Analysis of the Fuzzy Unordered Rule Induction Algorithm as a Method for Classification

Elaine Anita de Melo Gomes Soares[1], Lecidamia Cristina Leite Damascena[1], Luciana Moura Mendes de Lima[1], and Ronei Marcos de Moraes[1]

Graduate Program on Decision Models and Health - Federal University of Paraiba, Joao Pessoa, PB, Brasil

Abstract. Fuzzy Unordered Rule Induction Algorithm (FURIA) is a recent algorithm, proposed by Huhn and Hullermeier, responsible for the creation of fuzzy logic rules from a given database, and for the classification of it using the generated rules. This paper means to analyze the effectiveness of the FURIA as classification method apllied in different contexts. In order to do it, different databases from the Center for Machine Learning and Intelligent Systems repository were chosen and tested with Weka. It was found that for databases with a higher quantity of instances, quantitative or qualitative, this algorithm presented better performance; and in most of the cases, it resulted in a good agreement coefficient.

Keywords: fuzzy logic, classification, performance evaluation

1 Introduction

Fuzzy logic algorithms are known by virtue of their characteristics, such as rules of classification being easily understood by the reader, having the ability to process linguistic data, allowing expert opinion. Furthermore they can be used a as method for classification tasks [4]. Given that some information needed to solve classification problems is uncertain, the use of classical logic presents a challenge for these problems. Uncertainty can not be quantified by simple forms of probability and requires the use of more effective methods. The use of fuzzy logic assists in the processing of data with a certain degree of uncertainty. Fuzzy logic introduces the notion of gradual transition of elements between existing sets in the real world, transcending binary ”yes-no” logic [10]. This has been used in a number of areas, such as health [19], neurosciences [6], veterinary sciences [25], in the development of various types of computer systems, transport controllers, consumption products [10], among other utilities.

This diverse applicability of fuzzy logic extends the possibility of understanding certain phenomena that could not be explained by classical logic. In this study, the fuzzy logic will be applied using the Fuzzy Unordered Rule Induction Algorithm (FURIA) for classification. The FURIA algorithm, proposed by Huhn and Hullermeier [9], since it is based on fuzzy rules, models the decision limits by making them more flexible, with fuzzy intervals, generating non-ordered rules in place of the regular list of rules. There are few studies using this method. The present study aims to analyze the effectiveness of the FURIA algorithm as method classification in different types of databases.

2 Theoretical basis

2.1 Fuzzy Sets

The term fuzzy was proposed by Zadeh (1965) in order to process imprecise information, defining a fuzzy set as a class of objects with continuous degrees of association. This set is characterized by pertinence functions, which are assigned to each object of the set and which are defined in the interval [0, 1]. Hence, let X be a space of points, being x a generic element of X. A fuzzy set A in X is characterized by a membership function μ_A (x), which associates a real number in the interval [0,1] to each point in X, where μ_A (x) represents the degree of pertinence of x in A. Thus, $\mu_A(x) : [0, 1]$.

In fuzzy sets, it is possible to perform intersection and union operations as well as in classical logic, which can be represented by a conjunction (AND) and a disjunction (OR), respectively [14].

2.2 Fuzzy Logic and a System Based in Fuzzy Rules

Fuzzy logic consists of a specific type of logic that allows to deal with imprecision and uncertainties. One of the advantages of fuzzy logic is the ability to offer solutions to complex issues in all areas of life, as it resembles logical reasoning, aiding in classification process. The laborious development of the fuzzy rules can be mentioned as a disadvantage, as well as functions of pertinence and complex analysis of the outputs and generating numerous interpretations. In addition, it requires accuracy to construct a fuzzy system and a large amount of data is required [6].

The way to express knowledge in this logic is commonly through the rules of the condition-action type. Thus, a fuzzy rule is a unit capable of capturing some specific knowledge, while a set of fuzzy rules is able to describe a system in its different possibilities. Therefore, each rule is composed of an antecedent part (if) and a consequent part (then). The preceding part describes a condition being composed of the input variables and the consequent part describes a conclusion by the output variables [14]. For example:

IF (condition-1 OR condition-2) AND (condition-3) THEN (conclusion-1 AND conclusion-2)

Generally, a fuzzy rule is of the type:

IF (x is a_i) AND (y is b_i) OR ... THEN (z is c_i) AND (w is d_i)

where x and y are input linguistic variables, z and w are output linguistic variables and a_i, b_i, c_i and d_i are realizations of these variables, measured in the user's interaction with the system.

