Performance Evaluation of Predictive Classifiers for Pregnancy Care

Mário W. L. Moreira[1,2], Joel J. P. C. Rodrigues[1,3,4], Antonio M. B. Oliveira[2], Kashif Saleem[5], Augusto Neto[6]

[1] Instituto de Telecomunicações, University of Beira Interior, Covilhã, Portugal
[2] Federal Institute of Education, Science and Technology of Ceará (IFCE), Aracati, Brazil
[3] National Institute of Telecommunications (Inatel), Santa Rita do Sapucaí, MG, Brazil
[4] University of Fortaleza (UNIFOR), Fortaleza, CE, Brazil
[5] Center of Excellence in Information Assurance (CoEIA), King Saud University (KSU), Riyadh, Kingdom of Saudi Arabia
[6] Federal University of Rio Grande do Norte (UFRN), Natal, Brazil

mario.moreira@it.ubi.pt; joeljr@ieee.org; amauroboliveira@gmail.com; ksaleem@ksu.edu.sa; augusto@dimap.ufrn.br

Abstract— **Hypertensive disorders are the leading cause of deaths during pregnancy. Risk pregnancy accompaniment is essential to reduce these complications. Decision support systems (DSS) are important tools to patients' accompaniment. These systems provide relevant information to health experts about clinical condition of the patient anywhere and anytime. In this paper, a model that uses the Naïve Bayesian classifier is introduced and its performance is evaluated in comparison with the Data Mining (DM) classifier named J48 Decision Tree. This study includes the modeling, performance evaluation, and comparison between models that could be used to assess pregnancy complications. Evaluation analysis of the results is performed through the use of Confusion Matrix indicators. The founded results show that J48 decision tree classifier performs better for almost all the used indicators, confirming its promising accuracy for identifying hypertensive disorders on pregnancy.**

Keywords— e-Health; Hypertension; Decision support systems; Bayes methods; Decision trees; Data mining; Pregnancy

I. Introduction

Although many efforts have been performed to reduce deaths during gestation, about more than 800 women die every day in pregnancy complications according to the World Health Organization (WHO). Hypertensive disorders are the most common cause of these complications occurring in about 2-3% of pregnancies. The causes of these diseases in pregnancy have not been well established. Research shows that there is an association with hypertension, which can be chronic or specific pregnancy. This disease is a risk factor for future development of other complications. Even women who had normalization of blood pressure after childbirth in a long term is four times higher the risk of developing chronic hypertension. One way to avoid the aggravation of these problems is the careful and systematic prenatal care during pregnancy.

With the growing number of available data in healthcare, the DM techniques could be a very important tool in knowledge extraction of these data helping health experts in making decisions aimed at prevention and health promotion. DM is one of the most promising available technologies for information extraction in huge amount of data. This is due to the fact that companies spent lots of money in data collection and no useful information is identified. Before these techniques, the transformation process from data to information (and, after, into knowledge) was performed through manual processing by experts in order to produce reports for analysis. However, in most situations, due to the large volume of data, this process has become impractical. The knowledge discovery in databases (KDD) is an attempt to address this data overload.

Wu *et al.* discuss the introduction of the Big Data concept and as it is rapidly growing in all science and engineering domains [1]. A theorem that characterizes the Big Data expansion is based on DM perspectives. It involves the following four domains: demand aggregation, mining and analysis, modeling, and security and privacy. The open issues on this data-driven model and the growing topic Big Data are also analyzed. Mukhopadhyay *et al.* present the main characteristics of each DM technique used to build efficient predictive or descriptive models that use a large amount of data [2]. This paper presents a comprehensive survey addressing the recent developments of multiobjective evolutionary algorithms based on DM techniques, and tries to solve relative problems in a large amount of data. Some concepts related to optimization and DM for Internet of Things (IoT) are also proposed in [3]. It discusses the relationship between IoT and DM giving a brief review considering the features of these two concepts. Kalegele *et al.* discuss the DM usage for networks and systems managing during last forty years [4]. This research work discusses the perspective of the critical open issues for the effective application of DM in heterogeneous systems. DM presents itself as an efficient technique, responsive, reliable, and able to capture information, which is considered important hidden in large volumes of data.

