Performance Assessment of Decision Tree-based Predictive Classifiers for Risk Pregnancy Care

Mário W. L. Moreira[1,2], Joel J. P. C. Rodrigues[1,3,4], Neeraj Kumar[5], Jianwei Niu[6], Isaac Woungang[7]
[1] Instituto de Telecomunicações, Universidade da Beira Interior, Covilhã, Portugal •
[2] Instituto Federal de Educação, Ciência e Tecnologia do 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] Computer Science and Engineering Department, Thapar University, Patiala (Punjab), India •
[6] School of Computer Science and Engineering, Beihang University, Beijing, China
[7] Department of Computer Science, Ryerson University, Ontario, Toronto, Canada

mario.moreira@ieee.org; joeljr@ieee.org; neeraj.kumar@thapar.edu; niujianwei@buaa.edu.cn; iwoungan@scs.ryerson.ca

***Abstract*— The e-Health core concept includes Web usage in an integrated way with tools and services for healthcare. This definition improves access, efficiency, and clinical care quality process that are necessary for a service delivery improvement. Decision support systems (DSSs) belong to a plethora of e-Health concept dimensions. For these systems construction, it is important to find a reliable intelligent mechanism capable to identify diseases that can worsen the patient's clinical condition. Thus, this paper proposes the use of tree-based data mining (DM) techniques for the hypertensive disorders prediction in the risk gestation. It presents the modeling, performance evaluation, and comparison between the tree based classifiers ID3 and NBTree. The 5-fold cross-validation method realizes the performance comparison. Results show that the NBTree classifier obtained better performance, presenting F-measure 0.609, ROC area 0.753, and Kappa statistic 0.4658. This classifier can be a key to a smart system development capable to predict risk events in pregnancy. Therefore, DSSs are a leading solution for the reduction of both mother and fetal mortality.**

Keywords*— Decision support systems; Decision trees; Classification algorithms; Pregnancy; Hypertension; Measurement; Standards

I. INTRODUCTION

The fifth recommendation presented by the United Nations Millennium Development Goals (MDGs) relates to improving maternal health. To achieve this objective, it was proposed the reduction of the maternal mortality rate by about three-quarters, in 25 years [1]. According to the UN report published in 2014, the maternal death rate is about 45% between 1990 and 2015, presenting progress signs regarding maternal health improvement during pregnancy and childbirth (in several countries). However, the maternal mortality ratio remains unacceptably high. Advances are far from the required 5.5% annual decline needed to achieve the goal. Bleeding, sepsis, abortion in risk conditions, childbirth obstruction, and hypertensive disorders in pregnancy are the cause of more than 80% of maternal deaths. Unfortunately, the most part of all the losses are preventable. Access to adequate health services, equipment, and supplies as well as qualified health staff could prevent these deaths [2]. A comprehensive review of the millennium goals realized by the World Health Organization (WHO) has shown the need for a universal health coverage. This coverage implies that all the people have access to quality health services without risking their financial situation. The proposal includes access to prevention, protection, treatment, and palliative care. For this, it is necessary the development of smart health systems [3]. From the e-Health viewpoint, a single registry should join all the people from a particular country and provide access to relevant information to them. This purpose determines the need to develop e-Health strategies and applications that support prevention and health promotion, such as platforms and services that allow patients handling their personal health. In this context, it turns health policies viable for remote areas and the use of mobile solutions, telemedicine, wearable technologies, sensors, electronic health records (EHRs), and diagnosis are fundamental. The use of information and communication technologies (ICTs) focused on e-Health involves data integration and its transformation for application in decision support [4]. Data mining (DM) is a good solution for processing large volumes of data from health information systems [5]. This method can recognize useful patterns, enabling support in complex analyses of clinical data. To identify possible high-risk cases that can lead pregnant woman and/or fetus to death, this reliable inference mechanism is a valuable tool. Then, using this approach for DSSs can support maternal mortality reduction in a reliable, accessible, and low-cost way. Then, this paper proposes the use of tree-based DM techniques for the hypertensive disorders prediction in the risk gestation. It presents the modeling, performance evaluation, and comparison between the tree based classifiers ID3 and NBTree.

The rest of the paper is organized as follows. Section II presents related works discussing the used methods to identify hypertensive disorders in high-risk pregnancies. Section III shows the modeling proposal that uses tree-based classifiers capable to identify hypertensive disorders based on symptoms presented by patients. Performance evaluation, methods

978-1-5090-5019-2/17/$31.00 ©2017 IEEE

comparison, and the results analysis considering the proposed plans are shown in Section IV. Finally, Section V provides the conclusion and suggestions for further works.

