A Novel Weighted Hierarchical Adaptive Voting Ensemble Machine Learning Method for Breast Cancer Detection

Clemen Deng
Department of Electrical and Computer Engineering
Portland State University
Portland, OR, U.S.A.
clemen.deng@gmail.com

Marek Perkowski
Department of Electrical and Computer Engineering
Portland State University
Portland, OR, U.S.A.
mperkows@ee.pdx.edu

***Abstract*—A novel Weighted Hierarchical Adaptive Voting Ensemble (WHAVE) machine learning (ML) method was developed for breast cancer detection. It was constructed using three individual ML methods based on Multiple-Valued Logic: Disjunctive Normal Form (DNF) rule based method, Decision Trees, Naïve Bayes, and one method based on continuous representation: Support Vector Machines (SVM). Results were compared with other methods and show that the WHAVE method accuracy was noticeably higher than the individual ML methods tested. This paper demonstrates that the WHAVE method proposed outperforms all methods researched, and shows the advantage of using WHAVE method for ML in breast cancer detection.**

Keywords— Machine Learning; Ensemble; Majority Voting System; Multi-Valued Logic.

I. INTRODUCTION

Breast cancer is a type of cancer originating from breast tissue, and the first signs of it are breast lumps or an abnormal mammogram. If one develops a breast lump or has an abnormal finding on a mammogram, a biopsy maybe needed for breast cancer diagnosis. There are two main types of biopsies: needle biopsy and surgical biopsy. Needle biopsy types include fine needle biopsy (FNA), core needle biopsy, vacuum assisted breast biopsy. Through FNA, data on cytological characteristics can be obtained and assessed for breast cancer diagnoses.

There are many existing machine learning methods used for breast cancer detection, some with insufficient accuracy rates. One idea to increase the accuracy rate is to combine multiple machine learning methods into one. This is done through a majority voting system using ensembles. This method takes into account the outputs of the individual machine learning methods and produces a classification based on them.

Results of different machine learning algorithms for breast cancer detection can be found in existing papers [1-24]. Prediction accuracy ranges from 65% [1] to 99.54% [4].

Clark and Niblett tested the DNF method in 1989, giving a 65-72% accuracy [1]. Decision tree was tested by Quinlan in 1996, giving a 94.74% accuracy [2]. Naïve Bayes was tested by Bellaachia and Guven in 2006, resulting in an 84.5% accuracy rate [3]. Übelyi tested SVM in 2007, giving 99.54% accuracy [4]. This was the highest accuracy out of all the previous research methods that we found in literature. An ensemble method was tested in 2011 combining the methods Neural Fuzzy, K-Nearest Neighbors, and Quadratic Classifier, resulting in 97.14% accuracy [5]. This ensemble method did not use weights or have a hierarchical system of testing.

In this paper, we develop a Weighted Hierarchical Adaptive Voting Ensemble (WHAVE) machine learning method with a novel weights formula applied to the majority voting system. The method is unique in three aspects. First, the method is hierarchical since it employs a searching algorithm to always combine the most accurate individual Machine Learning (ML) method to an ensemble with other ML methods in each step. Second, the method applies a new weighting formula to the majority voting ensemble system and the formula can be adaptively adjusted to search for the optimal one that yields the highest accuracy. Third, the method is adaptive as it uses stopping criteria to allow the algorithm to adaptively search the optimal weights and hierarchy for the ensemble methods. It was also our intention to compare and combine methods based on two different representations of data: multiple-valued and continuous, with the belief that combining different types of methods should give better results.

The rest of the paper is organized as follows: Section 2 describes the breast cancer database used for testing the new method in this paper and details the methodology for the new method, Weighted Hierarchical Adaptive Voting Ensemble (WHAVE). Section 3 shows the experimental results of the new method and comparative results with individual ML methods as well as conventional Majority Voting System (MVS) method results. Section 4 concludes the paper.