A topic of greatest usage of decision support systems (DSS) is the healthcare, where these systems provide relevant medical information regarding the history/condition of patients. Wang *et al.* propose a learning framework to perform the mining of longitudinal heterogeneous event data [5]. The effectiveness of the proposed algorithm is validated with a healthcare dataset and shows that this optimization method can learn event

978-1-5090-1328-9/16/$31.00 ©2016 IEEE

patterns on a group. Tekin *et al.* propose a Web-based expert system to learn about the most relevant context to assign to each patient using DM [6]. This class of algorithms aims to discover the best clinic and expert given a patient's context. Results show that this proposed algorithm is capable to discover the optimal expert and clinic in a specific context. Yang and Kundakcioglu discuss novel opportunities to extract useful information from diverse and heterogeneous data sets in order to make better decisions in the medical field and improve the performance of systems on healthcare [7]. The Big Data concept is motivating a deep transformation on healthcare giving important opportunities for researchers in order to conduct innovative and transformative research, mainly, in DM techniques and health informatics [8]. This paper proposes a model that uses the Naïve Bayesian classifier to evaluate pregnant disorders. This model is evaluated in comparison with the well-known Decision Tree classifier J48 through a real dataset and the results are very promising.

The rest of the paper is organized as follows. Sections II and III address the use of the Naïve Bayes classifier and the J48 Decision Tree classifier. Section IV presents the performance evaluation study and results analysis considering the proposed methods. Finally, Section V provides the conclusion and suggestions for further works.

II. Tree-Based Classifiers on Healthcare

A Decision Tree (DT) is defined as a data structure with leaf nodes that indicate a class or decision nodes, which contains a test on the value of an attribute. Kelarev *et al.* make a comparison study of several methods based on decision trees and propose a novel application of sensor data processing for diabetes patients [9]. For this comparison, the accuracies of six algorithms based on DTs are given and the best performance is found for J48, NBTree, and SimpleCart. Hidayah *et al.* use a dataset from vertebral column to develop an classification model based on DTs [10]. This research joints the J48 decision tree classifier and Bagging algorithms as a classification model. The performance evaluation analysis of this model used 10-fold cross-validation and shown the proposed method improved the J48 classifier. Ludwig *et al.* investigate a fuzzy decision tree algorithm in order to classify gene expression data [11]. This algorithm is compared to classical DM algorithms that are applied for classification. Results of comparing the fuzzy decision tree with the J48 classifier show that this algorithm outperforms the J48 in terms of accuracy in some datasets.

The key factor for the large use of J48 algorithm in DM comes from the fact that it proves to be suitable for procedures involving continuous variables (data) and discrete qualitative, presented in several databases. The J48 algorithm is considered the one with the best results in the decision tree based approaches that use a set of training data. On this study, the J48 algorithm is used as it has a great accuracy rate. Algorithm I shows the proposed approach.

The information gain of an attribute *A* is calculated by the equation (1).

$$gain = info(T) - \sum_{i=1}^{s} \frac{|T_i|}{|T|} \times info(T_i) \tag{1}$$

, where *T* represents a set of cases and T_i (i=1 to s) are subsets of *T* and comprises distinct values for the attribute *A*. The term *info*(*T*) represents the entropy function described in Equation (2), as follows.

$$info(T) = -\sum_{j=1}^{Nclass} \frac{freq(C_j, T)}{|T|} \times \log_2\left(\frac{freq(C_j, T)}{|T|}\right) \tag{2}$$

Algorithm I. Pseudocode for J48 Algorithm.