II. Hypertensive Disorders Identification in High-risk Pregnancy

This study uses data from 25 hypertensive pregnant women obtained from gynecologist/obstetrics specialists. Table I shows the used attributes and includes the corresponding brief explanation of each one. The input variables are the symptoms and risk factors presented by a pregnant woman while the output variables (classes) are the primary hypertensive diseases in pregnancy. Table II gives the classes that represent the main hypertensive disorders in pregnancy based on the study presented in [6].

Table I. Main Symptoms and Risk Factors: a Brief Explanation.

Main symptoms	***Brief explanation***
Headache	Migraine can be related to an increased risk of preeclampsia.
Epigastric pain	Epigastric pain is a late symptom, which is attributed to stretching the hepatic capsule due to edema and/or hemorrhage
Nausea or vomiting	It usually begins in the first month of pregnancy and goes through the 14th or 16th week with increased hormones production in gestation. Some pregnant develop severe vomiting crisis, known as hyperemesis, which causes fluid loss and minerals by the body.
Blurred vision	It is possible to identify some signs of preeclampsia by the eyes. In 8% of pregnant women can occur temporary vision loss, increased sensitivity to light, blurred vision, and halos or flashes.
Giddiness	Persistent dizziness is an anemia sign, migraine, high or low blood pressure, and certain other illnesses. It occurs together with blurred vision and headache.
Hyperreflexia	Presences of headache, visual and neurological symptoms (hyperreflexia) have a high correlation with an increase of cerebral perfusion pressure, which would be predicting an increased risk of cerebral alterations (eclampsia).
Edema	Preeclampsia usually is accompanied by hand and face edema. The edema presence is not considered a reliable signal for preeclampsia diagnosis because it can occur in pregnant women without changes in their pressure levels.
Oliguria	Decreased urine production (urinary volume < 500ml/24 h) is a criterion for a severe preeclampsia identification.
Hypertension	Tension levels classify the hypertension in mild (BP around 140/90mmHg) and severe (BP greater than 160/110mmHg). Such classification identifies the patient of greater or minor risk during pregnancy, besides guiding the treatment conduction.
Proteinuria	Proteinuria occurs in three up to four weeks before changes in fetal development and/or in the worsening of the maternal clinical condition. The 2g/24hr limit of proteinuria is a severity criterion. The intensity degrees can be related to the worse maternal and perinatal prognosis.
Risk factors	***Brief explanation***
Gestational age	Gestational age can predict the risk of developing severe preeclampsia forms, such as eclampsia and HELLP Syndrome.

It is possible to create a decision tree for the proposed models based on this information. From the symptoms presented by a pregnant woman and her gestational age, the suggested model can infer to identify the hypertensive disorder in pregnancy and its severity degree.

Table II. Main Pregnancy Hypertensive Disorders.

Pregnancy Hypertensive Disorder	***Brief explanation***
Pre-existing hypertension complicating pregnancy, childbirth, and puerperium (CH)	It is a hypertension that is present before pregnancy or diagnosed before 20 weeks of gestation. The systolic blood pressure (BP) > 140mmHg and/or the diastolic BP > 90mmHg, measured in two occasions with 4-hour interval defines hypertension. Such a diagnosis is more difficult to perform in hypertensive women without a previous diagnosis, due to the presence of physiological lowering of BP, which occurs in the first half of pregnancy. It is also considered CH that was diagnosed for the first time during pregnancy and does not normalize in the postpartum period.
Pre-existing hypertensive disorder with superimposed proteinuria (PS)	It is the appearance of proteinuria (> 0.3g/24h) after the gestational age of 20 weeks in patients with CH. Other indicators are based on proteinuria increase in those who already had a previous increase, a sudden BP increase in those with controlled levels, clinical or laboratory abnormality characteristic of PE.
Gestational hypertension (pregnancy-induced) without significant proteinuria (GH)	It is the BP increase that occurs after 20 weeks of gestational age without the proteinuria presence. It can represent a PE that has not had time to develop proteinuria. Another definition is transient hypertension, i.e., if the BP returns to normal values after 12 weeks of delivery, or a CH if the BP persists.
Preeclampsia or eclampsia (PE)	Preeclampsia is a syndrome characterized by different generalized clinical compromises and laboratory abnormalities. Clinical findings can be expressed as a maternal syndrome (hypertension, proteinuria, and/or various symptoms) or as a fetal syndrome, or both. It occurs in 5% to 8% of pregnancies and it is the leading cause of maternal and perinatal death in developing countries. Eclampsia is the onset of seizures in patients with PE or GH. It is associated with increased maternal mortality and it is often followed by cerebral hemorrhage. Despite the large research in this area, its cause remains unknown.

