Prediction of mortality rates in heart failure patients with data mining methods

Jan Bohacik[1,2*], C. Kambhampati[1†], Darryl N. Davis[1‡], John G. F. Cleland[3§]

[1]*Department of Computer Science, University of Hull,
Hull, UK*
[2]*Department of Computer Science, University of Zilina
Zilina, Slovakia*
[3]*Department of Cardiology, University of Hull,
Hull, UK*

Abstract – Heart failure is one of the severe diseases which menace the human health and affect millions of people. Half of all patients diagnosed with heart failure die within four years. For the purpose of avoiding life-threatening situations and minimizing the costs, it is important to predict mortality rates of heart failure patients. As part of a HEIF-5 project, a data mining study was conducted aiming specifically at extracting new knowledge from a group of patients suffering from heart failure and using it for prediction of mortality rates. The methodology of knowledge discovery in databases is analyzed within the framework of home telemonitoring. Several data mining methods such as a Bayesian network method, a decision tree method, a neural network method and a nearest neighbour method are employed. The accuracy for the data mining methods from the point of view of avoiding life-threatening situations and minimizing the costs is discussed. It seems that the decision tree method achieves the best accuracy results and is also interpretable for the clinicians.

1   Introduction

Entirely all European countries are faced with an ageing population and so despite improved treatment the number of heart failure patients increases with subsequent large escalation in health care costs. This has led to increasing emphasis on improved

---

[*]J.Bohacik@hull.ac.uk, Jan.Bohacik@fri.uniza.sk

[†]C. Kambhampati@hull.ac.uk

[‡]D.N.Davis@hull.ac.uk

[§]J.G.Cleland@hull.ac.uk

care in the community (at home) in place of expensive hospital care. And so chronic disease management represents the greatest health care challenge in many locales [**1**]. Remote monitoring devices enable patients with serious problems to record their own health measures and send them electronically to clinicians. This keeps them out of doctor's offices for routine care, and thereby helps to reduce health care costs. It is probable that a new increase in heart failure will be observed in the future because in addition to the ageing population a trend leading to an increased prevalence of major heart risk factors such as obesity and diabetes can be observed [**2**]. Only in Europe alone there are over 3.5 million newly diagnosed people every year. At the same time, patients with known heart failure are at high risk of readmission [**3**].

Once a patient is known to suffer from heart failure, it is interesting to predict if the patient could die within a certain time so that an effective prevention can be employed. However, there is a lack of methods that accurately predict the death of a patient with heart failure and also, frequent misclassification is present. Existing clinical methods include the EFFECT Risk Scoring system [**4**], Emergency Heart Failure Mortality Risk Grade (EHMRG) [**5**] or Seattle Heart Failure Model (SHFM) [**6**]. Due to the massive development of information technologies, hospitals have possibilities to collect more and more data about patients in their databases. At the same time, they can employ remote monitoring devices directly at the heart failure patient's home and use them for data collection as well. The collected data can be used for making methods for prediction of mortality rates in the heart failure patient through the process of knowledge discovery in databases defined in [**7**]. Knowledge discovery in databases consists of several steps including data mining where data mining techniques such as classifiers are employed. Findings of this research are intended to contribute to an emerging body of literature, which may suggest that data mining methods outperform traditional epidemiological methods.

The paper is organized as follows. Knowledge discovery in databases is analyzed from the point of view of home telemonitoring and specifics of heart failure patients in Section 2. In Section 3, our heart failure database about 2032 patients is described. Section 4 contains information on employed data mining methods. The performance of the methods is compared and discussed in Section 5. Section 6 concludes the paper.

2 Knowledge discovery in databases

Knowledge discovery in databases is a non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in collected data [**7**]. The goal is to seek new relevant knowledge about an application domain. The domain influences knowledge extraction, including how data is stored, understood, transformed, the algorithms that are applied, evaluation and interpretation of the results etc. It consists of a sequence of steps, each of the steps being based upon the completion of the previous step. The steps are as follows: 1) Selection; 2) Preprocessing; 3)

