their relation with the prediction of triage waiting times, table 1 presents detailed information about some variables and their percentage of occurrence in the dataset. Accordingly, table 2 shows some statistical measures regarding the numerical variables of the study.

Table 1. Classes and occurrences of some variables used in the dataset.

Variable	Class	Percentage	Variable	Class	Percentage
Day of the Week	Sunday	10.23%	Trimester	First Quarter	25.14%
	Monday	18.69%		Second Quarter	27.94%
	Tuesday	14.56%		Third Quarter	24.51%
	Wednesday	15.09%		Last Quarter	22.41%
	Thursday	15.29%	Station	Winter	23.76%
	Friday	14.92%		Spring	28.84%
	Saturday	11.22%		Summer	24.81%
Part of the Day	Morning	44.01%		Autumn	22.59%
	Evening	55.73%	Triage Module	URG	49.02%
	Night	0.26%		ARGO	50.98%
Part of the Month	First Third	29.73%	Triage Result	50	50.98%
	Second Third	36.54%		52	48.99%
	Last Third	33.73%		54	0.03%

Table 2. Statistical measures of the numerical variables of the dataset.

Variable	Min	Max	Avg	Std Dev
NTP	1	63	5.44	3.47
NPW	1	61	1.77	1.92
Age	8	92	32.12	10.53
Gestation Weeks	0	46	11.23	15.40

The target variable *Triage Waiting Time* (TWT) was divided in different range approaches in order to obtained good data mining models. In a first approach, the target was simply distributed in two ranges, separated by the variable mean value. It allowed evaluating the relation between the selected variables and the study's aim, confirming its suitability to predict the triage waiting time, and subsequently, obtaining useful statistical results.

On the other hand, following a more clinical approach, the target TWT was organized having into account the categories of emergency of the Manchester Triage

System (MTS). The category of emergency is associated with a maximum waiting time until the patient is attended by the doctor, regarding their condition. Table 3 highlights these approaches' ranges and occurances in the dataset.

Table 3. Distributions of the target variable through the different assigned approaches.

ID	Approach	Class	Distribution	Percentage
1	Simple Split	0	0 – 17 minutes	96.28%
		1	18 – 516 minutes	3.72%
2	MTS	0	0 – 5 minutes	63.82%
		1	6 – 10 minutes	24.98%
		2	11 – 60 minutes	9.56%
		3	61 – 120 minutes	0.68%
		4	121 – 516 minutes	0.96%

4.3 Data Preparation

The cleaning and processing of a target dataset are important tasks of the data mining process, allowing the transformation of EHR and SAPE information in valuable information.

Firstly, the desired attributes are attained from raw records. For instance, the temporal variables were obtained by processing the entrance, triage and admission hours and dates. Similarly, the number of patients and triage professionals available at the emergency room for a particular record required creating procedures to concatenate the dataset entries given the record's specifications. The noise instances, such as duplicates, inconsistences and missing values are removed from the dataset.

As visible in table 3, there is a disproportion concerning the distribution of the target variable TWT in the three last approaches. In order to provide the study with a balance target dataset, the technique of oversampling was implemented. It consists in replicating the lower target ranges, until a composed dataset is obtained.

4.4 Modelling

Once the datasets are ready, the DM models are induced using the DM techniques presented in section 3: DT, NB, GML, SVM and NN, using R miner and the algorithms and configurations featured in table 1. The sampling method Holdout Sampling was applied, having 30% of the dataset composing the testing set, and the remaining 60% used for training. In order to identify which variables influence the triage waiting time the most, the dataset attributes were combined, generating ten different scenarios to test the DM techniques:

S0: {All variables}

S1: {Day of the week, Part of the day, Trimester, Hour, Season, Triage Module, Motive}

S2: {NTP, NPW, Triage Module, Triage Result}

S3: {Age, Gestation Weeks, Day of the week, Part of the day, Trimester, Hour, Season, NTP, NPW, Triage Module, Motive}

S4: {Month, Day of the month, Hour, ID professional, NPW, Motive, Triage Result}

S5: {Age, Day of the week, Part of the day, Day of the month, Part of the month, Season, NTP, Triage Module, Triage Result }

S6: {Month, Day of the month, Part of the month, Trimester, Season, Motive, Triage Module}

S7: {Day of the week, Part of the day, Hour, Module}

S8: {Age, Gestation Weeks, Day of the month, Season, NTP, NPW, Triage Module, Triage Result, Motive}

S9: {Age, Gestation Weeks, Trimester, NTP, NPW, Triage Result}

Each DM model (DMM) can be identified by equation 1.

$$DMM_m = DMT_y \times A_b \times T \times S_i \tag{1}$$

DMT_Y refers to the DM technique, A_b is the target approach, T_s represents the sampling method and S_i identifies the scenario. A total of 100 models were induced (10 scenarios * 5 techniques * 1 sampling method * 2 target approaches).