III. ID3 and NBTree Classifiers in Healthcare

Decision trees are models that use supervised training for data classification and prediction. The decision tree-based algorithms are among the most popular inference approaches. Several areas of knowledge have used these techniques, such as, in health, for medical diagnosis. Kishore *et al.* propose an ID3 algorithm variant in an anemic patients data set to reduce computational complexity and time required to construct a decision tree using large datasets [7]. Bashir *et al.* [8] compare several classification techniques using two data repositories benchmark to improve the diabetes diagnosis accuracy using decision trees. Results show that using the ID3 algorithm together with ensemble techniques can improve the decision tree performance regarding accuracy. Gomathi and Narayani [9] use the ID3 classification technique to predict systemic lupus erythematosus. Results show that the algorithm is capable to reduce the complexity and increase the computational performance.

The ID3 algorithm elaboration is based on inference systems and learning systems concepts. To construct the decision tree, the ID3 classifier separates a training set into subsets that contain a single class example. The algorithm divides this set through a single attribute selected from a statistical property, called information gain, which measures how informative an attribute is. The Algorithm I shows these steps.

ALGORITHM I. PSEUDO-CODE FOR ID3 ALGORITHM.

ID3 Algorithm

node LearnTree(examples, targetAttribute, attributes)
begin
  if all the examples have the same targetAttribute value,
    return a leaf with that value
  else if the set of attributes is empty
    return a leaf with the most common targetAttribute value among examples
  else begin
    A = the best attribute among attributes having a range of values $v_1, v_2, \ldots, v_k$
    Partition examples according to their value for A into sets $S_1, S_2, \ldots, S_k$
    Create a decision node N with attribute A
    for i = 1 to k
      begin
        Attach a branch B to node N with test V_i
        if S_i has elements (is non-empty)
          Attach B to LearnTree(S_i, targetAttribute, attributes – {A})
        else
          Attach B to a leaf node with most common targetAttribute
      end
    return decision node N
  end
end

The purity degree of a set defines its entropy. This concept represents the measure of lack information. Equation (1) shows the formula for the entropy calculation given a set S, with instances belonging to a class i, with probability p_i.

$$Entropy(S) = \sum p_i \log_2 p_i \qquad (1)$$

The information gain (IG) defines the entropy reduction. $IG(S, A)$ means the expected entropy decrease in S, ordering by the attribute A. Following, equation (2) calculates the IG.

$$IG(S, A) = Entropy(S) - \sum_{v \in values(A)} \frac{|S_v|}{|S|} . Entropy(S_v) \qquad (2)$$

The NBTree classifier is a hybrid classification model that combines decision trees with the Naïve-Bayes classifier. The nodes contain divisions considering a single attribute, as in regular decision trees, but the leaf nodes contain classifiers based on Bayesian networks. In the choice of a categorical attribute as the tree node, each possible attribute value receives an edge. In case the attribute is numeric, according to a threshold (the value that divides the examples), it is necessary to realize a binary division using the standard technique to minimize entropy, shown in (2). In health, it is possible to find this approach in research involving optimization to liver diseases classification [10], elderly well-being [11], and on essential proteins selection methods [12]. This method is present in the Algorithm II.

ALGORITHM II. PSEUDO-CODE FOR NBTREE CLASSIFIER.

NBTree Classifier [13]

for each attribute X_i, evaluate the information gain IG(Xi), of a split on attribute X_i
let J = AttMax(U_i). The attribute with the highest information gain
if U_j is not better than the information gain of the current node, create a Naïve Bayes classifier for the current node and return
Partition T according to the analysis on X_j.
  for all possible values do a multi-way split
For each child node, call the algorithm recursively on the portion of T that matches the test leading to the child.

Figures 1 and 2 show the decision trees created for both methods using the presented algorithms.

