

Prediction of 30-Day Mortality after a Hip Fracture Surgery Using Neural and Bayesian Networks

Dimitrios Galiatsatos¹, George C. Anastassopoulos¹, Georgios Drosos², Athanasios Ververidis², Konstantinos Tilkeridis, and Konstantinos Kazakos²

¹ Medical Informatics Laboratory, Medical School,
Democritus University of Thrace, Greece

dgaliatsatos2010@hotmail.com, anasta@med.duth.gr

² Department of Orthopedics, University Hospital of Alexandroupolis, Medical School,
Democritus University of Thrace, GR-68100

drosos@otenet.gr, athanasios@ververidis.net,
tilkerorth@yahoo.com, kazakosk@yahoo.gr

Abstract. Osteoporotic hip fractures have a significant morbidity and excess mortality among the elderly and have imposed huge health and economic burdens on societies worldwide. A medical database of 349 patients that have been operated for hip fracture has been analyzed. Two models of data were used in Multi-Layer Perceptrons, Radial Basis Function and Naïve Bayes networks, in order to predict the 30-day mortality after a hip fracture surgery and also to investigate which is the most appropriate risk factor between the New Mobility Score and Institution factor for the Greek population. The proposed method may be used as a screening tool that will assist orthopedics in the surgery of the hip fracture according to each different patient.

Keywords: hip fracture, artificial neural networks, Bayesian networks, 30-day mortality.

1 Introduction

Hip fractures are common in elderly people although their incidence varies among different countries and populations [1-3]. It has been estimated that according to the epidemiologic projections, the worldwide annual number of hip fractures will rise from 1.66 million in 1990 to 6.26 million by the year 2050 [4]. The mortality, morbidity social-economic costs after a hip fracture are significant [5-7]. Mortality in particular remains significant for decades although there is a geographical and Race/Ethnicity Differences [8-10]. Mortality rates are greatest within the first month [11-14]. The 30-day mortality has been used in several studies because the mortality during this time is directly related to the fracture, its therapy and complications, while the mortality rate beyond this time it may be related to other unrelated causes.

The ability to recognize patients at high risk of poor outcomes before operation would be an important clinical advance. Several preoperative risk factors for the high

mortality have been identified but the most predominant did not, and whether other factors could be considered as predictors are still unknown [15-17].

There are several reports concerning the pre-operative risk factors or scores like the Abbreviated Mental Test Score (AMTS) or the New Mobility Score that have been used as predictors for the 30-day mortality in patients with hip fractures. One of these scores is the Nottingham Hip Fracture Score (NHFS) [13, 18-19]. This score is using the following factors: age, sex, hemoglobin in admission, Abbreviated Mental Score (MMS), and living or not in an institution, number of co-morbidities (other illness) and suffering from malignancy.

Patients that live in an institute are those that cannot live alone and the factor “living in an institute” is related to the “mobility status” of the patients. In some countries –like our country- elderly people usually live with relatives (in their own homes or in their relatives’ homes). Therefore, the factor “living in an institute” is not a reliable factor. The New Mobility Score is a score related with the “mobility status” of these patients and has been found dependent with morbidity [20]. Patients with New Mobility Score 6 and above can walk out of the house and go for shopping even if they use a stick, while patients with score below 6 need help from another person. This score was used in this study in order to identify whether replacing the factor “living in an institute” with this score the results are more reliable or not.

The Artificial Neural Network (ANN) simulating high-level human brain functions, is a computational modeling tool that has become widely accepted for modeling complex real-world problems. Although it has been explored in many areas of medicine, such as nephrology, microbiology, radiology, neurology, cardiology, etc, its use in the orthopedic field is still rare [21].

The aim of this work was to apply three different ANN - based methods (Multi-Layer Perceptrons (MLPs), Radial Basis Function (RBF) Networks and Bayesian Networks) to evaluate their prediction of mortality using the factors of NHFS. In more details, the 3 methods were tested in a given number of patients with known characteristics (factors) and a known 30-day mortality. More specifically a decision support tool has been developed to help clinicians identify which people are at increased risk after the hip fracture surgery. This application area is considered as extremely important since it is associated with increased morbidity and mortality and high socio-economic costs. This paper was tried to focus on the three proposed methods and their performance in clinical data.

