



HAL
open science

A DATA MINING-BASED APPROACH TO PREDICT STRAIN SITUATIONS IN HOSPITAL EMERGENCY DEPARTMENT SYSTEMS

Sofia Benbelkacem, Farid Kadri, Sondès Chaabane, Baghdad Atmani

► **To cite this version:**

Sofia Benbelkacem, Farid Kadri, Sondès Chaabane, Baghdad Atmani. A DATA MINING-BASED APPROACH TO PREDICT STRAIN SITUATIONS IN HOSPITAL EMERGENCY DEPARTMENT SYSTEMS. 10ème Conférence Francophone de Modélisation, Optimisation et Simulation- MOSIM'14, Nov 2014, Nancy, France. hal-01081556

HAL Id: hal-01081556

<https://hal.science/hal-01081556>

Submitted on 9 Nov 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A DATA MINING-BASED APPROACH TO PREDICT STRAIN SITUATIONS IN HOSPITAL EMERGENCY DEPARTMENT SYSTEMS

S. BENBELKACEM¹, F. KADRI^{2,3}

S. CHAABANE^{2,3}, B. ATMANI¹

¹LIO Lab., University of Oran, Oran, Algeria

²Univ. Lille Nord de France, F-59000 Lille, France

³TEMPO Lab., PSI Team, F-59313 Valenciennes, France

sofia.benbelkacem@gmail.com

farid.kadri@univ-valenciennes.fr

sondes.chaabane@univ-valenciennes.fr

atmani.baghdad@gmail.com

ABSTRACT: *Nowadays, emergency departments are confronted to exceptional events such as epidemics problems or health crises. These situations increase the patient flow. The consequence of this influx of patients has resulted in problems of ED overcrowding which often increases the length of stay of patients (LOS) in EDs and leads to strain situations. To cope with such situations, ED managers must predict the LOS. In this paper, we propose a model for predicting the patient length of stay (LOS) in ED using data mining techniques. The used data was collected from the pediatric emergency department (PED) in Lille regional hospital centre, France. Our target is to illustrate with a real world case study, how data mining can be benefit to predict LOS and which its limitations.*

KEYWORDS: *data mining, emergency department, classification, prediction, length of stay.*

1 INTRODUCTION

Nowadays, with the growing demand for emergency medical care, the management of hospital emergency departments (EDs) has become increasingly important. The management of patient influx is one of the most crucial problems in EDs throughout the world (Kadri et al., 2014). To deal with this influx of patients, emergency departments require significant human and material resources, as well as a high degree of coordination among human and material elements (Kadri et al., 2014a; Kadri et al., 2013). Unfortunately, these resources are limited. The consequences of this influx of patients have resulted in problems of ED overcrowding (Boyle et al., 2012; Kadri et al., 2014b). This latter increases the length of stay of patients in EDs which lead to violence of angry patients against staff, patients leave the ED without being treated, reduced access to emergency medical services and increase in patient mortality (Sprivulis et al., 2006).

Thanks to information and computerized hospital, the data produced by emergency departments (EDs) is still increasing. With the use of data mining techniques it is possible to extract interesting and useful knowledge and regularities which may be used as a tool for decision making in such establishments, in order to respond to the needs of ED managers in their daily decision making activities. The aim of this study is to see how can we use data mining techniques and particularly classification methods to develop models for prediction of patients' length of stay (LOS) at the emergency department.

The remainder of this paper is organized as follows: Section 2 describes the problem. Section 3 summarizes the current data mining techniques and their applications in healthcare systems. Section 4 presents the approach used in this work applied to our study case. Section 5 discusses obtained results. Finally, section 6 reviews the main conclusions in this work and exposes the perspectives of the study.

2 PROBLEM DESCRIPTION

Emergency departments are confronted to exceptional events due to seasonal epidemics (in winter: influenza, colds, gastroenteritis, bronchiolitis, etc., in summer: trauma), health crises... These events can cause strain situations when disequilibrium between the care load flow (demand) and the care production capacity (supply) is observed. The management of patient flow in such situations is one of the most important problems managed by ED managers. To handle this influx of patients, emergency departments require significant human and material resources, but these are limited. To confront these situations and face these constraints, emergency departments have no choice but to adapt.