4.5 Evaluation

The evaluation of the models considered the accuracy statistic metric described in equation 2. The accuracy is estimated through the results provided by the confusion matrix (CMX) of each model.

$$Accuracy = TP / (TP + FT + TN + FN) \tag{2}$$

The CMX contains four types of results: the number of True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN). The CMX along with the accuracy result were obtained automatically using the package 'caret' in R Studio.

The best results concerning the DM techniques, scenarios and target approaches are exposed in table 4.

Table 4. Best accuracy results in view of the best DM technique and scenario for each target approach

Approach	DMT	Accuracy	DMT	Accuracy	DMT	Accuracy	DMT	Accuracy
1	*Scenario 0*		*Scenario 3*		*Scenario 4*		*Scenario 8*	
	DT	0.9359	DT	0.8982	DT	0.8670	DT	0.8614
	GML	0.8006	GML	0.7954	NB	0.7505	GML	0.7950
	NN	0.7846	NN	0.8041	GML	0.7915	SVM	0.7979
2	*Scenario 0*		*Scenario 3*		*Scenario 8*		*Scenario 9*	
	DT	0.7494	DT	0.7081	DT	0.7277	DT	0.6636

Overall, the best DM models were attained by inducing decision trees (DT). Regarding the attributes, scenario 0 presents best accuracy values, showing that the selected variables influence indeed the triage waiting times in the maternity emergency room.

4.6 Deployment

The best DM models, as well as the new knowledge obtained about the attributes that influence the triage waiting times in the GO unit of maternity emergency room are reported to the maternity care unit of CMIN, being implemented in the IDSS and the Business Intelligence (BI) platform in use in CHP [25]. The BI platform supports the clinical and administrative decision making process, concerning the care and the patients.

5   Discussion

As presented in table 4, the study achieved worthy results for all the persuaded target distributions, allowing sensible predictions of the triage waiting times and therefore, the support to the CMIN's triage process. It is up to the healthcare professionals to choose which target approach could be most beneficial to the melioration of the maternity emergency room patient flow. On the one hand, the first approach, which classifies the admissions in two classes according to the mean triage waiting time in the last 5 years, is the most opportune regarding statistical grounds. On the other hand, if the most clinical fitting approach would be classify the triage waiting times concerning the levels of severity; the second approach should be persuaded. Table 5 briefs the best models to predicting the triage waiting times.

Table 5. Top DM models that present the higher values of accuracy in view of the best target approaches

Target Approach	Scenario	DMT	Accuracy
1	0	DT	0.9359
2	0	DT	0.7494

Another useful contribution of the study is the identification of the dataset attributes as enhancers to the patients' triage waiting time. Variables as the number of patients present in the waiting room, the number of triage professionals working at the moment and some temporal variables can be used to identify outstanding situations in the triage process, and therefore, improve the healthcare services.

In a real time environment, physicians can rely on the DM models to send warnings informing about the triage waiting time itself, but also about workflow issues and uncommon or risk conditions. Consequently, the physicians can be observant and alert to special cases and can put the patients on watch accordingly, allowing the healthcare institution saving resources and time.

6   Conclusions and Future Work

By means of real data obtained from CMIN's information systems SAPE and EHR, it is possible to prove the viability of using DM models to predict the triage waiting

time, in maternity emergency room, through environmental and individual characteristics of the GO unit and patients. Clinically suitable results were achieved regarding the accuracy metric, by inducing the DM technique Decision Trees on data from scenario 0, achieving approximately 96% of accuracy, when using a five level target approach. The best DM models to predict the triage waiting times can also be used to implement a time tracker in the emergency room, since studies indicate that the awareness of waiting times increases the individuals' satisfaction and reduces anxiety and unattended leaves [26]. Accordingly, the best models and the results achieved will be included in the IDSS and the BI platforms, allowing the improvement of the GO patient flow and satisfaction, supporting the physicians' decision-making, leading to quality improvements in the maternity care.

Applying the present DM strategy on a different setting or environment could represent a significant step to evince the support decision solutions benefits on improving the quality of the emergency services, by aiding both health professionals and patients.