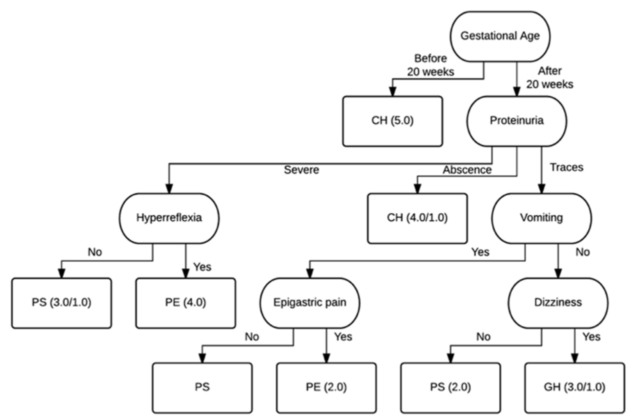

Fig. 1. Decision tree created according to the ID3 algorithm.

As there is a large set of possible hypotheses for the decision tree construction (Fig. 1), this study pruned the tree, removing the attributes that had information gain close to zero, *i.e.*, those that are not clearly relevant. First, the whole tree was created and, then, pruned.

IV. PERFORMANCE EVALUATION AND RESULTS ANALYSIS

The performance evaluation study uses the same database considered in Moreira *et al.* [14] adding the attribute "gestational age". This study also compares the J48 decision tree classifier. The database contains information from 25 simulated cases of pregnant women that suffered from hypertensive disorders in pregnancy. Obstetric/gynecologist physicians collaborated with this work constructing this dataset through their knowledge and experience to validate the proposed models.

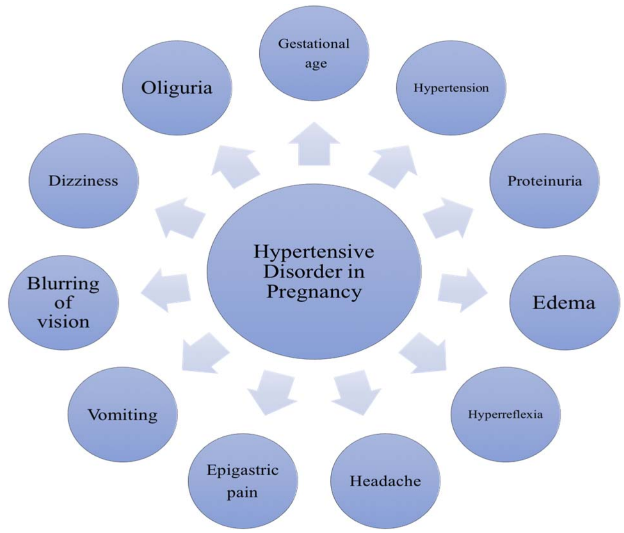

Fig. 2. Graph representing the relationship among NBTree classifier attributes.

A. *Standard metric measurements*

Performance evaluation of this study uses the 5-fold cross-validation method [15]. This technique divides the data set into 5 subsets (folds). The models learn, excluding the data from each subset, once until testing all folds. The separated fold, when applied to the model to test/evaluate, it estimates the error. Metrics for performance evaluation also consider the confusion matrix. Classifying the entire model cases into categories allow its construction, determining whether the predicted value corresponded to the real value. A confusion matrix is a conventional tool for statistical models' evaluation. The most significant misclassification types are the classification of a negative example as positive (false positive (FP)), also known as a false alarm (*e.g.*, identifying a high-risk pregnancy when the pregnant woman is healthy). The second type classifies a positive case as negative (false negative (FN)), *e.g.*, it diagnoses a high-risk pregnancy as healthy person. Based on this information, it is possible to calculate performance indicators, such as precision, recall, F-measure, and ROC curve. Precision indicates the percentage of positively cases that model predicts as negative. While the recall represents the rate that measures the number of negative cases, the model considers positive. The F-measure counts a harmonic average between recall and precision. It measures the system efficiency given the error in both classes. Equation (3) calculates this indicator.

$$F - measure = 2.\frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

B. *Experimental results*

Table III shows the performance evaluation of the proposed methods in comparison with the well-known J48 decision tree-based algorithm.

Table III. PRECISION, RECALL, F-MEASURE, AND ROC AREA VALUES OF THE ID3 AND NBTREE ALGORITHMS FOR EACH CLASSE.