	J48 Classifier
1:	Create a node N;
2:	**if** (T belongs to same category C) { leaf node=N; mark N as class C; return N; }
3:	**for** i=1 to n { Calculate Information_gain(A_i);}
2:	**if** (T belongs to same category C) { leaf node=N; mark N as class C; return N; }
3:	**for** i=1 to n { Calculate Information_gain(A_i);}
4:	t_a = testing attribute;
5:	N.t_a = attribute having highest information_gain;
6:	**if** (N.t_a==continuous) { find threshold; }
7:	**for** (Each T' in the splitting of T)
8:	**if** (T' is empty) { child of N is a leaf node;}
9:	**else** {child of N= dtree (T') }
10:	calculate classification error rate of node N;
11:	return N;

For choice of the attribute to experiment the current node, it is used the attribute that presents the highest gain information. This approach minimizes the expected number of necessary experiments to classify an object guaranteeing a simpler tree.

III. Bayesian Classifiers on Healthcare

The Naive Bayes classifier is probably the most widely used classifier in Machine Learning. This classifier assumes that attributes are conditionally independent. Despite this method has been considered a simplistic premise, this classifier reports the best performance in various classification tasks. Shaikh *et al.* propose an electronic recording system for heart disease prediction that uses this DM modeling technique [12]. This system is useful to extract hidden knowledge from the health database. Brüser *et al.* study the automatic detection of atrial fibrillation in order to identify cardiac arrhythmias [13]. This study compares seven machine learning algorithms. The Naïve Bayes classifier was slightly better than the more modern Support Vector Machine (SVM) algorithms, but in some classification problems, tree-based algorithms performed well.

The Naïve Bayesian algorithm [14] is presented in Algorithm II.

Algorithm II. PSEUDOCODE FOR NAÏVE BAYES ALGORITHM.

	Naïve Bayes Classifier
1:	training set D. Initialize X with one component;
2:	**if** $P(C_i/X) > P(C_j/X)$ for all $1 \leq j \leq m; j \neq i$; maximize $P(C_i/X)$
3:	compute $P(C_i/X) = \frac{P(X/C_i)P(C_i)}{P(X)}$
4:	$P(X/C_i)P(C_i)$ needs to be maximized;
5:	$P\left(\frac{X}{C}\right) = \prod_{k=1}^{n} P(X_k/C_i)$ $P(X_1/C_i)\times P(X_2/C_i)\times \ldots\times P(X_n/C_i)$ value of attribute A_k, for tuple X;
6:	**if** (A_k= categorical) **then** $P(X/C_i)P(C_i)$ **else** $P(X/C_i) = g(X_k, \mu C_i, \sigma C_i)$
7:	to predicate the class label X, $P(X/C_i)P(C_i)$ $P(X/C_i)P(C_i) > P(X/C_j)P(C_j)$ for all $1 \leq j \leq m; j \neq i$;
8:	output the classifier;

IV. PERFORMANCE EVALUATION

A. Standard metric measurements

The above-presented classifiers are evaluated with a healthcare dataset that includes discrete and categorical attributes. The dataset includes data gathered from experienced physicians organized in a database. In this research the experiments are executed considering the metrics of precision, recall, and F-measure for assess the classifiers performance. Performance metrics are calculated using a predictive classification table, known as Confusion Matrix. Performance evaluation of the classifiers used in the healthcare dataset is analyzed considering the confusion matrix. It uses the common standard indicators for measuring the performance classification of several models.

Precision (Prec.) is the proportion of the predicted relevant data sets that were correct. Recall (Rec.) represents the proportion of data sets that were correctly identified. Finally, the F-measure derives from precision and recall values as shown in Equation (3).

$$F\text{-}measure(\%) = \frac{2\times Rec.\times Prec.}{Rec.+Prec.} \quad (3)$$

This indicator is very significant because it only produces a good result when the Precision and Recall are both equilibrated.