II. MATERIALS AND METHODS

A. Breast Cancer Database

The Wisconsin Breast Cancer Database (WBCD) [25] was used for algorithm testing in this paper. The WBCD was produced by Dr. Wolberg based on a Fine Needle Aspiration (FNA). Nine cytological characteristics were extracted for each of the 699 patients tested, creating the data set known as the Wisconsin Breast Cancer Database.

0195-623X/15 $31.00 © 2015 IEEE
DOI 10.1109/ISMVL.2015.27

The features are displayed below:

#	Attribute	Domain
1.	Clump Thickness	1-10
2.	Uniformity of Cell size	1-10
3.	Uniformity of Cell shape	1-10
4.	Marginal Adhesion	1-10
5.	Single Epithelial Cell Size	1-10
6.	Bare Nuclei	1-10
7.	Bland Chromatin	1-10
8.	Normal Nucleoli	1-10
9.	Mitoses	1-10
10.	Class:	(2 for benign, 4 for malignant)

Number of instances:	699
Missing attributes:	16
Benign:	458
Malignant:	241

Each of these cytological characteristics of breast FNA in the database are multi-valued, graded from 1 to 10 at the time of sample collection, with 1 being closest to benign and 10 the most anaplastic [26].

B. Novel Weights Formula for MVS

The idea of majority voting system (MVS) is to use different machine learning algorithms to classify data, and choose the result that most of the algorithms predict [27]. This avoids any misclassifications done by any one method, hence improves the accuracy.

The case of equal voting outcome can also be avoided by using weighted majority voting. If one machine learning method performs better than others, the significance of the vote of that machine learning method increases. The resulting classification equation of weighted majority voting is in the form of:

*If (W_1 * M1 + W_2 * M2 + W_3 * M3 > Threshold)*
Then Classification = positive
Where:
M1, M2, M3 stands for individual Machine Learning method's classification results;
W_1, W_2, W_3 stands for the weights applied to individual ML method's classification results.
Threshold is the classification threshold value.

The conventional weights formula is as follows:

$$W_i = \frac{A_i}{\sum_{i=1}^{n} A_i} \qquad \text{Equation (1)}$$

Where,
A_i is the individual ML method accuracy;
W_i is the weights applied to the individual ML method.

This weighted majority voting scheme is proven to be more effective than the un-weighted majority voting.

In our system, in addition to using un-weighted majority voting and conventional weighted majority voting to find the optimal weights, we propose a novel weights method as shown below.

$$W_i = \frac{(1-A_i)^{-x}}{\sum_{i=1}^{n}(1-A_i)^{-x}} \qquad \text{Equation (2)}$$

Where
A_i is the individual ML method accuracy;
W_i is the weights applied to the individual ML method.
x is a parameter that can be adjusted and varied adaptively based on the accumulation of the database and ensemble methods to find the optimal weights.

When x = 1, the value of $(1-A_i)^{-1}$ increases as A_i increases. When x = 0, the weights for all classification methods are the same. When x equals a very large number, the individual ML classification method that has the highest accuracy will have the highest weight. It is important to note that the minimum and maximum values of x (x_{min} and x_{max}) are selected based on experimentation and the accuracy of each individual method. Fig. 1 below shows the relationship between relative weights vs. accuracy level at different value of x. Here, relative weights are calculated as the ratio of the weights of a given individual method to the weights of the highest accuracy individual method. The weights are calculated based on Equation (2) above. Weights are normalized such that the sum of the weights of all methods equals 1.

The adaptive nature of the WHAVE algorithm developed in this work allows the algorithm to search for the optimal x for the weight formula that yields the highest accuracy. The algorithm first selects x = 0, then x = x + Δx, where Δx is the step size to increase x to test the weights and accuracy. This is repeated until the accuracy stops improving.