Transformation; 4) Data mining; 5) Interpretation/Evaluation. A visual representation and the flows in the process are presented in more detail in Fig. 1.
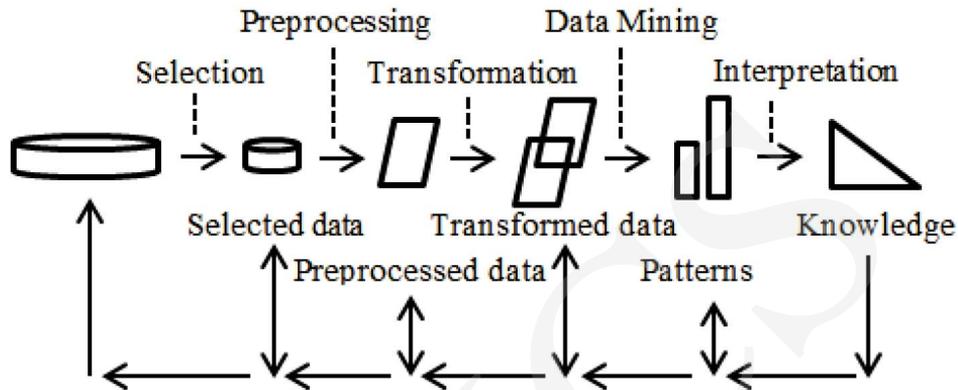

Fig. 1. Knowledge discovery in databases.

When data about heart failure patients are stored in a database, knowledge discovery in heart failure databases can be defined. Specific domain-related issues have to be considered. In our case, monitoring of heart failure patients and prevention of life-threatening situations with prediction of their mortality rates are especially important. As telemonitoring is a promising strategy for improving heart failure outcomes by making it possible to monitor patients remotely so that clinicians can intervene early if there is evidence of clinical deterioration [8], knowledge discovery in databases is likely to be partially based in the patients' homes (see Fig. 2).
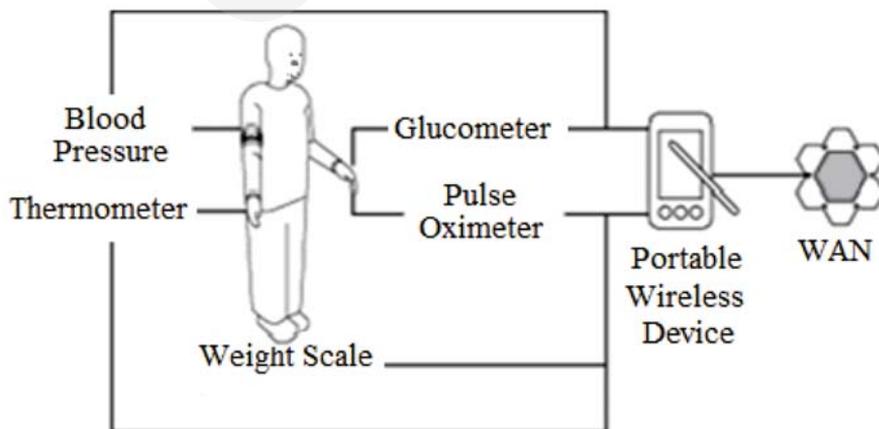

Fig. 2. An example of home telemonitoring.

Data about the patient is collected periodically. However, various pieces of data potentially have different time stamps. For example, blood pressure can be measured

every two hours while the patient's weight can be measured only once a day. After Selection, Preprocessing and Transformation of the data, a database (databases) is (are) given to a data mining method or to a collection of data mining methods whose predictions could be combined. Some of the data mining methods can run in the patient's Portable Wireless Device (such as an Android/iOS/Windows smartphone) and some can be employed directly in a server located in a hospital center. The patient's Portable Wireless Device can also present automatic predictions or predictions influenced by both the data mining methods and the clinicians. Emergency notifications can potentially be sent to the clinicians or emergency services as well. Overall, the potential for application of the process of knowledge discovery in databases in the area of home telemonitoring is enormous.

3   Heart failure database

As a database, a group of 2032 patients $\mathbf{p} \in \mathbf{V}$ classified into two levels of patient status and described by 9 attributes A as queries about clinical findings and physiological measurements is used. Instances are derived from Hull LifeLab which is a large, epidemiologically representative, information-rich clinical database [**9**]. The purpose of Hull LifeLab is studying patients with heart failure so that its definition and diagnosis, its natural history, its mechanisms and markers of progression, the associated costs to health services and society, and the delivery of proven treatment to patients are improved. The attributes in the database and their summary are in Table 1. Described attributes $\mathbf{A}$ are defined as $\mathbf{A} = \{A_1; \ldots; A_k; \ldots; A_9\}$. If A_k is a categorical attribute, $A_k = \{a_(k, 1); \ldots; a_(k, l); \ldots; a_(k, l_k)\}$ where $a_(k, 1); \ldots; a_(k, l); \ldots; a_(k, l_k)$ are possible categorical values. Class attribute C is used to classify instances into two possible categorical values c_1 and c_2 (*alive* and *dead*). It is denoted by $C = \{c_1; c_2\}$.