In this research, a Receiver Operating Characteristic (ROC) analysis was also performed. It uses sensitivity and specific indicators. The ROC analysis is a comparison of two characteristics: True Positive Rate (TPR) and False Positive Rate (FPR). The TPR measures the number of relevant classifications that were correctly identified while the FPR measures the percentage of samples misclassified as positive among all real negative.

B. Experimental Results

The healthcare dataset used for classification includes 25 cases of hypertension with four attributes. The categorical attributes considered for classification are shown in Table I.

Table I. ATTRIBUTES CATEGORICAL VALUES OF HYPERTENSIVE DISORDERS.

Naïve Bayes Classifier				
1	Hypertension	Normal	High	Very High
2	Proteinuria	Absence	Traces	Severe
3	Edema	Present		Absence
4	Hypertensive disease (heart or renal)	Present		Absence

Considering the attributes available in Table I, the hypertensive diseases are identified in accordance to the International Statistical Classification of Diseases and Related Health Problems [15]. According to WHO, there are seven hypertensive disorders, as follows: *i*) O10 - Pre-existing hypertension complicating pregnancy, childbirth and the puerperium; *ii*) O11 - Pre-eclampsia superimposed on chronic hypertension; *iii*) O12 - Gestational [pregnancy-induced] edema and proteinuria without hypertension; *iv*) O13 - Gestational [pregnancy-induced] hypertension; *v*) O14 - Pre-eclampsia; *vi*) O15 - Eclampsia; and *vii*) O16 - Unspecified maternal hypertension. In this context, this information was used in both classifiers and the performance measures of standard metrics were considered.

Table II. STANDARD METRICS VALUES OF NAÏVE BAYES CLASSIFIER COMPUTED ON CONFUSION MATRIX PREDICTIVE PARAMETERS.

TP Rate	FP Rate	Prec.	Rec.	F-measure	ROC Area	Class
0.75	0.048	0.75	0.75	0.75	0.917	O10
0.25	0.235	0.333	0.25	0.286	0.71	O11
0.667	0.045	0.677	0.677	0.677	0.955	O13
0.5	0.263	0.375	0.5	0.429	0.746	O14
0.75	0.048	0.75	0.75	0.75	0.72	O15

Table III. STANDARD METRICS VALUES OF J48 DECISION TREE CLASSIFIER COMPUTED ON CONFUSION MATRIX PREDICTIVE PARAMETERS.

TP Rate	FP Rate	Prec.	Rec.	F-measure	ROC Area	Class
0.75	0	1	0.75	0.857	0.863	O10
0.75	0.412	0.462	0.75	0.571	0.585	O11
1	0.045	0.75	1	0.857	0.962	O13
0	0.053	0	0	0	0.759	O14
0.75	0.048	0.75	0.75	0.75	0.786	O15

The 5-fold cross-validation method is also used in this study. It divides the total dataset into five mutually exclusive subsets and, from these, a subset is used for experiments and the remaining four cases are used for training. Then, the precision of the model is calculated. This process is performed five times alternating circular shaped test subset. The performance of the classifiers is presented in Tables II and III.

Figure 1 shows an illustration of the decision tree. It was created from the attributes of considered healthcare dataset.

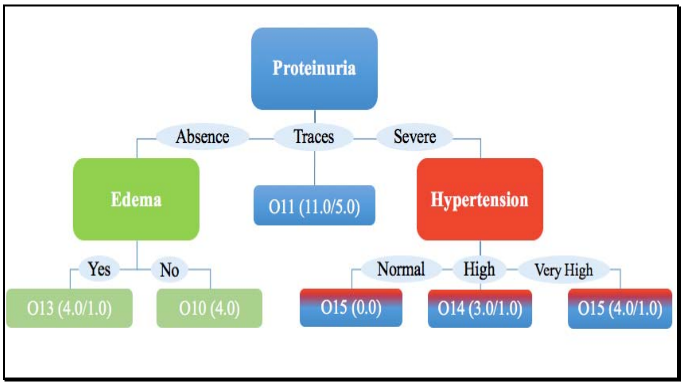

Fig. 1. Illustration of the Tree based on the used healthcare dataset for this study.