The paper is structured as follows: In the next section the hip fracture, as well as the risk factors that affect the 30-day mortality after the operation of the hip fracture are presented. Also, two models of data are described and three methods are illustrated. Section 3 analyzes the results, while in section 4 the conclusions are discussed.

2 Materials and Methods

2.1 Hip Fracture

This study assessed the predictive capability of Neural and Bayesian networks for 30-day mortality after surgery for hip fracture in the Greek population.

Data were collected from 349 patients that have been operated for hip fracture in the Orthopedic department of the University Hospital of Alexandroupolis. All these patients were followed-up and the 30-day mortality was recorded. The number of patients who died after 30-day of monitoring was 33. Apart from patients demographics, other factors that may be related to patients' mortality were also recorded and categorized as follows; Age (between 65 and 85 and above 85 years), Sex (male or female), Hemoglobin (Hb) in admission (below 10 g/dl and above 10 g/dl), Mini Metal Score (MMTS) (below 6 or above 6), Living or not in an Institute, Number of co-morbidities (other illness) (less than 2 or more than 2), suffering from Malignancy or not, and the New Mobility Score (score 6 and below or above 6).

Two models of data set have been used as follows: The first model (Model 1) depends on the factors: Age, Sex, Hb, MMTS, New Mobility Score, Number of co-morbidities and Malignancy. In the second model the New Mobility Score had been replaced by the Institute factor (Model 2). This model is using the factors that are being used in the Nottingham Hip Fracture Score. Both models consisted of seven almost identical attributes (independent variables), while the death was the dependent variable.

2.2 Neural and Bayesian Networks

Trying to find the best model for 30-day mortality after surgery for hip fracture, two most common algorithms of Artificial Neural Networks have been used, the Multilayer Perceptron and the Radial Basis Function. In addition to these ANN algorithms, the simplest Bayesian algorithm (Naïve Bayes) has been utilized. As a validation measurement of the data the k-fold cross validation has been utilized in all of the above methods.

All data were analyzed and processed with the assistance of Weka, which is a machine learning software written in Java and is an open source application that is freely available. Weka contains tools for data processing, classification, clustering, feature selection and visualization. The greatest capability of this machine learning is to perform useful information after learning from training data. [22].

Multilayer Perceptron is a feed-forward algorithm with one or more hidden layers between the input and the output layer. It utilizes a supervised learning technique called back propagation for training the network. In the medical field, MLP algorithm is widely used and gives great classification accuracy for some diseases such as breast cancer [23-24].

Radial Basis Function networks are feed-forward networks trained using a supervised training algorithm. They are similar to back propagation networks in many aspects although RBF have a few advantages. The main advantage in the biomedicine applications is that they train much faster and more accurate than back propagation networks [25-26]. In contrast with the time, the models are less robust than other methods (for example logistic regression) [21, 27].

Bayesian Networks are very efficient classifiers in many health related datasets (Natural Language Processing, Computer Vision, Medical Diagnosis, Bioinformatics, etc) [28]. Naïve Bayes (NB) Classifier is the simplest Bayesian approach algorithm assuming conditional independence between the variables in the models [29]. This simplicity makes naïve bayes techniques attractive and suitable in many domains. The

connection in this classifier appears only between the variables and the main node in each model (output/dependent variable). The main advantages are its easy construction (known a priori) and its effective classification process. However there is a strong disadvantage that makes it unsuitable for every set of data. The conditional independence assumptions between nodes rarely are true in the most real world applications. Despite although strong dependencies between nodes, researchers had shown that naïve bayes performs good classification [30]. An updated form of the naïve bayes algorithm such as the augmented naïve bayes algorithm overcomes the conditional independence and it is an innovative algorithm based on the original one. In recent years many researchers had tried to evolutes naive bayes classifiers as they select feature subset and relax independence assumptions [31-32].