Table 1. Database description.

Attribute	Data Type	Values
Blood Uric Acid Level (A_1)	Numerical	0.11 - 1.06
Blood Sodium Level (A_2)	Numerical	123 - 148
Blood Creatinine Level (A_3)	Numerical	37 - 1262
NT-proBNP Level (A_4)	Numerical	0.89 - 18236
Age (A_5)	Numerical	27 - 96
Sex (A_6)	Categorical	*female* ($a_{6,1}$)
		male ($a_{6,2}$)
Pulse Rate (A_7)	Numerical	38 - 150
Weight (A_8)	Numerical	29.80 − 193.80
Height (A_9)	Numerical	1.2 − 1.96
Patient Status (C)	Categorical	*alive* (c_1)
		dead (c_2)

Blood Uric Acid Level (A_1) denotes the amount of uric acid in millimoles per liter of the patient's blood. Uric acid is a heterocyclic compound of carbon, nitrogen, oxygen, and hydrogen. It forms when purines break down in the body. If the patient's body produces too much uric acid or if it does not remove enough of it, the patient can get sick. There can be a lack of uric acid caused by low dietary zinc intakes.(A_2) is the amount of sodium in millimoles per litre of the patient's blood. Sodium is the major positive ion in fluid outside of cells. Too much or too little sodium can cause cells to malfunction, and extremes in the blood sodium levels can be fatal. *Blood Creatinine Level* (A_3) is the amount of creatinine in micromoles per liter of the patient's blood. Creatinine is a chemical waste molecule which is generated from muscle metabolism. Since the muscle mass in the body is relatively constant from day to day, the creatinine production remains essentially unchanged on a daily basis. *NT-proBNP Level* (A_4) represents the amount of the N-terminal prohormone of brain natriuretic peptide (NT-proBNP) in picograms per milliliter of the patient's blood. It may be useful to establish prognosis in heart failure as it is typically higher in patients with a higher possibility of heart failure. *Age* (A_5) represents the age of the patient in years. Heart failure risk increases with advancing age and heart failure is also the most common reason for hospitalization in old people. *Sex* (A_6) indicates if the patient is female or male. Men have a higher incidence of heart failure, but the overall prevalence rate is similar in both sexes, since women survive longer after the onset of heart failure [**10**]. *Pulse Rate* (A_7) is the rate of the patient's pulse measured by tactile on the outside of an artery in beats per minute. It is good when the rate of the patient's pulse is low, but when it drops too low, it means the patient's heart has to work harder to pump her/his blood. *Weight* (A_8) denotes the patient's weight in kilograms. High weight is linked to hypertension and an enlarged left ventricle which causes an increased risk for heart failure. Low weight causes the patient's body begins to conserve its energy, which leads to lowered blood pressure and eventual permanent heart damage. *Height* (A_9) is the patient's height in meters. A patient's height is the distance from the bottom of the feet to the top of the head in a human body standing erect. According to [**11**], shorter people are about twice as likely as taller people to experience heart attack. Furthermore, taller men are less likely than shorter men to report a diagnosis of heart failure [**12**]. *Class attribute Patient Status* (C) is used to classify patients into two possible categorical levels of patient status c_1 and c_2 (alive and dead, respectively). It is denoted by $C = c_1; c_2$.

4   Data mining methods

Four principally different data mining methods employed in the prediction of mortality rates in heart failure patients are described here. Concretely, the methods are as follows: a) Bayesian Network method; b) Decision Tree method; c) Neural Network method; and d) Nearest Neighbor method. The Bayesian Network method uses a Bayesian network B over **A** which is a network structure B_S defined as a

directed acyclic graph over describing attributes $\mathbf{A}$ and a set of probability tables $B_P = P(A_k|pa(A_k))|A_k \in \mathbf{A}$ where $pa(A_k)$ is the set of parents $A_k \in \mathbf{A}$ in network structure B_S [13]. The Bayesian network represents a probability distribution $P(\mathbf{A}) = \prod_{A_k \in \mathbf{A}} P(A_k|pa(A_k))$. It is learnt from our group of 2032 patients $\mathbf{p} \in \mathbf{V}$. First the network structure is learnt, then the probability tables are learnt. Prediction of mortality rates in heart failure with the Bayesian network consists of classifying class attribute C given $\mathbf{A}$. Classification is based on calculating $\text{argmax}_{c_j \in C}\{P(c_j|\mathbf{A})\}$ using the probability distribution $P(\mathbf{A})$ represented by the Bayesian network.