It is worth mentioning that the linguistic variable has a qualitative value expressed by a linguistic term (words or phrases) and quantitatively by its pertinence function. One of the main reasons for the successful application of this logic in intelligent systems is the ability to combine numeric symbolic variables [14].

3 Methodology

To analyze the efficacy of this algorithm, ten different public data banks were used, these being: breast cancer, vote, zoo, hypothyroid, postoperative patient data, arrhythmia, diabetes, ecoli, dermatology, nursery, from the repository Center for Machine Learning and Intelligent Systems.

3.1 FURIA

FURIA is a recent algorithm responsible for generating fuzzy logical rules from a user-supplied database, and sorting using generated rules [9]. For the better execution of FURIA, the Waikato Environment for Knowledge Analysis (WEKA) software allows the customization of some variables of the algorithm. The choice between product and minimum of t-norm is offered, which is used by this algorithm to implement the concepts of fuzzy in logical rules. It presented effective results when compared to other statistical methods [9].

FURIA is an advanced derivative of the RIPPER algorithm. It produces fuzzy rules instead of conventional rules in order to model more flexible classification boundaries. The fuzzy rules are generated by replacing fuzzy intervals through a trapezoidal relevance function in combination with the sophisticated rule induction technique employed by the original RIPPER algorithm [23].

It presents an advantageous feature which is the elongation of the rule. The elongation consists of the generalization of the rules covering any possibility. It is a local strategy that explores information in the vicinity of the query. The minimum generalization of a rule is obtained simply by excluding all antecedents that are not satisfied by the query [9]. The pseudo-code for a single rule r is presented below [9].

Confusion matrix, Kappa coefficient and its confidence intervals were the measures used to evaluate the results generated by FURIA.

Let A be the set of numeric antecedents of r
while $A \neq \emptyset$ **do**
    $a_{max} \leftarrow$ null (a_{max} denotes the antecedent with the highest purity)
    $pur_{max} \leftarrow 0$ (pur_{max} is the highest purity value, so far)
    for $i \leftarrow 1$ **to** *size(A)* **do**
        compute the best fuzzification of $A[i]$ in terms of purity
        $pur_{A[i]} \leftarrow$ be the purity of this best fuzzification
        if $pur_{A[i]} > pur_{max}$ **then**
            $pur_{max} \leftarrow pur_{A[i]}$
            $a_{max} \leftarrow A[i]$
        end
    end
    $A \leftarrow A \backslash a_{max}$ Update r with a_{max}
end

Algorithm 1: Generation of a single rule r [9].

3.2 Confusion Matrix

The confusion matrix is used to measure the accuracy of the model classification. A simple way to measure the percentage of correct classifications is from the sum of the values in the main diagonal of the matrix, divided by the sum of all values of the matrix [3]. In Table 1, it is possible to observe the confusion matrix generated for the Nursery database.

Table 1. Matrix of confusion generated by WEKA software for the database Nursery.

a	b	c	d	e	← classified as
852	0	0	0	0	a = not_recom
0	0	0	0	0	b = recommend
0	0	47	22	0	c = very_recom
0	0	4	810	49	d = priority
0	0	0	5	803	e = spec_prior

3.3 Kappa Coefficient

Cohen [2] presented the Kappa Coefficient as a measure of the agreement between the hit and error rates achieved, using the confusion matrix. Its calculation proceeds as follows.

$$K = \frac{P_0 - P_c}{1 - P_c} \tag{1}$$

where P_0 and P_c are given by:

$$P_0 = \frac{\sum_{i=1}^{M} n_{ii}}{N} \quad and \quad P_c = \frac{\sum_{i=1}^{M} n_{i+}n_{+i}}{N^2} \tag{2}$$

where n_{ii} is the sum of the main diagonal of the confusion matrix, n_{i+} is the sum of all values of the line i and n_{+i} is the sum of the values of column i, M is the total number of classes and N is the value of possible classifications present in the classification matrix.