In k-fold cross validation the data set separated in k equal size subsets. From the k subsets, a single subset preserved as a validation data and the remaining k-1 samples are utilized as training data. The cross validation process repeated k times with each of the k subsets used exactly once. Then the k results are averaged and came of the estimation. The advantage of this method is that each observation is used in training and validation process. In this work, 3, 10 and 20 fold cross validation has been performed in each technique. In experiments, such as in Hyperthyroid disease, a 6-fold cross validation performed maximum accuracy [33].

3 Results

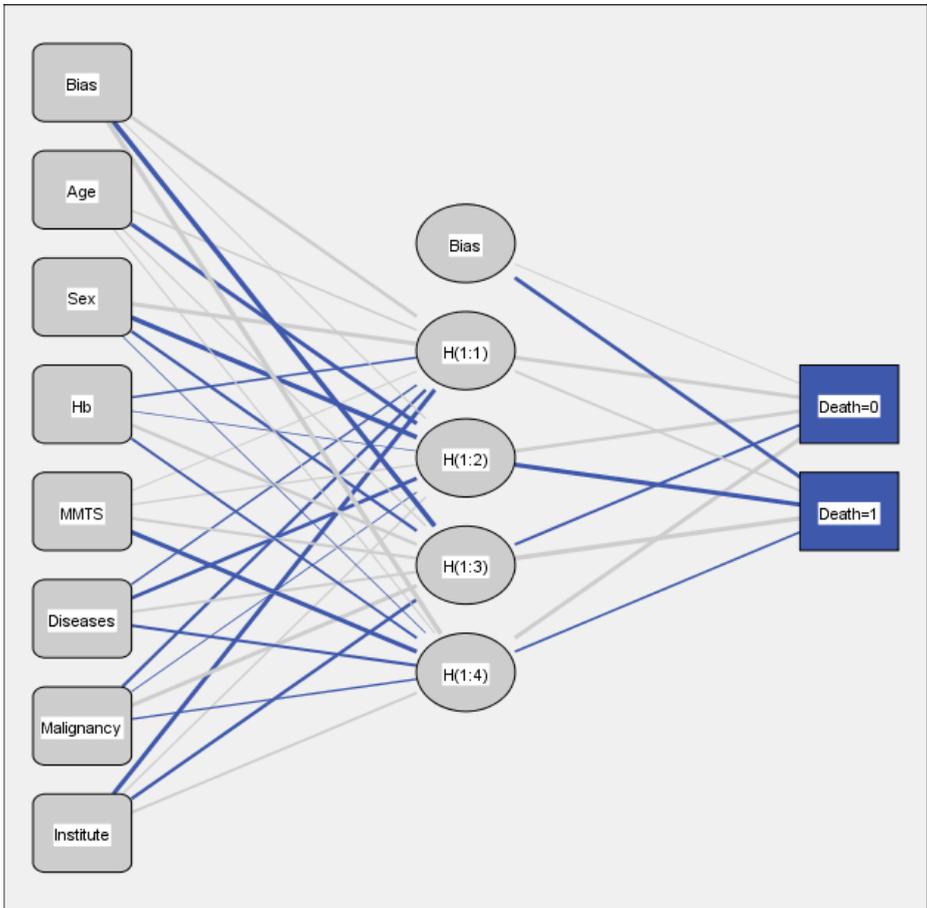
For each one of the two models, the entire set of 349 data records were used, in order to evaluate the best classification method between multilayer perceptron, radial basis function and naïve bayes networks. For the validation of the data, 3, 10 and 20-fold cross validation for each classified algorithms has been used. For each neural network (MLP and RBF) many experiments have been made using a lot of different topologies in terms of hidden layer etc. The results presented below correspond to the best topology of each method which had the best performance.

In figure 1a, MLP topology of the second model is presented. Seven factors are presented in the input layer, whereas four nodes consist the hidden layer. The same topology had been used in the RBF network.