The Decision Tree method uses a decision tree learnt from our group of 2032 patients $p \in V$. The decision tree consists of the following nodes: a) the root associated with an $A_k \in \mathbf{A}$; b) internal nodes associated with an $A_k \in \mathbf{A}$; b) leaf nodes associated with a $c_j \in C$. At each node of the tree (except the leaf nodes), the attribute of the data that most effectively splits its set of patients into subsets enriched in *alive* and *dead* is chosen. The splitting criterion is the normalized information gain. The attribute with the highest normalized information gain is chosen to make the decision. Each non-leaf node has an outgoing branch for each possible value $a_{(k,l)} \in A_k$, $A_k \in \mathbf{A}$ is an attribute associated with the node. Numerical attributes $A_k \in \mathbf{A}$ are discretized. The decision tree is pruned after creation and some branches are removed in an attempt to help advance prediction with replacement of branches with leaf nodes. Patients are classified by sorting them based on attribute values for $A_k \in \mathbf{A}$. Each non-leaf node in the decision tree represents an attribute in $\mathbf{A}$ for a patient to be classified, and each branch represents a value that the node can assume. Instances are classified starting at the root node and sorted based on their attribute values [14].

The Neural Network method is a neural network based on multilayer perceptron which focuses on building intelligent codes that mimic the learning mechanism of a human brain by constituting a parallel-connected network model [15]. It consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. With the exception of the input nodes, each node has a nonlinear activation function developed to model the frequency of firing, of biological neurons in the brain. The nodes with the function are called neurons. The input nodes receive the values of attributes in $\mathbf{A}$. There are three or more layers (an input and an output layer with one or more hidden layers) of nodes. Each node in one layer connects to all nodes in the following layer with a certain weight. The neural network is learnt through backpropagation where connection weights are changed after the values of attributes in $\mathbf{A}$ for a patient $\mathbf{p} \in \mathbf{V}$ are processed, based on the amount of error in the prediction of mortality rates compared to the expected rates. Once the neural network is learnt, predictions of mortality rates can be calculated as a functional mapper using the last updated connections weights.

The Nearest Neighbour method assumes our heart failure patients in $\mathbf{V}$ correspond to points in space $R^{M(\mathbf{A})}$, $M(\mathbf{A})$ is the cardinality of $\mathbf{A}$. It creates exemplars on the basis of all patients $\mathbf{p} \in \mathbf{V}$ where exemplars are $M(\mathbf{A})$–dimensional regions covering a finite area of the space [16] and are associated with either class value *alive* or class

value dead. One of the goals is to have much fewer exemplars than $M(\mathbf{V})$, $M(\mathbf{V})$ is the cardinality of $\mathbf{V}$. At the same time, the exemplars do not overlap one another. When patients from $\mathbf{V}$ are added, they may be incorporated into existing exemplars, used for creation of new exemplars or discarded completely if already covered by an exemplar. Exemplars can also be pruned if there could be a conflict with the class value. Exemplars and attributes in $\mathbf{A}$ also have weights which are potentially updated when a patient $\mathbf{p} \in \mathbf{V}$ is added. Growth of the number of exemplars is rapid at first, but decreases over time until the size becomes fairly static, with new added patients resulting in minor changes to, rather than growth of, the number of exemplars. The method predicts the mortality rates for a patient by determining the nearest neighbour in the created exemplars using a modified Euclidian distance function taking exemplar and attribute weights into consideration. Having computed the distance between the patient whose mortality is predicted and all exemplars, the method chooses the class of the closest one. In the event of a tie, it chooses the class with the most exemplars at the minimum distance.