Class	TP Rate	FP Rate	Precision	Recall	F-measure	ROC Area
ID3 Algorithm						
CH	0.6	0.105	0.6	0.6	0.6	0.697
PS	**0.333**	0.278	0.286	**0.333**	0.308	0.535
GH	0.333	**0.167**	0.4	0.333	0.364	0.588
PE	0.571	0.185	0.571	0.571	0.571	0.702
NBTree classifier						
CH	**0.667**	**0**	**1**	**0.667**	**0.8**	**0.904**
PS	**0.333**	**0.211**	**0.333**	**0.333**	**0.333**	**0.553**
GH	**0.667**	0.211	**0.5**	**0.667**	**0.571**	**0.719**
PE	**0.714**	**0.111**	**0.714**	**0.714**	**0.714**	**0.825**

The NBTree classifier presented the best performance for all the considered metrics and in all classes than the ID3 algorithm. Only the FP Rate indicator (where the best values are around zero) for PS class, the ID3 algorithm showed a better performance. Table IV shows the average standard metrics for all decision tree-based methods.

Table IV. AVERAGE STANDART METRICS VALUES OF ID3, NBTREE, AND J48 METHODS.

Standard Metrics	ID3	NBTree	J48
Prec.	0.463	**0.64**	0449
Rec.	0.458	**0.6**	0.44
F-Measure	0.459	**0.609**	0.441

On average, the NBTree classifier also performed better. This result shows that this method is an excellent predictor for all different types of hypertensive disorders during pregnancy. The ID3 algorithm presented performance slightly better than the J48 algorithm.

A Receiver Operating Characteristic (ROC) curve represents a robust graphical method, representing the variation of specificity (FP Rate) and sensibility (TP Rate). Table V shows the area value under ROC curve.

Table V. ROC AREA THAT CHARACTERIZES THE CLASSIFICATION MODELS PERFORMANCE.

Standard Metrics	ID3	NBTree	J48
TP Rate	0.417	**0.56**	0.52
FP Rate	0.198	**0.145**	0.169
ROC Area	0.603	**0.758**	0.738

Figure 3 shows the relation between the FP rate and TP rate by the ROC curve. The NBTree classifier presented a ROC curve closest to the point (0, 1) compared to the other methods. Curves close to this point indicate an excellent predictive performance.

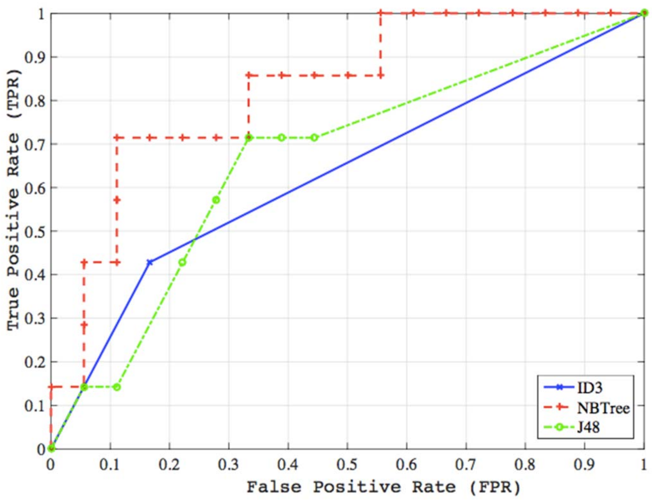

Fig. 3. ROC curve for the decision tree-based approaches.

Kappa statistic expresses an agreement measure used in nominal scales that gives an idea about how much differs the observations from those expected. Table IV shows the k statistic for the proposed methods.

Table VI. CALCULATION OF THE KAPPA STATISTIC FOR INTER-RATER RELIABILITY.

Kappa Statistic	ID3	NBTree	J48
k	0.2186	**0.4124**	0.3548

Results show that NBTree classifier is moderate regarding reliability and ID3 and J48 classifiers present acceptable results.

V. CONCLUSION

This paper presented tree-based classifiers known as ID3 and NBTree, comparing their behavior with the J48 decision tree classifier. NBTree classifier showed a great performance for the F-measure indicators (0.609), an excellent performance in the ROC Area (0.758), and provided average performance in the Kappa statistics (0.4124) compared to the other approaches presented.

Further works perspective indicates the study of new classification methods to improve the system reliability, such as artificial neural networks. Other search-based algorithms also need to be evaluated, such as decision trees and regression, and decision rules (OneR, Coverage, Top-Down, and Bottom-up algorithms, among others). Bayesian classifiers represent another important topic in classification problems. This approach has presented numerous applications in both academic and industry research.

e-Health concept is a relatively recent term regarding healthcare practice, but its solutions can reach unassisted women, mostly in developing countries where a large proportion of maternal deaths occurs. The contributions presented on this paper confirm its importance and usefulness.