For the evaluation of the models except from Artificial Neural Networks (MLP and RBF), Bayesian neural networks (naïve Bayes) had utilized in order to find the best classified model.

In figure 1b the direct dependencies between the dependent variable (Death) and the independent variables of the first model in the naïve bayes networks are presented. As it can be observed this specified algorithm has given no direct dependencies between the independent variables. For the better explanation of the Bayesian network, probabilities were concluded in each confusion matrix between variables given the output. The same topology with figure 1a has the naïve bayes form of the second model. This is due to the assumption of non-existence of direct dependencies between the independent variables. Even though the similarity of the graphs, there are strong differences between probabilities in the factors. The most important in the first model which is illustrated below are the probabilities of New Mobility Score which are above or equal to six given that the patient is dead ($Pr=0.809$) and the number of

diseases which are over or equal to three given death ($Pr=0.962$). In the second model the highest scoring probability corresponds to Number of co-morbidities factor given death ($Pr=0.926$). The lowest scoring in both models is the probability between institute given death ($Pr=0.126$). The probability of having a New Mobility Score below 6 given Death has a huge difference than the probability living in an Institution given death. The real differentiation between the two pre-mentioned results reflected into the patients with hip fracture in the Greek society. A probability result of 0.809 in New Mobility score translated that a patient with score below 6 has over 80% possibility to die in contrast with the patient who was living in an Institute who has approximately 12%. Due to this reason, the institution factor in the Greek population is unsuitable for study although the classification of the Bayesian network are higher in the model which included.



(a)

Fig. 1a,b. Neural and Naïve Bayes Networks

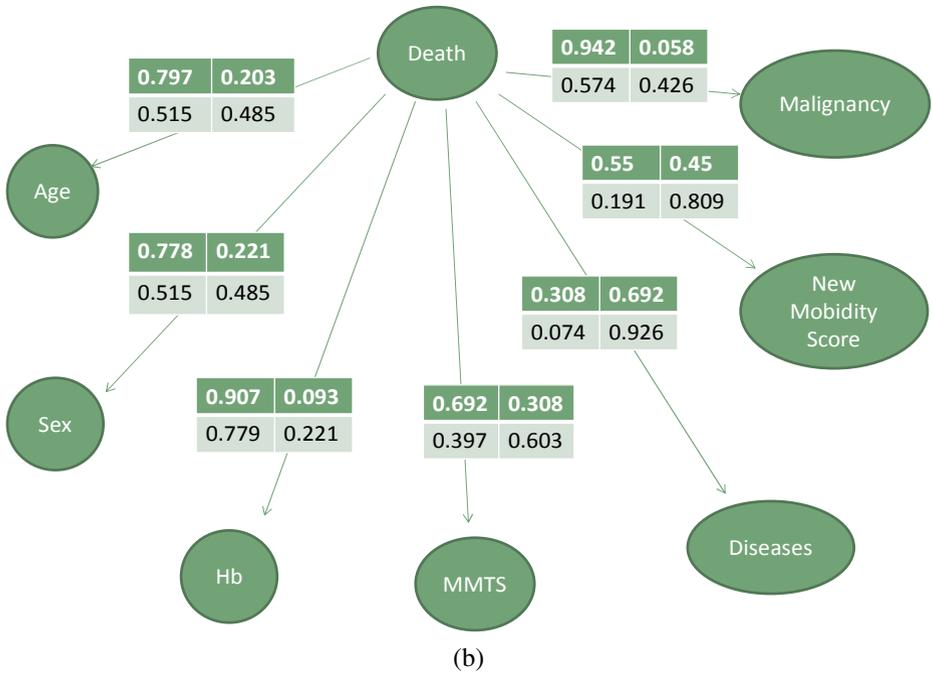


Fig. 1. (Continued.)