The variance of the Kappa Coefficient K, denoted by σ_K^2 is described by Moraes and Machado [15] as:

$$\sigma_K^2 = \frac{P_0(1-P_0)}{N(1-P_c)^2} + \frac{2(1-P_0)(2P_0P_c-\theta_1}{N(1-P_c)^3)} + \frac{(1-P_0)^2(\theta_2-4P_c^2)}{N(1-P_c)^4} \tag{3}$$

where θ_1 and θ_2 are given by:

$$\theta_1 = \frac{\sum_{i=1}^{M} n_{ii}(n_{i+}+n_{+i})}{N^2} \quad and \quad \theta_2 = \frac{\sum_{i=1}^{M} n_{ii}(n_{i+}+n_{+i})^2}{N^3} \tag{4}$$

Additionally, the 95% confidence interval can be obtained using the following [21]:

$$IC_{1-\alpha}(K) = \left[K - Z_{\frac{\alpha}{2}}\sqrt{\sigma_K^2}, K + Z_{\frac{\alpha}{2}}\sqrt{\sigma_K^2}\right] \tag{5}$$

with $\alpha = 0.05$. This interval represents the range of possible values for the Kappa coefficient given the unpredictability of the data.

$$\theta_1 = \frac{\sum_{i=1}^{M} n_{ii}(n_{i+}+n_{+i})}{N^2} \quad and \quad \theta_2 = \frac{\sum_{i=1}^{M} n_{ii}(n_{i+}+n_{+i})^2}{N^3} \tag{6}$$

According to Landis and Koch[13], the closer to 1, the greater the degree of agreement of the classification model, as shown in Table 2.

Table 2. Degree of Agreement for Kappa Coefficient [13].

Kappa Coefficient	Degree of Agreement
<0.0	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

3.4 WEKA

As already mentioned, the WEKA software was used, version 3.8, developed at the University of Waikato, New Zealand. It is open source and has several algorithms of classifiers (J48, JRip, Decision Table, Naive Bayes, among others) and regression [1].

An important feature is how the test database is defined. WEKA has four options, which are the use of a test database, training database, cross-validation and percentage split. Since the databases used in this work do not have available test databases, only the last three options are available. Several values were used for cross-validation, from 2 to 40 and 70, 80, 90% percentage split were used.

Cross-validation reserves a certain amount of testing and uses the rest for training. The data is randomly divided into a number of equal parts, ten parts for example, and the class has approximately the same proportions in each stratum as in the complete database. Each part is tested and the other nine are trained. In this way, the learning process is performed ten times in the various training sets. The percentage split performs part of the database for testing and the remaining part for training.

4 Results

In order to obtain the best results, the authors executed the classification of the following databases in many different ways. Various parameters were changed, both from the method and the setting of the database (training, cross-validation, percentage split) with different values for each case. After all these exhaustive runs, the ones with the best results were selected and they are presented in the following.

4.1 Arrhythmia

This database was created for the work of Guvenir [7], which presented a supervised machine learning algorithm to determine the type of arrhythmia from the electrocardiogram exam results [7]. This base has 279 attributes, being 206 numerals and 73 nominal, and 16 classes, which represents the possibility of having no arrhythmia, one of the 14 types of arrhythmia or a extra class for unclassified cases. Although it has 16 classes, the database only has elements of 13 classes. Given the relevance of the variables from the electrocardiogram, it was not possible to withdraw any of the attributes present in this bank.

Using 10-fold cross validation, 68.5841% were classified correctly, with 310 instances correctly classified and 142 incorrectly classified, few type II errors were present in the confusion matrix (8 false negatives), and 25 rules were generated. Resulting on a Kappa coefficient of 0.4683 with variance of 9.6890×10^{-4}. Additionally, the 95% confidence interval for this database's Kappa was $[0.4073, 0.5293]$ and, despite the absence of data in the bank, it presented a substantial agreement index.

4.2 Breast Cancer

This bank was obtained from the Ljubljana University Medical Center, Institute of Oncology, Yugoslavia referring to the occurrence of breast cancer [26]. This data set has 286 instances, nine nominal attributes and two classes (recurrence or not of the event).

Using 11-fold cross-validation, 75.8741% were correctly classified, with 217 instances ranked correctly and 69 ranked incorrectly, and four rules were generated. This base resulted a Kappa coefficient of 0.2617 with variance of 0.0001, a fair agreement was observed. In addition, the 95% confidence interval for Kappa was $[0.2360, 0.2874]$.

4.3 Dermatology

This bank was produced by the Gazi University School of Medicine, Department of Computer Engineering and Information Science, Bilkent University, both in Ankara, Turkey (1998). It presents 366 instances, 34 attributes of which 33 are numerical and only one nominal, with the objective of performing the differential diagnosis of erythematous-squamous diseases in six classes [8].

Using 13-fold cross-validation, 95.6284% were classified correctly, with 350 instances correctly and 19 incorrectly, generating 24 rules. Resulting in a Kappa coefficient of 0.9450 with variance of 0.0001. In addition, the 95% confidence interval for Kappa was $[0.9188, 0.9712]$, there was almost perfect agreement.