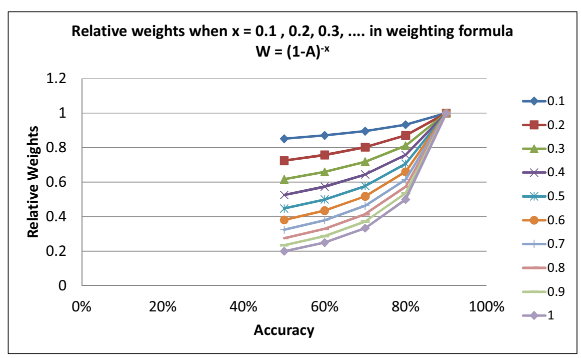

Fig. 1. Relative weights vs. accuracy when x = 0.1 to 1, accuracy range from 50% to 90%

Fig. 1 shows that:

(1) The higher the accuracy of the ML method, the larger the relative weights it carries.
(2) When x = 0, relative weights of all individual ML method are the same. So each method has equal weights.
(3) When 0 < x < 1, the relative weights of the less accurate individual method increases as x decreases.

C. *Weighted Hierarchical Adaptive Voting Ensemble (WHAVE) Method*

The WHAVE method developed in this paper selects all possible groups of ***n*** machine learning algorithms to ensemble (in the example below, we use $n = 3$). Each group requires including the highest accuracy algorithm. Weighted majority weighted voting is applied to each of these groups, and the three methods are trained and tested. The group with the highest accuracy is deemed to be the optimal method.

The steps are as follows using n=3 as an illustration example:

1. Select the method with the highest accuracy
2. Create ensembles by selecting all permutations of two other methods and putting them in ensembles with the method selected in step 1.
3. Using the 3-method weighted ensembles, train and test on the data (each ensemble applies majority voting and weights for each of its 3 methods). Set x = 0 in the weighted ensemble formula.
4. Select the ensemble with the highest accuracy.
5. Compare the highest accuracy from Step 4 with the highest individual method accuracy. If it is greater than the highest individual method accuracy, then do the next level of ensemble by selecting the ensemble combination that yields the highest accuracy and ensemble that with each of the remaining individual method. Repeat Step 4 and 5 until the accuracy of ensemble stops improving.
6. Vary x = x + 0.25, repeat Step 3 – Step 5, until the accuracy of the ensemble method stops improving.

Fig. 2 illustrates how the hierarchical ensemble method works using six individual ML classification methods as an example.

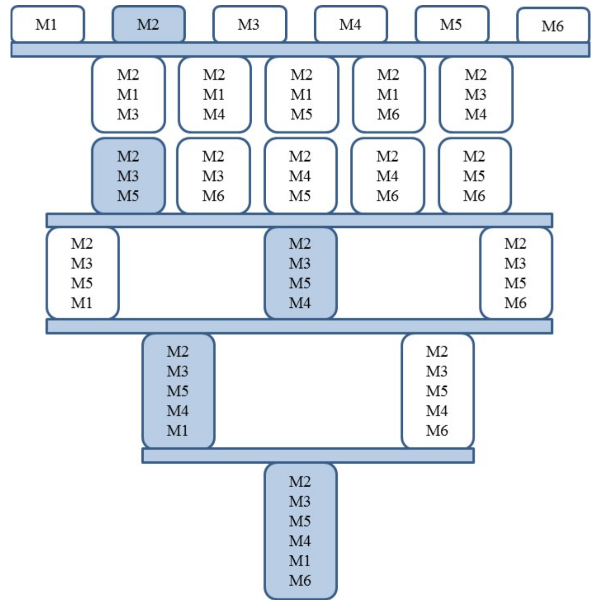

Fig. 2. Hierarchical ensemble method illustration

In Fig. 2, M1, M2, M3, M4, M5, M6 stand for the six ML methods. The blue colored methods are the ones with the highest accuracy rates in that level. Each level is separated by a light blue line. It employs a searching algorithm to always combine the most accurate ML method to ensemble with the remaining other ML methods in each level. If the accuracy stops improving after a certain level, the method stops there and does not go to the next level. The reason an exhaustive search is not done is because as more machine learning methods are added to the ensemble, the computation time for an exhaustive search increases exponentially. WHAVE allows the method to function effectively on a large number of machine learning methods, particularly when dealing with a large database without risking high computation time.