To help manager to resolve this problem (Farid Kadri et al., 2014b) propose a decision support system. One of the principle part is to Prevent and predict strain situations with a predictive model. They identify and develop ten relevant indicators validated by professionals. One of the important indicators is the length of stay defined as the time between when the patient completes their ad-

ministrative registration at the ED and their discharge (departure) from the ED. In strain situations, LOS increases considerably and causes violence of angry patients. Thus, the estimation and prediction of the patient length of stay (LOS) is important to detect the beginning of strain situations to help systems to react to such situations. It helps or managing patients and internal resources in EDs, and predict sufficient hospitalization capacity downstream.

Our purpose is to see how data mining models and techniques can be used to predict LOS indicator and which limitations of such approaches.

3 RELATED WORK

Koh and Tan (2005) present a data mining as a relatively recently developed methodology and technology, coming into prominence only in 1994 (Trybula, 1997). It aims to identify valid, novel, potentially useful, and understandable correlations and patterns in data by combing through copious data sets to sniff out patterns that are too subtle or complex for humans to detect. In this paper, authors explored data mining applications within healthcare in diverse areas involving: evaluation of treatment effectiveness, management of healthcare, customer relationship management, and detection of fraud. This review shows that data mining has been largely applied in the hospital sector to provide useful information to improve hospital establishment efficiency. Milovic (2012) mentioned data mining applications for a variety of healthcare areas, such as doctors who use patterns by measuring clinical indicators, quality indicators, customer satisfaction and economic indicators, performance of physicians from multiple perspectives to optimize use of resources, cost efficiency and decision making based on evidence, identifying high-risk patients and intervene proactively, optimize health care, etc.

More recently, Esfandiari et al. (2014) presented a review study, including papers published between 1999 and 2013, in the context of medical data mining applications. Five data mining approaches were considered: classification, regression, clustering, association and hybrid.

In this section, the literature review is focused on the supervised classification approaches. It refers to supervised methods that determine target class value of unseen data. It is used to identify profiles of classes in terms of their attributes. The most popular supervised classification algorithms in medical data mining based on the literature (Esfandiari et al. 2014) are : decision tree (DT), multi-layer perception type of artificial neural network model (MLP), support vector machine (SVM) and naive bayes (NB).

Walczak and Scharf (2000) used artificial neural network to predict the quantity of transfusion units that are required by surgical patients for a specific operation. The results show that artificial neural networks are capable of reducing the overall quantity of blood units ordered for operations. Thus, the artificial neural networks offer a means to reduce patient costs while maintaining a high level of patient care.

Chae et al. (2003) used a decision tree to analyze the healthcare quality indicators in order to develop quality improvement strategies. The authors identified the important factors influencing the inpatient mortality: length of stay, disease classes, discharge departments, and age groups. The results provided cumulative statistics that show how the authors found the inpatient mortality.

Kraft et al. (2003) are the first to use nursing diagnosis and neural networks to predict the length of stay of patients with spinal cord injuries. Four types of neural networks were developed for building predictive models for length of stay: dynamic network, prune network, multilayer perceptron, and radial basis function network. The results show that the radial basis function network model neural network was selected as the best model.

Walsh et al. (2004) used artificial neural network ensembles to predict the disposition and length of stay in children presenting to the Emergency Department with bronchiolitis. The results show that the neural network ensembles correctly predicted disposition in 81% of test cases. The authors concluded that artificial neural network ensembles can predict disposition for infants and toddlers with bronchiolitis; however, the prediction of length of hospital stay is not as good.

To illustrate a data mining application in healthcare, Koh and Tan (2005) used the decision tree to find out how certain variables are associated with the onset of diabetes. The purpose of this healthcare data mining application is to identify high-risk individuals. The results show that age is the most important factor associated with the onset of diabetics.

Liu et al. (2006) applied decision trees and naive Bayesian classifiers to a geriatric hospital dataset in order to predict inpatient length of stay. Results show that naive Bayesian models performed better in comparison with the C4.5 algorithm of decision tree. So, applying naive Bayesian imputation models to handle a considerable amount of missing data can greatly increase the classification accuracy of predicting length of stay.