Table 1 presents the comparisons between multilayer perceptron, radial basis function and naïve bayes networks between model 1 and model 2 for different cross validation folds. Results had shown that in hip fracture models, the 10-fold cross validation was the most appropriate for the most of the cases. In the first and second columns of the table 1, the experiments of the neural and Bayesian networks are presented for each model, respectively. For the accuracy of the networks a correctly classified and an incorrectly classified index was used. The first index sums the true positive and the false negatives values of the confusion matrix and shows the proportion of the total number of predictors that were correctly identified, whereas the second index sums false positives and true negatives values and indicates the predictors that were incorrectly recognized. The time is the processing time value that is required to build the model, measured in seconds.

The processing time of the MLP algorithm is much higher than that of RBF and naïve bayes network algorithms. The same results are illustrated graphically in figure 2 for the 10-fold cross validation. Although both the neural networks classifiers had resulted that the first model are the most appropriate to predict the 30-day mortality after surgery for hip fracture, the results from naïve Bayes classifier had given the impression that the second model is the most suitable.

Table 1. Evaluation results of MLP, RBF and NB for the two models

3-fold cross validation	Model	Correctly Classified Instances	Incorrectly Classified Instances	Time (sec)
MLP	1	320 (91.6905%)	29 (8.3095%)	1.28
	2	322 (92.2636%)	27 (7.7364%)	0.95
RBF	1	320 (91.6905%)	29 (8.3095%)	0.59
	2	324 (92.8367%)	25(7.1633%)	0.05
NB	1	317(90.8309%)	32(9.1691%)	< 0.01
	2	322(92.2636%)	27(7.7364%)	< 0.01
10-fold cross validation	Model	Correctly Classified Instances	Incorrectly Classified Instances	Time (sec)
MLP	1	322 (92.26%)	27 (7.74%)	1.09
	2	321 (91.97%)	28 (8.03%)	0.92
RBF	1	324 (92.83%)	25 (7.17%)	0.48
	2	322 (92.26%)	27 (7.74%)	0.03
NB	1	316 (90.54%)	33 (9.46%)	0.01
	2	325 (93.12%)	24 (6.88%)	< 0.01
20-fold cross validation	Model	Correctly Classified Instances	Incorrectly Classified Instances	Time (sec)
MLP	1	320 (91.6905%)	29 (8.3095%)	0.87
	2	327 (93.6963%)	22 (6.3037%)	1.13
RBF	1	319 (91.404%)	30 (8.596%)	0.03
	2	323 (92.5501%)	26 (7.4499%)	0.03
NB	1	314 (89.9713%)	35 (10.0287%)	< 0.01
	2	325 (93.1232%)	24 (6.8768%)	< 0.01

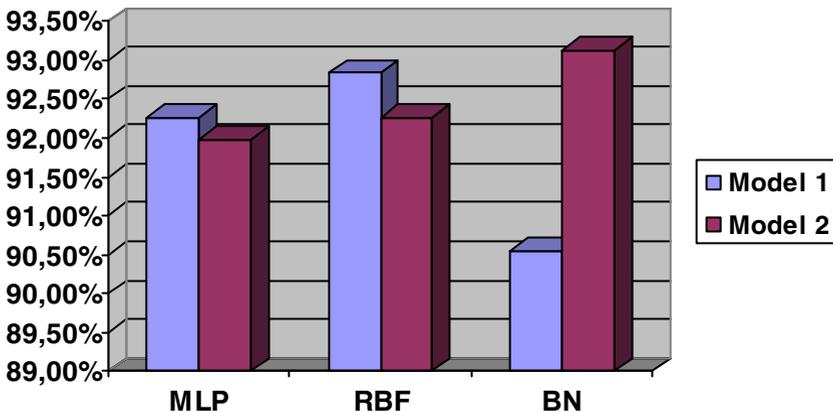


Fig. 2. Percentage of Correctly Classified Instances

From table 1 and figure 2 it is denoted that the instances in the first model were classified in a better way than in the second model for MLP (92.26% in the first, whereas 91.97% in the second model) and RBF (92.83% in the first, whereas 92.26% in the second model) algorithms. The naïve Bayes classifier evaluated significantly good classification for the second model (93.12%) in hip fracture data (all the classification results are above 90%). Also, the processing time for the classification is quite smaller in NB networks. The first model which included the New Mobility Score matched better in the citizens of Greece than the model included institution factor.