Let N_1 be the total number of ensemble for the WHAVE method and N_2 be the total number ensemble for the exhaustive ensemble method.

$$N_1 = C_{n-1}^{2} + \sum_{i=1}^{n-3} C_i^1 \qquad \text{Equation (3)}$$

$$N_2 = \sum_{i=3}^{n} C_n^i \qquad \text{Equation (4)}$$

Where n is the total number of ML methods. When n = 6, the total number of ensembles from the WHAVE method is N_1 = 16, while the total number of ensembles from the exhaustive ensemble method is N_2 = 42. In this case, WHAVE reduces the total number of ensembles needed to be computed by 61%, and saves computation time and power significantly. The more individual ML methods, the more computation time WHAVE can save.

The above hierarchical adaptive ensemble method will keep searching for the better ensemble model together with its best x value until the resulting accuracy stops improving.

D. Adaptive Method and Stopping Criteria

There are two aspects relating to the adaptive nature of the WHAVE method. First, the method finds the optimal x value by increasing x until the accuracy stops improving. Second, the hierarchical ensemble method will keep creating the next level of ensemble until either the accuracy stops improving or the end of the ensemble tree is reached.

The above adaptive hierarchical ensemble method will keep searching for the better ensemble model and best x value until the resulting accuracy stops improving.

E. Implementation of WHAVE Method for Breast Cancer Detection

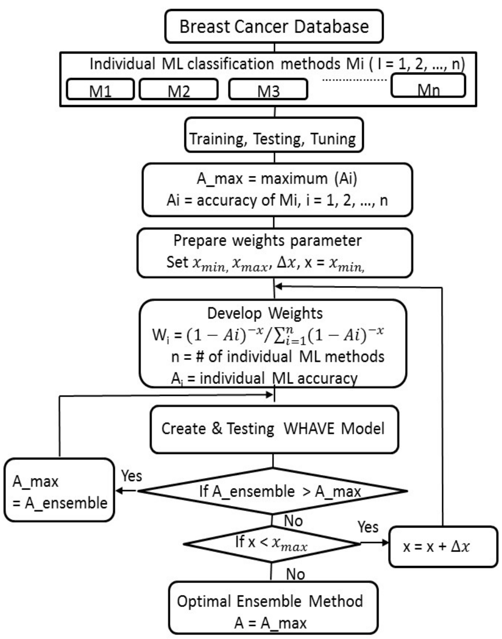

Fig. 3. Implementation flowchart of WHAVE method for breast cancer detection

Fig. 3 is the program implementation flowchart of the WHAVE method for breast cancer detection. First the individual machine learning classification methods are trained, tested and tuned on the database [28, 29]. In this work we picked 4 different machine learning methods to use as individual ML methods therefore n = 4: Disjunctive Normal Form (DNF) rule based method (CN2 learner) [5, 28], Decision Tree [28, 29], Support Vector Machines (SVM) [28, 29] and Naïve Bayes [28, 29, 30]. Each ML classification method goes through training, testing and tuning phases.

Three methods are based on multiple-valued logic: DNF, Decision Tree and Naïve Bayes. DNF rule based method (CN2 learner) is a logic-based method and uses binary function minimization. Decision Trees learning method is a practical inductive inference method and is based on creating a decision tree to classify the data. Naïve Bayes is a probabilistic based Machine Learning method and assumes each attribute of the data is unrelated to any other attribute. SVM is a non-probabilistic binary linear classifier and operating on continuous representation, it selects the optimal hyper plane used as the threshold for classifying the data. We intentionally selected different types of methods and different representations, believing that this should improve the results.

Each machine learning classification method is trained using a certain randomly chosen portion of the data. The methods are trained on 90%, 80%, 70%, ... , 10% of the data randomly chosen [29]. This is done to see the various accuracies of the methods when changing how much data the method is given.

The trained methods are then tested on a portion of the data [29]. Testing is always done on a randomly selected 10% of the data. For each trained method, the method is tested 100 times on different randomly selected 10% portions of the data, and then averaged.

After training, testing and tuning, the best individual method is determined and inputted into the WHAVE system. The best ensemble method is determined.