De Toledo et al. (2009) used decision trees techniques to build a model of outcome prediction for subarachnoid hemorrhage that makes the knowledge discovered from the data explicit and communicable to domain experts. The algorithms used in this paper were C4.5, fast decision tree learner, partial decision trees, nearest neighbor

with generalization, and ripple down rule learner. Compared to other algorithms, the C4.5 obtained the best results

Delen et al. (2009) built a classification model to analyze the healthcare coverage using artificial neural networks and decision trees. The developed model was used to predict if an individual has healthcare coverage or not based on the specific information's about socio-demographic and lifestyle, and the importance of the various factors in the classification model. The results indicated that the most accurate classifier for this phenomenon was the multi-layer perception type of artificial neural network model.

Aljumah et al. (2013) used data mining techniques to discover patterns that identify the best mode of treatment for diabetes across different age groups. In this study, predictions on the effectiveness of different treatment methods were elucidated using a support vector machine algorithm. These results indicated that dietary treatment is more effective for patients in the old age group than patients in the young group.

Different applied methods and application fields are summarized in table 1.

Method	Reference	Application field
DT	Chae et al. (2003)	Identify factors influencing inpatient mortality
	Koh and Tan (2005)	Find variables associated with onset of diabetes to identify high-risk individuals
	De Toledo et al. (2009)	Build a model of outcome prediction of subarachnoid hemorrhage
ANN	Walczak and Scharf (2000)	Predict quantity of transfusion units required by surgical patients
	Kraft et al. (2003)	Predict length of stay of patients with spinal cord injuries
	Walsh et al. (2004)	Predict length of stay in children presenting bronchiolitis to ED
DT ANN	Delen et al. (2009)	Predict if an individual has healthcare coverage
SVM	Aljumah et al. (2013)	Predict the effectiveness of treatments of diabetes
NB	Liu et al. (2006)	Predict inpatient length of stay of geriatric patients

Table 1: Classification algorithms applied in healthcare systems

4 DATA MINING BASED PROCESS : A CASE OF STUDY

Lille Regional Hospital Centre (CHR-Lille) serves four million inhabitants in Nord-Pas-de-Calais, a region characterized by one of the largest population densities in France (7% of the French population). The paediatric emergency department (PED) is open 24 hours a day and receives 23 900 patients a year on average. Besides its internal capacity, the PED shares many resources, such as administrative patient registration and clinical laboratories, with other hospital departments. In this study the data mining based approach is presented in the figure 1 and applied to predict the patient's length of stay at the PED.

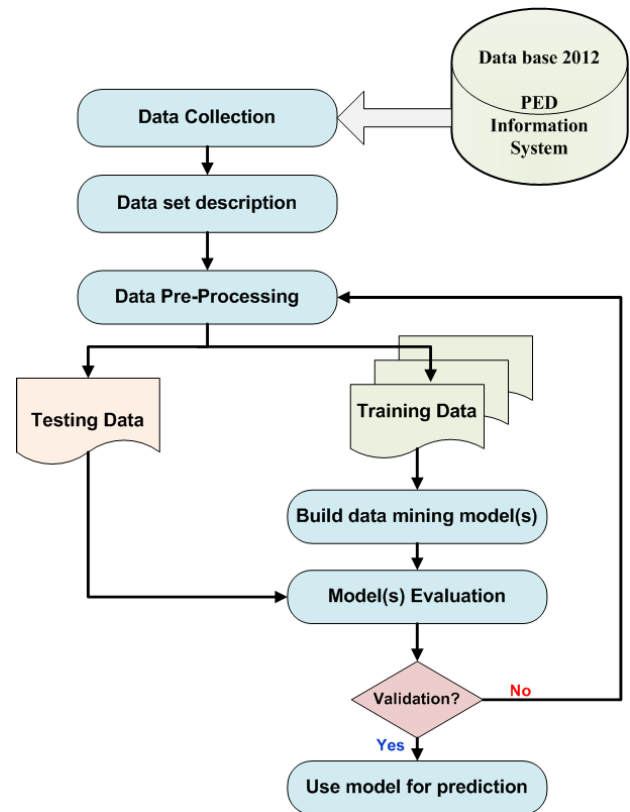


Figure 1: General architecture of the proposed approach

4.1 Data Collection

The first step of the proposed approach is the collection and the preparation of the data. This step is composed of several sub-steps: (1) data sources are located, accessed, and selected. (2) Selected data is put into a tabular format in which instances and variables take place in rows and columns, respectively (Giudici, 2005).