4 Conclusion

The Artificial Neural Networks and the Bayesian Networks are subfields of the computational intelligence. Although they are widely used in most research areas, their use in the field of medicine is still limited, especially the use of Bayesian Networks is almost nonexistent in the orthopedic field.

In this paper, the application of MLP, RBF and NB networks that have been applied to medical data, in order to predict the 30-day mortality after surgery for hip fracture in the Greek population is presented. In addition, two models of the medical data have been used, in order to investigate who is the most appropriate risk factor for the Greek population, between the New Mobility Score and the Institution factor, as it is described in the NHFS.

All the methods performed prediction of 30-day mortality over 90%. That means that all these methods are suitable for designing and developing a decision support tool to help clinicians identify which people are at increased risk after the hip fracture surgery. Moreover, the neural networks had resulted that the first model are the most appropriate to predict the 30-day mortality after surgery for hip fracture, but the results from naïve Bayes network had given the impression that the second model is the most suitable.

In the future, more computational intelligence methods have to be investigated (especially Bayesian Networks) in more clinical records from the Greek population, in order to find the best method that could predict the 30-day mortality after hip fracture surgery with maximum reliability. In addition, the institution factor in accordance with the New Mobility Score must be further investigated in order to decide which one of these factors is the most suitable for the Greek population.

References

1. Chang, K.P., Center, J.R., Nguyen, T.V., Eisman, J.: Incidence of hip and other osteoporotic fractures in elderly men and women: Dubbo Osteoporosis Epidemiology Study. *J. Bone Miner. Res.* 19, 532–536 (2004)
2. Court-Brown, C.M., Caesar, B.: Epidemiology of adult fractures: a review. *Injury* 37, 691–697 (2006)
3. Dhanwal, D.K., Dennison, E.M., Harvey, N.C., Cooper, C.: Epidemiology of hip fracture: Worldwide geographic variation. *Indian J. Orthop.* 45, 15–22 (2011)