4.4 Diabetes

This database was developed by the National Institute of Diabetes and Digestive and Kidney Diseases and made available by Johns Hopkins University in 1990 to predict the onset of diabetes mellitus in a high-risk population [22]. It presents 768 instances, eight attributes (all numeric), which were classified into two categories:positive test for diabetes and negative test for diabetes.

With all the variables present and 15-fold cross-validations, five rules were generated. As a result a Kappa coefficient of 0.4529 with variance of 0.0002 and confidence interval $[0.4235, 0.4823]$ was obtained. The percentage of correct classifications was 75.5208% with 580 instances correctly classified e 188 misclassified, showing moderate agreement for this data.

4.5 Ecoli

This database was used by Nakai and Kanehisa [16] in their work on the creation of a probabilistic classification system for predicting the location of proteins in cells. This set has 336 instances, eight numerical attributes and a total of eight classes representing the cell areas.

As a result of this database, 20 rules were generated, through which the classification was performed using 15 folds cross-validation. From this, a Kappa coefficient of 0.8007 was achieved with variance of 6.559×10^{-4} and 95% confidence interval of $[0.750465, 0.85086]$. Additionally, a hit rate of 82.2% was obtained, with 288 instances correctly classified and 48 erroneously. The confusion matrix presented a greater number of errors in the outer membrane, not being such errors relevant. Based on the obtained Kappa, we can affirm that this algorithm obtained a substantial index of agreement.

4.6 Hypothyroid

This base includes data for thyroid disease provided by the Garavan Medical Research Institute and J. Ross Quinlan, Sydney, Australia [18]. This set of data has 3772 instances, 29 variables between nominal and numerical, which decision attribute had four classes (negative, primary hypothyroidism, secondary hypothyroidism, compensated hypothyroidism).

Using 8-fold cross-validation, the FURIA method correctly classified 3756 instances with a percentage of 99.5758 of correct classifications, generating 18 rules. It presented a Kappa coefficient of 0.9708 with a variance of 2.24984×10^{-5}, which corresponds to an almost perfect agreement, with a confidence interval of 95% for Kappa $[0.9615, 0.9801]$.

4.7 Nursery

This base was created by Vladislav Rajkovic et al. (1997) and derived from a hierarchical decision model idealized to categorize childcare requests. The outcome depended on three conditions of work of the parents and the child's nursery, family structure and financial capacity, social and family health. It has 12960 instances and eight attributes [17].

A 8 folds cross-validation test was used and 293 rules were obtained, which classify the need to install day-care centers in 5 groups, ranging from not recommended to special priority. As a result, 97.3071% of the instances were correctly classified (12611) and only 2.6929% (349) were misclassified. The Kappa result was 0.9546 with variance of 7.8942×10^{-6}, with a confidence interval on $[0.9491, 0.9601]$, presenting almost perfect agreement.

4.8 Postoperative Patient Data

This database aims to determine where the patients in a postoperative recovery area should be sent, since hypothermia is a significant concern following surgery [24]. This data set has 90 instances, eight nominal attributes, which are classified on three classes (refer to the Intensive Care Unit, home or to walk in the Hospital).

Using 14-fold cross-validation, 71.1111% were classified correctly and three rules were generated. It showed a Kappa coefficient of 0.0487 with variance of 0.02474329, which represents a slight concordance agreement. In addition, the 95% confidence interval was $[0, 0.3571]$.

4.9 Vote

This data set includes the vote of each congressional representative in the 98th Congressional Quarterly Almanac [20], in the United States on the 16 main themes. This data set has 435 instances, 16 nominal attributes and attribute with two classes (Republican or Democrat).

Using 5-fold cross-validation, 96.3218% were classified correctly, with 419 instances correctly and 16 incorrectly, obtaining seven rules. A Kappa coefficient of 0.9220 with variance of 0.0003 was obtained, and its 95% confidence interval of $[0.8847, 0.9594]$, presenting almost perfect agreement.

4.10 Zoo

This database was created by Richard Forsyth and provided by Richard S. Forsyth, Mapperley Park, Nottingham, 1990. It presents 101 instances, 18 attributes (two numerical and 16 nominal).

With 4-fold cross-validation, this test resulted on a Kappa coefficient of 0.8664 with variance of 0.0015 and confidence interval of $[0.7901, 0.9428]$. Additionally, 91 (90.0990%) instances were correctly classified and 10 (9,9010%) incorrectly classified, presenting a almost perfect agreement for the analyzed data.