Our study case was conducted utilizing a dataset extracted from the database of PED of CHR-Lille. As the target is to predict the patient's LOS in strain situations,

a winter period from January to March 2012 including 6135 records is selected.

4.2 Data set Description

The aim of this step is the description of the dataset identified in previous step. Let $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ denote the set of objects taken into account for the training. Each object ω_i is described by a set of variables X_1, X_2, \dots, X_p called descriptive variables or exogenous variables. The value taken by a descriptive variable X_j is called modality or value of the attribute X_j of an individual ω_i . An attribute C , called class variable or endogenous variable, is associated with every individual ω_i . C takes its values in the set of classes $C = \{c_1, c_2, \dots, c_m\}$ (Atmani and Beldjilali, 2007).

Consider our case of study to predicting the LOS at the PED. This problem can be described by 12 exogenous variables $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}$ and detailed in table 2.

Exogenous variables	Signification	Description
X_1	Arrival time	Patient arrival time at the PED (hour of day)
X_2	Arrival mean	Arrival mean of patients
X_3	Orientation	Orientation of patient at the end of emergency management
X_4	Age	Age of patients
X_5	Sex	Sex of patients
X_6	CCMU	Clinical classification of emergency patients called CCMU (1-6)
X_7	Echography	If the patient had an echography
X_8	Scanner	If the patient had a scanner
X_9	Radiology	If the patient had a radiology
X_{10}	Biology	If the patient had a biology test
X_{11}	Diagnostic	Diagnostic of patients
X_{12}	GEMSA	Multicentric Emergency Department Study Group called GEMSA (1-6)

Table 2: Exogenous variables, signification and description

Each individual is associated with a class C corresponding to the LOS using a discretized method in WEKA tool. A set of classes $C = \{c_1, c_2, c_3\}$ is obtained according bounds:

- Class c_1 ($LOS < 289$ minutes): represents the *short length of stay* (LOS);
- Class c_2 ($289 < LOS < 432$ minutes): represents the *average LOS*;
- Class c_3 ($LOS > 432$ minutes): represents the *long LOS*.

The last class is the most interesting for prediction particularly for to detect strain situations in PED.

After the description, the data set is separated in two subsets:

- *Training data* (90% of records): is used to build a prediction model by applying a data mining methods;
- *Testing data* (10% of records): is used to evaluate the resulting model and to test the accuracy of predictions.

4.3 Data Preprocessing

Real-world data may be incomplete, noisy, and inconsistent, which can disguise useful patterns. Thus, the pre-processing step is required to generate quality data and improve the efficiency of data mining models (Zhang et al., 2003). First, anomalies are removed and duplicate records are eliminated. Then, numeric attributes are discretized by cutting its range of values in a finite number of intervals. After that, a classifier subset evaluator (Hall, 1999) is used to select attributes. It evaluates the predictive ability of attributes preferring sets of attributes that are highly correlated with the class. The fact that many attributes depend on one another often unduly influences the accuracy of prediction models (Kotsiantis, 2007). For our case of study, the open source tool Weka (Witten and Frank, 2005) to do this step. Weka is a collection of data mining algorithms and data pre-processing methods for a wide range of tasks. The result of this step is very important because we define the set of data that will be used in the remaining steps.

4.4 Build classification Models

In this step, in order to find and select the best models, we applied diverse supervised classification techniques to the training data. As presented previously, the most popular classification algorithms of data mining applied in healthcare based on the literature are used:

- Naïve Bayes (*NB*) and BayesNet ;
- Support vector machine learner (*SVM*) ;
- Decision tree algorithms including: *Id3*, *C4.5*, *REPTree*, *BFTree* and *PART*. *Id3* and *C4.5* are the most well-know algorithms in the literature for building decision trees; *REPTree* is a fast decision tree learner whereas *PART* is a rule induction algorithm and *BFTree* is best-first decision tree learning.

All these algorithms are available in Weka tool.