5 Experimental analysis

The main purpose of the experimental study is to compare the performance of the data mining methods with each other. Experiments were carried out with our Java software tool. The core algorithms of the data mining methods were implemented in Weka [**17**]. The performance is measured with sensitivity $= \frac{tp}{tp+fn}$ and specificity $= \frac{tn}{tn+fp}$. In the formulas, $tp/fp/fn/tn$ is the number of true positives/false positives/false negatives/true negatives. "C is *alive*" is considered negative and "C is *dead*" is considered positive. Values tp, fp, fn and tn are computed during 10-fold cross-validation where the dataset is partitioned into 10 folds of patients. The partition is random, but all folds contain roughly the same proportions of alive and dead patients. A patient is considered alive/dead in the dataset if the value assigned to class attribute "C is *alive*"/"C is *dead*". Of the 10 folds, a single fold is retained as the testing dataset for evaluation, and the remaining 9 folds are used as the learning dataset. The learning dataset is analyzed by the data mining method for the purpose of knowledge discovery. The cross-validation process is repeated 10 times, with each of the 10 folds used exactly once as the testing dataset.

The results of our experiments are in Table 2. BNM denotes a Bayesian Network method implemented in Weka as class BayesNet. DTM is a C4.5 decision tree method implemented in Weka as J48. MLPM is a Neural Network method using multilayer perception (a feedforward network) implemented in Weka as class MultilayerPerception. NNM is a Nearest Neighbor method using non-tested generalized examplars implemented in Weka as class NNge. Sensitivity is sensitivity in percentages, Specificity is specificity in percentages and Sum is the sum of Sensitivity and Specificity.

Interpretation of the results in Table 2 in the form of a ROC graph is in Fig. 3. Each tested method has a plot in the ROC graph. The distance from the random guess line

Table 2. Experimental results.

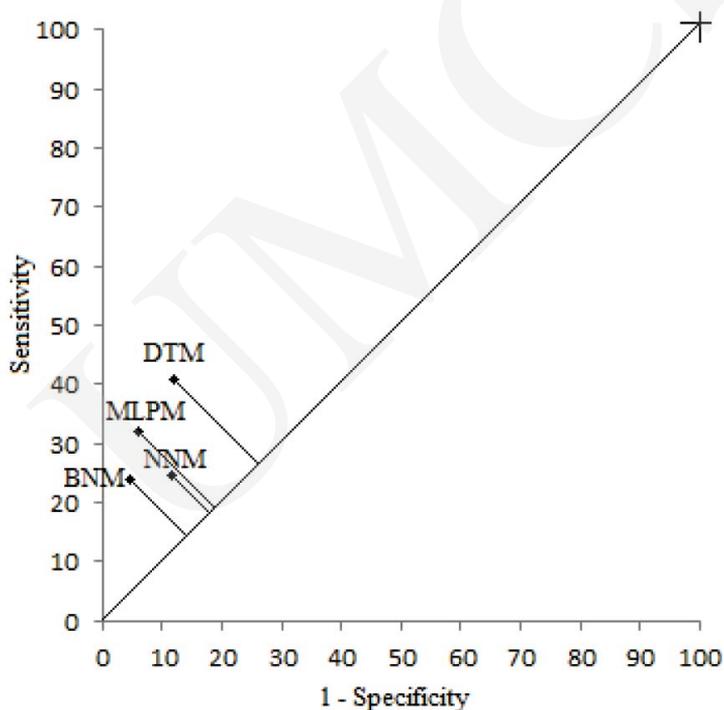| Method	Sensitivity	Specificity	Sum
BNM	24.0385	95.3704	119.4089
DTM	40.9615	87.8968	128.8583
MLPM	32.1154	93.9815	126.0969
NNM	24.6154	88.3598	112.9752

Fig. 3. ROC space for particular methods.

is an indicator of how well the method predicts mortality rates in heart failure patients. A similar indicator is the sum of sensitivity and specificity. It is very important to avoid classification of dead patients as alive as it would lead to life-threatening situations. On the other hand, many alive patients classified as dead ones would increase the running costs of the treatment considerably. The best performance is achieved by the C4.5 decision tree method (DTM) sensitivity 40.9615% and specificity 87.8968