III. Results and Discussion

Table and graphs of the accuracy results of the WHAVE methods, each individual ML method, and the un-weighted majority voting method are produced to show the minimum, maximum and average accuracies of each method. Graphs that show the difference in accuracy between 90% and 10% training data and the variation of each method are generated.

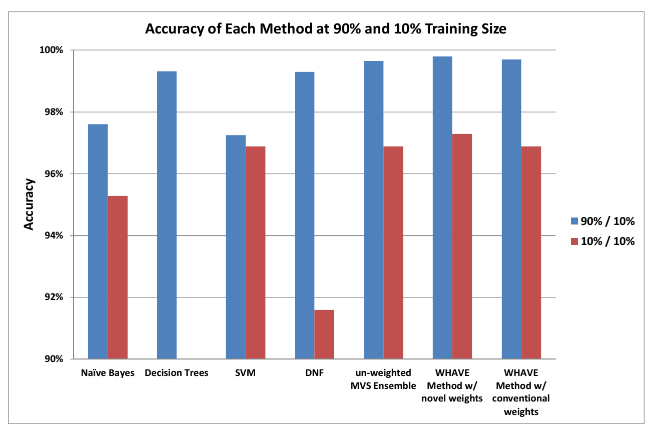

Fig. 4. Accuracy of each method at 90% & 10% training size

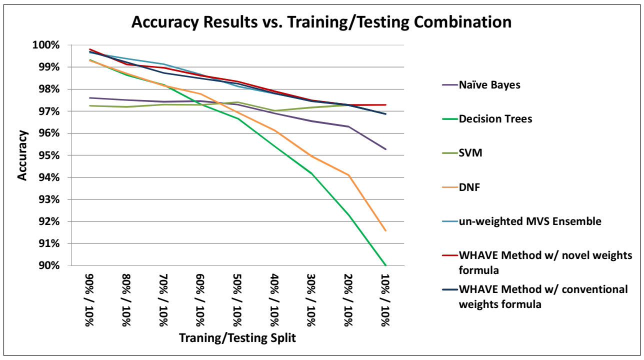

Fig. 5. Training/testing combination results

The results in Fig. 4 and 5 show:

1. The highest accuracy occurs at 90% training size.
2. WHAVE method with novel weights formula produces the highest accuracy of 99.8%, better than any individual ML method accuracy and hierarchical un-weighted MVS as well as the WHAVE with conventional weights.
3. All three ensemble methods produce better accuracy results than any individual ML method tested.
4. All three ensemble methods are reasonably stable regardless of training size with 2.5-3.2% variation of accuracy for 90% vs.10% training sizes, although SVM is most stable regardless of training size and has the least accuracy variation.
5. Given 90% training size, DNF and Decision Trees have the highest accuracy of 99.3% among the four tested individual ML methods.
6. Given 10% training size, SVM has the highest accuracy of 97% among the four tested individual ML methods.

Accuracy Evaluation

In order to compare the accuracy results of each classification method, the following evaluation parameters were used:

False positive (FP): An input without breast cancer is incorrectly diagnosed as having cancer.

False negative (FN): An input with breast cancer is incorrectly diagnosed as having no cancer.

Sensitivity = TP / (TP + FN) %
Specificity = TN / (TN + FP) %
Where,
True positive (TP): An input is correctly diagnosed as a patient with breast cancer.

True negative (TN): An input is correctly diagnosed as a patient without breast cancer.

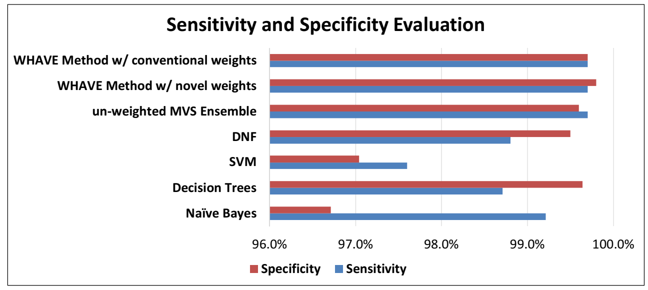

Fig. 6. Specificity and Sensitivity results comparison on top result of all training/testing combination