4.5 Models Evaluation

In this step, the testing data is used in order to validate the chosen model(s). The information discovered in the modelling is evaluated to assess its utility and reliability. To obtain reliable results, the extracted knowledge should be evaluated by a comparison of results obtained with various supervised classification methods and using several measures (Dunham, 2006). Performance evaluation of classifiers can be measured by hold-out, random sub-sampling, cross-validation and bootstrap (Esfandiari et al., 2014). The most common method is the cross validation (Browne, 2000). Furthermore, performance measures can be used to analyze predictive models. They based on four values of the confusion table (Figure 2): true positives (*TP*), false positives (*FP*), true negatives (*TN*), false negatives (*FN*).

		Observed	
		True	False
Predicted	True	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

Figure 2: Format of confusion table

Five performance measures are used:

- **Accuracy:** is a rate of correct classification defined by $\frac{TP+TN}{TP+TN+FP+FN}$
- **Precision:** is a rate of true positive classification defined by $\frac{TP}{TP+FP}$
- **Recall:** is the evaluation and ranking of each sample based on positive class defined by $\frac{TP}{TP+FN}$
- **Kappa statistic:** measures the agreement of prediction with the true class labels. A score value of 1.0 signifies complete agreement, and a value greater than 0 means that the classifier is doing better than pure random behavior (Azari et al., 2012).
- **ROC:** gives the trade-off between true positive rate defined by $\frac{TP}{TP+FN}$ (*recall*) and false positive rate defined by $\frac{FP}{FP+TN}$ for a given model (Han et al., 2011).

If the evaluation results are not enough performing then the process is resumed from the data pre processing step.

5 RESULTS AND DISCUSSION

The valuation of models is carried out using 10-fold cross validation on the PED data set. It is based on the partition of the original sample into ten subsamples, using nine as training data and one for testing data. The cross validation process is then repeated ten times with each of the ten samples, averaging the results from the ten folds to produce a single estimation (De Toledo et al., 2009).

Table 3 shows the performance of the prediction models obtained with each classifier in terms of accuracy, precision, recall, kappa statistic, and ROC Area.

	Accuracy	Precision	Recall	Kappa statistic	ROC Area
NB	78.463	0.76	0.785	0.351	0.829
BayesNet	78.286	0.763	0.783	0.360	0.83
SVM	79.942	0.725	0.799	0.237	0.609
Id3	72.291	0.758	0.784	0.3	0.691
PART	78.576	0.742	0.786	0.298	0.778
C4.5	79.701	0.748	0.797	0.282	0.706
REPTree	78.945	0.724	0.789	0.210	0.728
BFTree	79.170	0.739	0.792	0.245	0.735

Table 3: Performance evaluation of LOS prediction models

The maximum accuracy score equal to 79.942 is obtained by SVM and the lowest one equal to 72.291 is obtained by Id3. The maximum kappa score belongs to 0.36 with BayesNet while the lowest kappa score belongs to 0.210 with REPTree. The highest precision score again belongs to BayesNet (0.763). Furthermore, we can see that the recall was slightly better with SVM. The maximum score of ROC Area belongs to 0.83 with BayesNet and the lowest score is for SVM (0.609). According this result, BayesNet and SVM give the best results for the performance measures.

Let us define the *Vindicator* (*i,j*): is the value of the indicator $i \in \{Accuracy, Precision, recall, Kappa statistic, ROC area\}$ for the model $j \in \{NB, BayesNet, SVM, Id3, PART, C4.5, REPTree, BFTree\}$. Figure 3 gives the comparison of all models according the *RateOfMax* (*i*) and defined by:

$$RateOfMax(i) = \frac{Vindicator(i,j)}{Max_{j \in SetofModels}(Vindicator(i,j))}$$

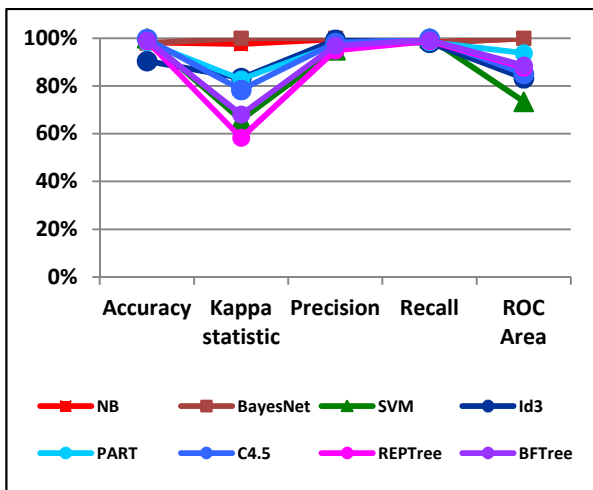


Figure 3: Rate of maximum

The results show that BayesNet and NB give the best result according all performance measures. However DT models are for the most, more than 80% closer to the maximum. The conclusion is that decision trees give comparable results compared to NB and BayesNet. In this case, we preferred decision trees because they are reputed to be several orders of magnitude faster than neural networks and support vector machine.