4.11 Analysis and discussion of results

Table 3. Summary of the results presented on this section.

Database	Size	# of Attributes	Test Type	# of Rules	Kappa	Hit Rate
Arrhythmia	452	279	10-fold	25	0.4683	68.5841%
Breast Cancer	286	9	11-fold	4	0.2617	75.8741%
Dermatology	366	34	13-fold	24	0.9451	95.6284%
Diabetes	768	8	15-fold	5	0.4529	75.5208%
Ecoli	336	7	15-fold	25	0.8007	85.7143%
Hypothyroid	3772	29	8-fold	18	0.9709	99.5758%
Nursery	12960	8	8-fold	293	0.9546	97.3071%
Postop. Patient Data	90	8	14-fold	3	0.0488	71.1111%
Vote	435	16	5-fold	7	0.9221	96.3218%
Zoo	101	17	4-fold	6	0.8664	90.0990%

The databases that had a greater number of instances, such as hypothyroid and nursery, presented the best classification results, consequently good kappa coefficient and rate of correctness. With the confidence interval, the randomness of the results can be observed. It is worth mentioning that six out of ten banks resulted in a Kappa coefficient with substantial or almost perfect agreement. All the final results, as well as the characteristics of the databases, can be visualized in Table 3.

Regarding the type of attribute, numerical and nominal, one can influence the results, but this is not a fact that can be confirmed based on the databases analyzed in this paper. Additionally, FURIA presents a good performance for problems with multiple classes. This was observed in data sets of three classes in an experiment performed by Huhn and Hullermeier [9], and confirmed here by the Ecoli and Dermatology databases.

The applicability of FURIA occurs in several areas of knowledge, responding to complex questions such as the accuracy of a medical diagnosis or simple issues of belonging to an animal class. Gasparovica and Aleksejeva [4] used FURIA to compare a few types of cancer (gastric, breast, prostate), gastric intestinal disease and healthy individual data sets. As a result a good classification accuracy was obtained, especially when the missing values were replaced by mean values of the same class. This algorithm has also been used in the categorization of text [11], in the evaluation of cardiac arrhythmia evidencing a hit rate of 92,12 [12], and in the identification and confirmation of cases of coronary artery disease [23].

Few studies were found regarding the subject, but those who studied pointed out that FURIA presented effective results when compared to other statistical methods [9]. Experiments demonstrate that FURIA outperforms the original RIPPER and classifiers such as C4.5 significantly relative to the classification accuracy [12]. The FURIA has the advantage of using the rules elongation that supports the classification of new records, previously unseen, guaranteeing greater classification precision [4].

The tests used for the database were cross-validation. Cross-validation generated some good results, as is also observed in a study with Leukemia [5]. The same authors affirm that cross-validation assists in generating better rules and records classified in a more correct way. The same authors affirm that cross-validation assists in generating better rules and records classified in a more correct way. Using k-folds (two to ten), FURIA presented scores above 81% [23].

5 Conclusion

This paper had as goal to analyze accuracy of the FURIA as a classification method, when apllied in different contexts. Several different databases from UCI were used in order to provide it. FURIA presented more than 80% of correct classification rate for more than 50of the databases. It was observed that its best applicability is directed to the databases with high number of instances or high number of classes.

In this investigation the best results were presented through exhaustive attempts, and it was possible to verify that the impact of the customization of the variables of the existing method is minimal. The use of nominal and numerical attributes may influence the elaboration of results. However, it was not evidenced in this study.

As future works, an evaluation about influence of nominal and/or numerical attributes in the final results deserves attention. Additionally, a comparison between FURIA and other classification methods would be enriching.

Acknowledgments

This project is partially supported by CAPES. It is partially supported also by grants 132170/2017-5 and 308250/2015-0 of the National Council for Scientific and Technological Development (CNPq) and is related to the National Institute of Science and Technology Medicine Assisted by Scientific Computing (465586/2014-4) also supported by CNPq.