6 Conclusions

Knowledge discovery in databases and specifics for monitoring of heart failure patients were described. Several data mining methods were employed on our Hull LifeLab data within the process of knowledge discovery in databases. Our Hull LifeLab data contained 2032 patients described by 9 attributes and classified into *alive* and *dead*. Class values *alive* and *dead* were used for predicting mortality rates of heart failure patients. The following four principally different data mining methods were employed: a) Bayesian Network method; b) Decision Tree method; c) Neural Network method; and d) Nearest Neighbour method. The accuracy of the data mining methods was evaluated from the point of view of avoiding life-threatening situations and minimizing the costs. Life-threating situations appear when patients classified as *dead* are considered *alive*, which is measured by sensitivity. These situations should be minimized and so sensitivity should be maximized. Costs are increased when *alive* patients are treated as if they were classified as *dead*, which is measured by specificity. For the costs to be minimized specificity should be maximized. For our Hull LifeLab data, the best results were achieved by the Decision Tree method. Concretely, its sensitivity was 40.9615% and its specificity was 87.8968%. The Decision Tree method was also easily interpretable for the clinicians. However, the understandability of the decision tree was not optimal. The divisions of the numerical attributes in discretization were not according to the understanding of the clinicians. This could be changed through the adoption of the notions of fuzzy logic and fuzzification where the numerical attributes would be transformed by the clinicians. And then a decision tree with comparable accuracy could be created, which is a part of our future research. Also, the accuracy of the data mining methods should be improved. It seems real-world medical data contains lots of noise and uncertainties that have to be addressed, possibly through the adoption of fuzzy logic.

Acknowledgements

References

[1] West D., How mobile devices are transforming healthcare, Issues in Technology Innovation (2012).

[2] Lopez-Sendon J., The heart failure epidemic, Medicographia 33(4) (2011): 363.

[3] Vinson J. M., Rich M. W., Sperry J. C., McNamara T. C., Early readmission of elderly patients with congestive heart failure, Journal of the American Geriatrics Society 38(12) (1990): 1290.

[4] Lee D. S., Austin P. C., Rouleau J. L., Liu P. P., Naimark D., Tu J. V., Predicting mortality among patients hospitalized for heart failure: derivation and validation of a clinical model, JAMA 290(19) (2003): 2581.

[5] Lee D. S., Stitt A., Austin P. C., Stukel T. A., Schull M. J., Chong A., Newton G. E., Lee J. S., Tu J. V., Prediction of heart failure mortality in emergent care: a cohort study, Annals of Internal Medicine 156(11) (2012): 767.

[6] Ketchum E. S., Jacobson A. F., Caldwell J. H., Senior R., Cerqueira M. D., Thomas G. S., Agostini D., Narula J., Levy W. c., Selective improvement in Seattle Heart Failure Model risk stratification using iodine-123 metaiodobenzylguanidine imaging, Journal of Nuclear Cardiology 19(5) (2012): 1007.

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

[8] Chaudhry S. I., Mattera J. A., Curtis J. P., Spertus J. A., Herrin J., Lin Z., Phillips C. O., Hodshon B. V., Cooper L. S., Krumholz H. M., Telemonitoring in patients with heart failure, The New England Journal of Medicine 363(24) (2010): 2301.

[9] Clinical Effectiveness and Evaluation Unit of the Royal College of Physicians, Managing chronic heart failure, learning from best practice, The Lavenham Press Ltd, UK (2005).

[10] Stromberg A., Martensson J., Gender differences in patients with heart failure, European Journal of Cardiovascular Nursing 2(1) (2003).

[11] Walker M., Shaper A. G., Phillips A. N., Cook D. G., Short stature, lung function and risk of a heart attack, International Journal of Epidemiology 18(3) (1989): 602.

[12] Akinkuolie A. O., Aleardi M., Ashaye A. O., Gaziano J. M., Djousse L., Height and risk of heart failure in the Physicians' Health Study, American Journal of Cardiology 109 (7) (2011): 994.

[13] Bouckaert R. R., Bayesian Network Classifiers in Weka for Version 3-5-7, University of Waikato, New Zealand (2008).

[14] Kotsiantis S. B., Supervised machine learning: a review of classification techniques, Informatica 31 (2007): 249.

[15] Cakır A., Demirel B., A software tool for determination of breast cancer treatment methods using data mining approach, Journal of Medical Systems 35(6) (2011): 1503.

[16] Martin B., Instance-based learning: Nearest neighbour with generalization, University of Waikato, New Zealand (1995).

[17] Witten I. H., Frank E., Hall M. A., Practical machine learning tools and techniques (3rd edition), Morgan Kaufman Publishers, USA (2011).