As shown by results, this study can be benefit to PED manager to predict LOS and detect the beginning of strain situation. This information can be used to make the decision more proactive. However as presented in Koh and Tan (2005), this study presents some limitations.

First, this study can be limited by the accessibility of data because of heterogeneity of source of data (administration, laboratory, PED ...). Hence, data must be collected and integrated before used. One of solution proposed in the literature is the use of data warehouse. But it is a costly and time consuming project.

Secondly, data can be missing, or non-standardized such as pieces of information recorded in different format. For example, the variable X3 (Orientation) we can find two different records "RETURN HOME with external consultation next 7 days" and "RETURN HOME with external consultation next 8 days". This item can be standardized as "RETURN HOME" and have another column specifying the delay for external consultation.

Thirdly, "the successful application of data mining requires the knowledge of the domain area as well as in data mining methodology and tools" (Koh and Tan 2005). This is true in our case, for example the definition of endogenous variables (classes) C is obtained by WEKA tool. The question is how can we trust this result? Did professionals have the same bounds? To com-

plete this study and response to such question, it is important to involve professionals in the whole data mining process.

6 CONCLUSION

In this paper, we presented an approach based on supervised classification methods in order to predict the patient length of stay at the pediatric emergency department in Lille regional centre, France. We described the necessary steps that should be taken to preprocess and prepare the data set. After that, we tested and compared the performance of experiment results with different classification methods. The results indicated that BayesNet and NB give the best results. However decision tree methods obtain comparable results and can successfully be suitable for the prediction of patient's length of stay. This latter, may help EDs managers in their daily decision making activities.

This study illustrates the benefit of classification methods to make PED management system more proactive face to strain situations. However some limitations must be taken into account. Some perspectives can be considered. First, additional data set must be considered and new experiments must be conducted. Second, other data mining methods must be tested with a larger dataset in order to extract knowledge's and to build models that may help PED managers in decision making in the management of PED activities. Finally, professionals must be involved in the process to give the knowledge and information about the domain.

REFERENCES

- Aljumah, A.A., Ahamad, M.G., Siddiqui, M.K., 2013. Application of data mining: Diabetes health care in young and old patients. *J. King Saud Univ. - Comput. Inf. Sci.* 25, 127–136.
- Atmani, B., Beldjilali, B., 2007. Neuro-IG: A Hybrid System for Selection and Elimination of Predictor Variables and Non Relevant Individuals. *Informatica* 18, 163–186.
- Azari, A., Janeja, V.P., Mohseni, A., 2012. Predicting Hospital Length of Stay (PHLOS): A Multi-tiered Data Mining Approach, in: *IEEE 12th International Conference on Data Mining Workshops (ICDMW)*. pp. 17–24.
- Boyle, A., Beniuk, K., Higginson, I., Atkinson, P., 2012. Emergency Department Crowding: Time for Interventions and Policy Evaluations. *Emerg. Med. Int.* 2012.
- Browne, M.W., 2000. Cross-validation Methods. *J Math Psychol* 44, 108–132.
- Chae, Y.M., Kim, H.S., Tark, K.C., Park, H.J., Ho, S.H., 2003. Analysis of healthcare quality indicator using data mining and decision support system. *Expert Syst. Appl.* 24, 167–172.