Table IV shows the average standard metrics values of both the classifiers on healthcare dataset. These results shown that the J48 tree-based classifier has performed well in all indicators mentioned above, that is, it is a better predictor than the Naïve Bayes classifier on identify positive or negative conditions in patients who have or not certain hypertensive disorder.

Table IV. AVERAGE STANDARD METRICS VALUES OF NAÏVE BAYES AND J48 DECISION TREE CLASSIFIERS.

Standard Metrics	J48 Decision Tree	Naïve Bayes
Prec.	**0.518**	0.517
Rec.	**0.6**	0.52
F-measure	**0.543**	0.514

The ROC analysis with TPR and FPR characteristics is shown in Table V. For the TRP indicator the optimal values are the highest as possible, while the optimal values for FRP indicator are the lowest. Despite the Naïve Bayes classifier to have a greater ROC area, the J48 Decision Tree classifier has a better performance on TPR and FPR indicators, *i.e.*, it identifies positive results that occur among all positive samples correctly and, on the other hand, it also identifies incorrect positive results that occur among all negative samples.

Table V. TRUE POSITIVE RATE AND FALSE POSITIVE RATE.

Standard Metrics	J48 Decision Tree	Naïve Bayes
TPR	**0.6**	0.52
FPR	**0.157**	0.159
ROC Area	0.748	**0.782**

Figure 2 shows the ROC curve. It allows studying the variation of the sensitivity and specificity for different parameters. Models that present curves closest to the point (0,1) are considered optimal models.

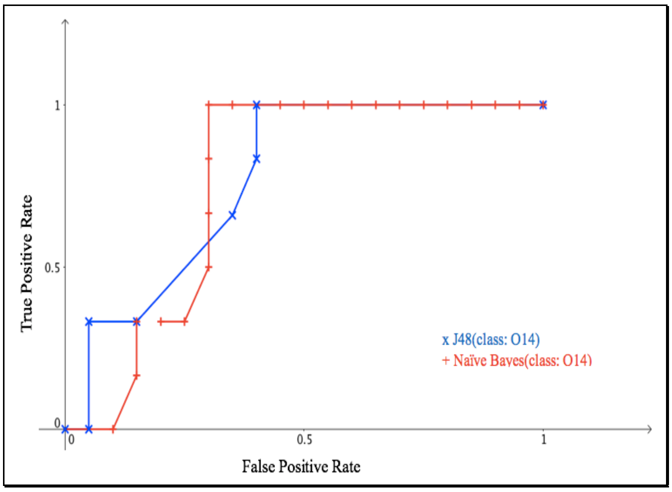

Fig. 2. Performance comparison of ROC curve considering J48 Decision Tree (x) and Naïve Bayes (+) classifiers for Preeclampsia class.

From the ROC analysis of Naïve Bayesian classifier and J48 Decision Tree classifier, it was found that J48 Decision Tree classifier is a more accurate technique than the Naïve Bayes classifier.

Figure 3 shows the performance evaluation of J48 Decision Tree classifier and Naïve Bayes classifiers on standard measures while Figure 4 shows the ROC analysis. This analysis provides a path to select optimal models and to discard suboptimal ones.

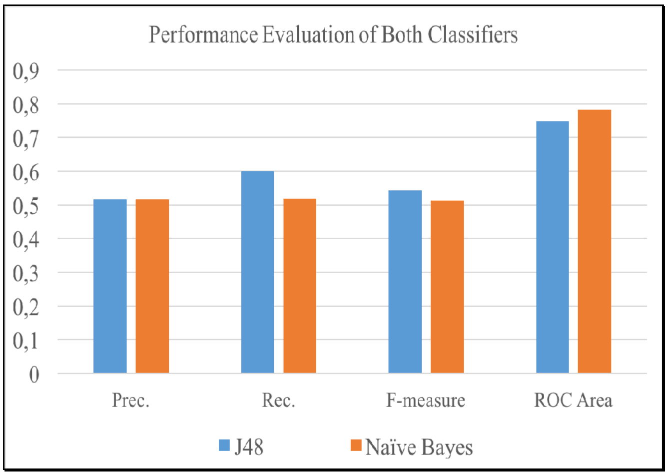

Fig. 3. Performance comparison considering ROC curves of both classifiers on standard measures.