ACKNOWLEDGMENTS

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 *Centro de Referência em Radiocomunicações* - CRR project of the *Instituto Nacional de Telecomunicações* (Inatel), Brazil, and by *Ciência sem Fronteiras* of CNPq, Brazil, through the process number 207706/2014 − 0.

REFERENCES

[1] D. A. Campbell, “An update on the United Nations Millennium Development Goals.” Journal of Obstetric, Gynecologic & Neonatal Nursing, vol. 46, no. 43, pp. e48-e55, May-Jun. 2017.

[2] L. Say, D. Chou, A. Gemmill, Ö. Tunçalp, A. B. Moller, J. Daniels, A. M. Gülmezoglu, M. Temmerman, and L. Alkema, “Global causes of maternal death: A WHO systematic analysis,” Lancet Global Health, vol. 2, no. 6, pp. 323–333, Jun. 2014.

[3] C. Dye, J. C. Reeder, and R. F. Terry, “Research for Universal Health Coverage,” WHO, vol. 5, no. 199, pp. 1–3, Aug. 2013.

[4] J. Andreu-Perez, C. C. Poon, R. D. Merrifield, S. T. Wong, and G. Z., “Big data for health,” IEEE journal of biomedical and health informatics, v. 19, no. 4, p. 1193-1208, Jul. 2015.

[5] W. Fan and A. Bifet, “Mining Big Data : Current Status, and Forecast to the Future,” ACM sIGKDD Exploration Newsletter, vol. 14, no. 2, Dec. 2012, pp. 1–5.

[6] L. A. Magee, A. Pels, M. Helewa, E. Rey, P. von Dadelszen, and SOGC, “Diagnosis, Evaluation, and Management of the Hypertensive Disorders of Pregnancy: Executive Summary,” Journal of Obstetrics and Gynaecology Canada, vol. 36, no. 7, pp. 575–576, May 2014.

[7] C. R. Kishore, K. P. Rao, and G. R. S. Murthy, “Performance Evaluation of Entorpy and Gini using Threaded and Non Threaded ID3 on Anaemia Dataset,” 2015 Fifth International Conference on Communication Systems and Network Technologies (CSNT 2015), Gwalior, MP, India, Apr. 4-6, 2015, pp. 1080–1084.

[8] S. Bashir, U. Qamar, F. H. Khan, and M. Y. Javed, “An Efficient Rule-based Classification of Diabetes Using ID3, C4.5 & CART Ensembles,” 12th International Conference on Frontiers of Information Technology (FIT 2014), Islamabad, Pakistan, Dec. 17-19, 2014, pp. 226–231.

[9] S. Gomathi and V. Narayani, “Systemic Lupus Erythematosus manifestation using ID3 Algorithm – A clinical Analysis,” 2014 International Conference on Data Mining and Intelligent Computing (ICDMIC 2014), Delhi, India, Sep. 5-6, 2014, pp. 1–5.

[10] N. Novita and T. Mantoro, “Data Mining Techniques For Optimatization of Liver Disease Clasification,” International Conference on Advanced Computer Science Applications and Technologies (ACSAT 2013), Sarawak, Malaysia, Dec. 22-24, 2013, pp. 379–384.

[11] A. Cufoglu and J. Chin, “Towards an understanding classification of well-being for care of older people,” IEEE 13th International Conference on Industrial Informatics (INDIN 2015), Cambridge, United Kingdom, Jul. 22-24, 2013, pp. 1494–1499.

[12] J. Zhong, J. Wang, W. Peng, Z. Zhang, and M. Li, “A Feature Selection Method for Prediction Essential Protein,” Tsinghua Science and Technology, vol. 20, no. 5, pp. 491–499, Oct. 2015.

[13] M. Hussein, “Analyzing NB, DT and NBTree Intrusion Detection Algorithms,” Journal of Zankoy Sulaimani-Part A, vol. 16, no. 1, pp. 69–76, Feb. 2014.

[14] M. W. L. Moreira, J. J. P. C. Rodrigues, A. M. B. Oliveira, K. Saleem, and A. Neto, “Performance Evaluation of Predictive Classifiers for Pregnancy Care,” IEEE Global Communications Conference (GLOBECOM 2016), Dec. 4-8, 2016, Washington, DC, USA, pp. 1–5.

[15] C. Bernau, M. Riester, A. Boulesteix, G. Parmigiani, C. Huttenhower, L. Waldron, and L. Trippa, “Cross-study validation for the assessment of prediction algorithms,” Bioinformatics, vol. 30, no. 12, pp. 105–112, Jun. 2014.