- De Toledo, P., Rios, P.M., Ledezma, A., Sanchis, A., Alen, J.F., Lagares, A., 2009. Predicting the Outcome of Patients With Subarachnoid Hemorrhage Using Machine Learning Techniques. *IEEE Trans. Inf. Technol. Biomed.* 13, 794–801.
- Delen, D., Fuller, C., McCann, C., Ray, D., 2009. Analysis of Healthcare Coverage: A Data Mining Approach. *Expert Syst Appl* 36, 995–1003.
- Dunham, M.H., 2006. *Data Mining: Introductory And Advanced Topics*. Pearson Education.
- Esfandiari, N., Babavalian, M.R., Moghadam, A.-M.E., Tabar, V.K., 2014. Knowledge discovery in medicine: Current issue and future trend. *Expert Syst. Appl.* 41, 4434–4463.
- Giudici, P., 2005. *Applied Data Mining: Statistical Methods for Business and Industry*. John Wiley & Sons.
- Hall, M.A., 1999. Correlation-based Feature Selection for Machine Learning.
- Han, J., Kamber, M., Pei, J., 2011. *Data Mining: Concepts and Techniques: Concepts and Techniques*. Elsevier.
- Kadri, F., Chaabane, S., Harrou, F., Tahon, C., 2014. Modélisation et prévision des flux quotidiens des patients aux urgences hospitalières en utilisant l'analyse de séries chronologiques, in: 7ème Conférence de Gestion et Ingénierie Des Systèmes Hospitaliers (GISEH), pp. 8.
- Kadri, F., Chaabane, S., Harrou, F., Tahon, C., 2014a. Time series modelling and forecasting of emergency department overcrowding. *J. Med. Syst.*, 38(9):107. doi: 10.1007/s10916-014-0107-0
- Kadri, F., Chaabane, S., Tahon, C., 2014b. A simulation-based decision support system to prevent and predict strain situations in emergency department systems. *Simul. Model. Pract. Theory* 42, 32–52.
- Kadri, F., Pach, C., Chaabane, S., Berger, T., Trentesaux, D., Tahon, C., Sallez, Y., 2013. Modelling and management of the strain situations in hospital systems using an ORCA approach, *IEEE IESM*, 28-30 October. RABAT - MOROCCO, p. 10.
- Koh, H.C., Tan, G., 2005. Data mining applications in healthcare. *J. Healthc. Inf. Manag. JHIM* 19, 64–72.
- Kotsiantis, S.B., 2007. Supervised Machine Learning: A Review of Classification Techniques, in: *Proceedings of the Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*. IOS Press, Amsterdam, The Netherlands, The Netherlands, pp. 3–24.
- Kraft, M.R., Desouza, K., Androwich, I., 2003. Data mining in healthcare information systems: case study of a veterans' administration spinal cord injury population, in: *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*. Presented at the Proceedings of the 36th Annual Hawaii International Conference on System Sciences.
- Liu, P., Lei, L., Yin, J., Zhang, W., Naijun, W., El-Darzi, E., 2006. Healthcare Data Mining: Prediction Inpatient Length of Stay, in: *2006 3rd International IEEE Conference on Intelligent Systems*. Presented at the 2006 3rd International IEEE Conference on Intelligent Systems, pp. 832–837.
- Milovic, B., 2012. Prediction and decision making in Health Care using Data Mining. *Int. J. Public Health Sci. IJPHS* 1.
- Sprivilis, P.C., Da Silva, J.-A., Jacobs, I.G., Frazer, A.R.L., Jelinek, G.A., 2006. The association between hospital overcrowding and mortality among patients admitted via Western Australian emergency departments. *Med. J. Aust.* 184, 208–212.
- Trybula, W., 1997. Data Mining and Knowledge Discovery. *Annu. Rev. Inf. Sci. Technol.* 32, 197–229.
- Walczak, S., Scharf, J.E., 2000. Reducing surgical patient costs through use of an artificial neural network to predict transfusion requirements. *Decis. Support Syst.* 30, 125–138.
- Walsh, P., Cunningham, P., Rothenberg, S.J., O'Doherty, S., Hoey, H., Healy, R., 2004. An artificial neural network ensemble to predict disposition and length of stay in children presenting with bronchiolitis. *Eur. J. Emerg. Med. Off. J. Eur. Soc. Emerg. Med.* 11, 259–264.
- Witten, I.H., Frank, E., 2005. *Data Mining: Practical Machine Learning Tools and Techniques*, Second Edition. Morgan Kaufmann.
- Zhang, S., Zhang, C., Yang, Q., 2003. Data preparation for data mining. *Appl. Artif. Intell.* 17, 375–381.