It is noticed that there is a little difference on each performance indicator, but the tree-based J48 classifier has a significant better performance regarding the Recall indicator. Despite the insignificant difference on the performance among the J48 and Naïve Bayes models, the J48 decision tree obtained a better performance in almost all predictors with highest TPR e lower FPR.

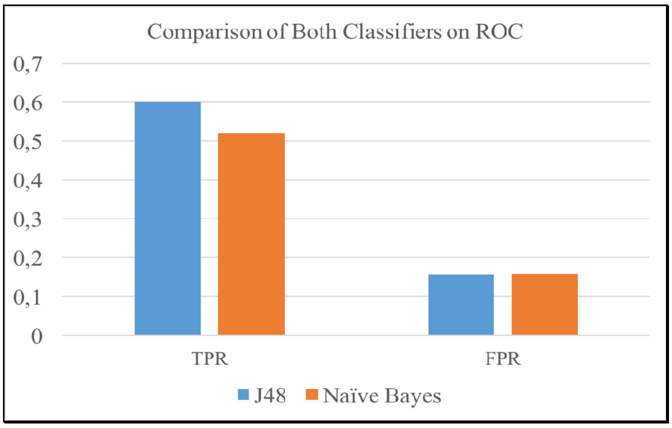

Fig. 4. Receiver Operating Characteristic (ROC) Analysis consiering the True Positive Rate (TPR) and False Positive Rate (FPR) for the J48 and Naïve Bayes classifiers.

V. Conclusion

In this paper the Naïve Bayes and the J48 Decision Tree classifiers were used on a real healthcare dataset to identify several hypertensive disorders. Comparison evaluation of these classifiers was performed, in detail, with confusion matrix and using predictive parameters. The performance evaluation of these two classifiers was analyzed from standard metric measures in order to classify the healthcare dataset following the International Statistical Classification of Diseases and Related Health Problems.

The classification results analyzed in this study shows that J48 Decision Tree classifier is a more accurate technique than the Naïve Bayes classifier. Although the results of these classifiers are very close, both are used as good predictors to decision-making problems. Therefore, the J48 Decision Tree classifier is proposed in medical field for knowledge discovery from the healthcare datasets to support physicians in problems that need more attention.

Further works aim to improve classification accuracy through better and large datasets using other kinds of classifiers. Different types of Bayes-based and Tree-based algorithms will be considered. A study based on Multi-instance learning may also be considered. SVM and K-Neural Networks are also important tools to improve the accuracy of predictive models. Although the diagnosis of hypertensive disorders is complex, many efforts have been carried out in order to improve smart decision support systems that support health experts in uncertain moments (medical diagnosis).

Acknowledgements

Part of this work has been partially supported by the *Instituto de Telecomunicações*, Next Generation Networks and Applications Group (NetGNA), Portugal, by National Funding from the FCT - *Fundação para a Ciência e a Tecnologia* through the UID/EEA/50008/2013 Project, by Finep, with resources from Funttel, Grant No. 01.14.0231.00, under the Radiocommunication Reference Center (*Centro de Referência em Radiocomunicações* - CRR) project of the National Institute of Telecommunications (*Instituto Nacional de Telecomunicações* - Inatel), Brazil, by the National Plan of Science, Technology and Innovation (MAARIFAH), King Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia, Award Number (12-INF2817-02), by *Laboratório de Redes de Computadores* (LAR-A), Aracati, Brazil, and *Ciências sem Fronteiras of Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq), Brazil through the process number 207706/2014-0.