Acknowledgments

References

1. Kawamoto, K., Houlihan, C., Balas, E., Lobach, D.: Impriving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ, 1-8 (2005)

2. Haynes, R., Wilczynski, N.: Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: Methods of a decision-maker-researcher partnership systematic review. Implementation Science 5 (12), 1-8 (2010)

3. Kaushal, A., Zhao, Y., Peng, Q., Strome, T., Weldon, E., Zhang, M., Chochinov, A.: Evaluation of fast track strategies using agent-based simulation modeling to reduce waiting time in a hospital emergency department. Socio-Economic Planning Sciences 50, 18-31 (2015)

4. Bergs, J., Verelst, S., Gillet, J.-B., Deboutte, P., Vandoren, C., Vandijck, D.: The number of patients simultaneously present at the emergency department as an indicator of unsafe waiting times: A receiver operated curve-based evaluation. International Emergency Nursing 22 (2014)

5. Abelha, A., Pereira, E., Brandão, A., Portela, F., Santos, M., Machado, J., Braga, J.: Improving Quality of Services in Maternity Care Triage System. International Journal of E-Health and Medical Communications 6 (2), 10-26 (2015)

6. Cabral, A., Pina, C., Machado, H., Abelha, A., Salazar, M., Quintas, C., Portela, F., Machado, J., Neves, J., Santos, M.: Data Acquisition Process for an Intelligent Decision Support in Gynecology and Obstetrics Emergency Triage. CENTERIS, 223-232 (2011)

7. Pereira, E., Brandão, A., Salazar, M., Portela, F., Santos, M., Machado, J., Abelha, A., Braga, J.: Pre-Triage Decision Support Improvement in Maternity Care by means of Data Mining. Integration of Data Mining in Business Intelligence Systems, 175-192 (2014)

8. Abelha, A., Machado, J., Santos, M., Allegro, S., Rua, F., Paiva, M., Neves, J.: Agency for Integration, Diffusion and Archive of Medical Information. IASTED (2002)

9. Abelha, A., Analide, C., Machado, J., Neves, J., Santos, M., and Novais, P.: Ambient intelligence and simulation in health care virtual scenarios. Establishing the Foundation of Collaborative Networks , 461-468 (2007)

10. Peixoto, H., Santos, M., Abelha, A., Machado, J.: Intelligence in Interoperability with AIDA. 20th International Symposium on Methodologies for Intelligent Systems (2012)

11. Cardoso, L., Marins, F., Portela, F., Santos, M., Abelha, A., Machado, J.: The Next Generation of Interoperability Agents in Healthcare. Int. J. Environ. Res. Public Health 11, 5349-5371 (2014)

12. Portela, F., Cabral, A., Abelha, A., Salazar, M., Quintas, C., Machado, J., Santos, M.: Knowledge Acquisition Process for Intelligent Decision Support in Critical Health Care. Healthcare Administration: Concepts, Methodologies, Tools, and Applications, 270 (2014)

13. Frawley, W., Piatetsky-Shapiro, G., Matheus, C.: Knowledge discovery in databases: An overview. AI magazine 13(3), 57 (1992)

14. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. AI magazine 17 (3), 37 (1996)

15. Maimon, O., Rokach, L.: Introduction to Knowledge Discovery and Data Mining. Data Mining and Knowledge Discovery Handbook, 1-17 (2005)

16. Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J.-F., Hua, L.: Data Mining in Healthcare and Biomedicine: A Survey of the Literature. J Med Syst 36, 2431–2448 (2012)

17. Obenshain, M.: Application of data mining techniques to healthcare data. Infection Control 25(08), 690-695 (2004)

18. Braga, P., Portela, F., Santos, M., Rua, F.: Data mining models to predict patient's readmission in intensive care units. In : ICAART (2014)

19. Portela, F., Santos, M., Machado, J., Abelha, A., Silva, Á.: Pervasive and Intelligent Decision Support in Critical Health Care Using Ensembles. Information Technology in Bio-and Medical Informatics Springer Berlin Heidelberg, 1-16 (2013)

20. Veloso, R., Portela, F., Santos, M., Sila, Á., Rua, F., Abelha, A., Machado, J.: A Clustering Approach for Predicting Readmissions in Intensive Medicine. Procedia Technology 16, 1307-1316 (2014)

21. Pereira, S., Portela, F., Santos, M., Machado, J., Abelha, A.: Predicting Type of Delivery by Identification of Obstetric Risk Factors through Data Mining. Procedia Computer Science - HCIST 2015 - Healthy and Secure People (2015)

22. Brandão, A., Pereira, E., Portela, F., Santos, M., Abelha, A., Machado, J.: Managing Voluntary Interruption of Pregnancy using Data Mining. Procedia Technology 16, 1297–1306 (2014)

23. Shafique, U., Qaiser, H.: A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA). International Journal of Innovation and Scientific Research (2014)

24. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRISP-DM 1.0 Step-by-step data mining guide. (2000)

25. Pereira, E., Brandão, A., Portela, F., Santos, M., Machado, J., Abelha, A.: Business Intelligence in Maternity Care. In : IDEAS (2014)

26. Shaikh, S., Witting, M., Winters, M., Brodeur, M., Jerrad, D.: Support for a waiting room time tracker: a survey of patients waiting in an urban ED. Journal of Emergency Medicione (2013)