References

1. Bouckaert, R.R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., Scuse, D.: Weka manual for version 3-7-8. Hamilton, New Zealand (2013)
2. Cohen, J.: A coefficient of agreement for nominal scales. educational and psychosocial measurement, 20, 37-46 (1960)
3. Foody, G.M.: Status of land cover classification accuracy assessment. Remote sensing of environment 80(1), 185–201 (2002)
4. Gasparovica, M., Aleksejeva, L.: Using fuzzy unordered rule induction algorithm for cancer data classification. Breast cancer (BrCa) 13, 1229 (2011)
5. Gasparovica, M., Tuleiko, I., Aleksejeva, L.: Influence of membership functions on classification of multi-dimensional data. Scientific Journal of Riga Technical University. Computer Sciences 45(1), 78–84 (2011)
6. Godil, S., Shamim, M., Enam, S., Qidwai, U.: Fuzzy logic: A" simple" solution for complexities in neurosciences? Surgical neurology international 2, 24 (2011)
7. Guvenir, H.A., Acar, B., Demiroz, G., Cekin, A.: A supervised machine learning algorithm for arrhythmia analysis. In: Computers in Cardiology 1997. pp. 433–436. IEEE (1997)
8. Güvenir, H.A., Demiröz, G., Ilter, N.: Learning differential diagnosis of erythemato-squamous diseases using voting feature intervals. Artificial intelligence in medicine 13(3), 147–165 (1998)
9. Hühn, J., Hüllermeier, E.: Furia: an algorithm for unordered fuzzy rule induction. Data Mining and Knowledge Discovery 19(3), 293–319 (2009)
10. Jensen, R., de Moraes Lopes, M.H.B.: Enfermagem e lógica fuzzy: uma revisão integrativa. Revista Latino-Americana de Enfermagem 19(1), 195–202 (2011)
11. Karabulut, M.: Fuzzy unordered rule induction algorithm in text categorization on top of geometric particle swarm optimization term selection. Knowledge-Based Systems 54, 288–297 (2013)
12. Kumari, V., Kumar, P.R.: Fuzzy unordered rule induction for evaluating cardiac arrhythmia. Biomedical Engineering Letters 3(2), 74–79 (2013)
13. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. biometrics pp. 159–174 (1977)
14. Massad, E., ORTEGA, N.R., SILVEIRA, P.S.: Métodos quantitativos em medicina. Editora Manole Ltda (2004)

15. Moraes, R.M., Machado, L.S.: Psychomotor skills assessment in medical training based on virtual reality using a weighted possibilistic approach. Knowledge-Based Systems 70, 97–102 (2014)
16. Nakai, K., Kanehisa, M.: A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics 14(4), 897–911 (1992)
17. Olave, M., Rajkovic, V., Bohanec, M.: An application for admission in public school systems. Expert Systems in Public Administration 1, 145–160 (1989)
18. Quinlan, J.R., Compton, P.J., Horn, K., Lazarus, L.: Inductive knowledge acquisition: a case study. In: Proceedings of the Second Australian Conference on Applications of expert systems. pp. 137–156. Addison-Wesley Longman Publishing Co., Inc. (1987)
19. de Sá, L.R., de Almeida Nogueira, J., de Moraes, R.M.: Modelo de decisão sobre o perfil demográfico para o controle da tuberculose usando lógica fuzzy. Revista Eletrônica de Enfermagem 17(2), 223–37 (2015)
20. Schlimmer, J.: United states congressional voting records database, congressional quarterly almanac. In: 98th Congress, 2nd session. vol. 40 (1984)
21. Sheskin, D.: Handbook of Parametric and Nonparametric Statistical Procedures: Third Edition. CRC Press (2003), https://books.google.com.br/books?id=bmwhcJqq01cC
22. Smith, J.W., Everhart, J., Dickson, W., Knowler, W., Johannes, R.: Using the adap learning algorithm to forecast the onset of diabetes mellitus. In: Proceedings of the Annual Symposium on Computer Application in Medical Care. p. 261. American Medical Informatics Association (1988)
23. Verma, L., Srivastava, S., Negi, P.: A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data. Journal of Medical Systems 40(7), 1–7 (2016)
24. Woolery, L., Grzymala-Busse, J., Summers, S., Budihardjo, A.: The use of machine learning program lers-lb 2.5 in knowledge acquisition for expert system development in nursing. Computers in nursing 9(6), 227–234 (1990)
25. Zaninelli, M., Tangorra, F.M., Costa, A., Rossi, L., DellOrto, V., Savoini, G.: Improved fuzzy logic system to evaluate milk electrical conductivity signals from on-line sensors to monitor dairy goat mastitis. Sensors 16(7), 1079 (2016)
26. Zwitter, M., Soklic, M.: Breast cancer data. Institute of Oncology, University Medical Centre Ljubljana, Yugoslavia (1988)