References

[1] X. Wu; X. Zhu; G. Q. Wu; and W. Ding, "Data mining with big data," *IEEE Transactions on Knowledge and Data Engineering,* vol. 26, nº 1, pp. 97-107, jan. 2014.

[2] A. Mukhopadhyay; U. Maulik; S. Bandyopadhyay; and C. A. C. Coello, "A Survey of Multiobjective Evolutionary Algorithms for Data Mining: Part I," *IEEE Transactions on Evolutionary Computation,* vol. 18, nº 1, pp. 4-19, feb. 2014.

[3] C. W. Tsai; C. F. Lai; M. C. Chiang; and L. T. Yang, "Data Mining for Internet of Things: A Survey," *IEEE Communications Surveys & Tutorials,* nº 16, p. 1, first quarter 2014.

[4] K. Kalegele; K. Sasai; H. Takahashi; G. Kitagata; and T. Kinoshita, "Four Decades of Data Mining in Network and Systems Management," *IEEE Transactions on Knowledge and Data Engineering,* vol. 27, nº 10, pp. 2700-2716, oct. 2015.

[5] F. Wang; N. Lee; J. Hu; J. Sun; S. Ebadollahi; and A. F. Laine, "A Framework for Mining Signatures from Event Sequences and Its Applications in Healthcare Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 35, nº 2, pp. 272-285, feb. 2013.

[6] C. Tekin; O. Atan; and M. Van Der Schaar, "Discover the Expert: Context-Adaptive Expert Selection for Medical Diagnosis," *IEEE Transactions on Emerging Topics in Computing,* vol. 3, nº 2, pp. 220-234, jun. 2015.

[7] H. Yang and E. Kundakcioglu, "Healthcare Intelligence: Turning Data into Knowledge," *IEEE Intelligent Systems,* vol. 29, nº 3, pp. 54-68, may-Jun. 2014.

[8] U. Srinivasan and B. Arunasalam, "Leveraging Big Data Analytics to Reduce Healthcare Costs," *IT Professional,* vol. 15, nº 6, pp. 21-28, nov.-dec. 2013.

[9] A. V. Kelarev; A. Stranieri; J. L. Yearwood; and H. F. Jelinek, "Empirical Study of Decision Trees and Ensemble Classifiers for Monitoring of Diabetes Patients in Pervasive Healthcare," em *15th International Conference on Network-Based Information Systems,* Melbourne, VIC, USA, September 26-28, 2012, pp. 441-446.

[10] I. Hidayah; A. Erna P.; and M. A. Kristy, "Application of J48 and bagging for classification of vertebral column pathologies," em *International Conference on Information Technology and Multimedia (ICIMU)*, Putrajaya, Malaysia, November 18-20, 2014, pp. 314-317.

[11] S. A. Ludwig; D. Jakobovic; and S. Picek, "Analyzing gene expression data: Fuzzy decision tree algorithm applied to the classification of cancer data," em *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Istanbul, Turkey, August 2-5, 2015, pp. 1-8.

[12] S. Shaikh; A. Sawant; S. Paradkar; and K. Patil, "Electronic recording system-heart disease prediction system," em *International Conference on Technologies for Sustainable Development (ICTSD)*, Mumbai, India, February 4-6, 2015, pp. 1-5.

[13] C. Bruser; J. Diesel; M. D. H. Zink; S. Winter; P. Schauerte; and S. Leonhardt, "Automatic Detection of Atrial Fibrillation in Cardiac Vibration Signals," *IEEE Journal of Biomedical and Health Informatics,* vol. 17, nº 1, pp. 162-171, jan. 2013.

[14] H. K. S. Doreswamy, "Performance Evaluation of Predictive Engineering Materials Data Sets," *Open Access Library Journal,* vol. 3, nº 3, pp. 1-7, mar. 2011.

[15] World Health Organization, "International Statistical Classication of Diseases and Related Health Problems," vol. 2, nº 10, 2010.