4. Kannus, P., Parkkari, J., Sievänen, H., Heinonen, A., Vuori, I., Järvinen, M.: Epidemiology of hip fractures. *Bone* 18(suppl. 1), 57S–63S (1996)
5. Keene, G.S., Parker, M.J., Pryor, G.: Mortality and morbidity after hip fractures. *BMJ* 307(6914), 1248–1250 (1993)
6. Braithwaite, R.S., Col, N.F., Wong, J.: Estimating hip fracture morbidity, mortality and costs. *J. Am. Geriatr. Soc.* 51, 364–370 (2003)
7. Abrahamsen, B., van Staa, T., Ariely, R., et al.: Excess mortality following hip fracture: a systematic epidemiological review. *Osteoporos Int.* 20, 1633–1650 (2009)
8. Giverson, I.: Time trends of mortality after first hip fractures. *Osteoporos Int.* 18, 721–732 (2007)
9. Haleem, S., Lutchman, L., Mayahi, R., Grice, J.E., Parker, M.: Mortality following hip fracture: trends and geographical variations over the last 40 years. *Injury* 39, 1157–1163 (2008)
10. Sterling, R.: Gender and race/ethnicity differences in hip fracture incidence, morbidity, mortality, and function. *Clin. Orthop. Relat. Res.* 469(7), 1913–1918 (2011)
11. Bass, E., French, D.D., Bradham, D.D., Rubenstein, L.: Risk-adjusted mortality rates of elderly veterans with hip fractures. *Ann. Epidemiol.* 17, 514–519 (2007)
12. Giverson, I.: Time trends of mortality after first hip fractures. *Osteoporos Int.* 18, 721–732 (2007)
13. Maxwell, M.J., Moran, C.G., Moppett, I.: Development and validation of a preoperative scoring system to predict 30-day mortality in patients undergoing hip fracture surgery. *Br. J. Anaesthesia* 101, 511–517 (2008)
14. Hu, F., Jiang, C., Shen, J., Tang, P., Wang, Y.: Preoperative predictors for mortality following hip fracture surgery: a systematic review and meta-analysis. *Injury* 43(6), 676–685 (2012)
15. Hannan, E.L., Magaziner, J., Wang, J.J., et al.: Mortality and locomotion 6 months after hospitalization for hip fracture: risk factors and risk-adjusted hospital outcomes. *JAMA* 285, 2736–2742 (2001)
16. Johansen, A., Mansor, M., Beck, S., et al.: Outcome following hip fracture: post-discharge residence and long-term mortality. *Age Ageing* 39, 653–656 (2010)
17. Ozturk, A., Ozkan, Y., Akgoz, S., et al.: The risk factors for mortality in elderly patients with hip fractures: postoperative one-year results. *Singapore Med. J.* 51, 137–143 (2010)
18. Moppett, I.K., Parker, M., Griffiths, R., Bowers, T., White, S.M., Moran, C.: Nottingham Hip Fracture Score: longitudinal and multi-assessment. *Br. J. Anaesth.* 109(4), 546–550 (2012)
19. Gunasekera, N., Boulton, C., Morris, C., Moran, C.: Hip fracture audit: the Nottingham experience. *Osteoporos Int.* 21(suppl. 4), S647–S653 (2010)
20. Parker, M.J., Palmer, C.: A new mobility score for predicting mortality after hip fracture. *J. Bone Joint Surg. Br.* 75(5), 797–798 (1993)
21. Tseng, W.-J., Hung, L.-W., Shieh, J.-S., Abbod, M.F., Lin, J.: Hip fracture risk assessment: artificial neural network outperforms conditional logistic regression in an age- and sex-matched case control study. *BMC Musculoskeletal Disorders* 14, 207 (2013), doi:10.1186/1471-2474-14-207
22. <http://www.cs.waikato.ac.nz/ml/weka/> (last accessed April 21, 2014)
23. Breast Cancer Diagnosis on Three Different Datasets Using Multi-Classifiers. *International Journal of Computer and Information Technology*
24. Othman, M.F.B., Yau, T.M.S.: Comparison of Different Classification Techniques. Using WEKA for Breast Cancer

25. Venkatesan, P., Anitha, S.: Application of a radial basis function neural network for diagnosis of diabetes mellitus. Tuberculosis Research Centre, ICMR, Chennai 600 031, India
26. Del_can, Y., Özyilmaz, L., Yildirim, T.: Evolutionary Algorithms Based RBF Neural Networks For Parkinson's Disease Diagnosis. In: ELECO 2011 7th International Conference on Electrical and Electronics Engineering, Bursa, Turkey, December 1-4 (2011)
27. Escaño, L.M.E., Saiz, G.S., Lorente, F.J.L., Fernando, Á.B., Ugarriza, J.M.V.: Logistic Regression Versus Neural Networks For Medical Data. *Monografías del Seminario Matemático García de Galdeano* 33, 245–252 (2006)
28. Comparison of Different Classification Methods Reihaneh Rabbany Khorasgani, Department of Computing Science, University of Alberta, Edmonton, Canada
29. Tomar, D., Agarwal, S.: A survey on data mining approaches for healthcare. *International Journal of BioScience and Biotechnology* 5(5), 241–266 (2013)
30. Zhang, H.: The Optimality of Naïve Bayes, Faculty of Computer Science, University of New Brunswick, Fredericton, New Brunswick, Canada
31. Langseth, H., Nielsen, T.: "Classification using Hierarchical Naïve Bayes models". *Machine Learning* 63(2), 135–159 (2006)
32. Zhang, N.: Hierarchical latent class models for cluster analysis. In: 18th National Conference on Artificial Intelligence, pp. 230–237 (2002)
33. Diagnosis And Classification Of Hypothyroid Disease Using Data Mining, Techniques. *International Journal of Engineering Research & Technology (IJERT)*