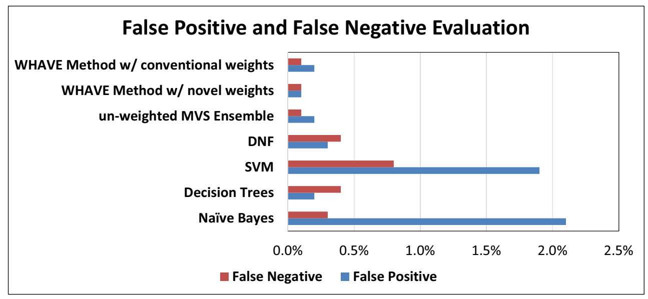

Fig. 7. False Positives and False Negatives results comparison on top result of all training/testing combination

Fig. 6 and Fig. 7 show that the WHAVE method gives the highest sensitivity and specificity values, and also the lowest false negative and positive values compared to individual ML methods.

IV. Conclusion

A new machine learning method Weighted Hierarchical Adaptive Voting Ensemble (WHAVE) was developed and implemented for breast cancer detection. The Wisconsin database was used in this work in order to compare the algorithm with existing methods. This WHAVE ensemble method includes a novel weights formula in addition to the conventional weights formula and the un-weighted MVS. A system of programs was developed that implements and compares the seven Machine Learning methods for breast cancer detection. In contrast to the previously published research, several ML methods for breast cancer were compared in a uniform and detailed way. Results showed that given a 90%/10% training/testing combination, the WHAVE method with novel weights gave the highest accuracy of 99.8%, better than any of the 4 individual ML method tested and hierarchical un-weighted MVS as well as hierarchical MVS with conventional weights. The hierarchical ensemble methods produce better accuracy results than any individual ML method tested at both 90% and 10% training size. All 3 ensemble methods are reasonably stable regardless of training size with 2.5-3.2% variation between accuracy for 90% vs.10% training sizes, while SVM is most stable regardless of training size and has least accuracy variation. WHAVE gave higher accuracy than any other methods tested in this paper, and its accuracy is also higher than any other ML method for breast cancer detection using WBCD documented in literature.

The WHAVE method in this paper is demonstrated to yield the highest accuracy out of all ML algorithms for breast cancer detection using WCBD. This method can be applied to other medical diagnosis/decision making problems, such as melanoma cancer. The method can also be applied to other decision making problems beyond medical fields such as weather forecasting.

REFERENCES

[1] Clark, EE., & Niblett, T. (1989). "The CN2 induction algorithm". Machine Learning, 3, pp. 261-284

[2] Quinlan, J. R. (1996). "Improved use of continuous attributes in C4.5". Journal of Artificial Intelligence Research, 4, 77909.

[3] Abdelghani Bellaachia, Erhan Guven (2006) "Predicting Breast Cancer Survivability Using Data Mining Techniques." Department of Computer Science, George Washington University.

[4] Übeyli, E. D. (2007). "Implementing automated diagnostic systems for breast cancer detection". Expert Systems with Applications, 33(4), 10541062. Of Medical Systems. 36 (5), pp. 2841–2847.

[5] Hsieh, S.L., et al. (2012). "Design ensemble machine learning model for breast cancer diagnosis". J. Med. Syst. 36 (5), 2841–2847.

[6] O.L. Mangasarian and W.H. Wolberg: "Cancer diagnosis via linear programming", *SIAM News*, Vol. 23, Number 5, September 1990, pp 1 & 18.

[7] Hamiton, H. J., Shan, N., & Cercone, N. (1996). "RIAC: A rule induction algorithm based on approximate classification". In International conference on engineering applications of neural networks, University of Regina, pp 17.

[8] Ster, B., & Dobnikar, A. (1996). "Neural networks in medical diagnosis: Comparison with other methods". In Proceedings of the international conference on engineering applications of neural networks, pp. 427–430.

[9] A. Sharkey. (1998). "On combining artificial neural nets," Connection Science, vol. 8, pp. 299-313, 1996.

[10] Nauck, D., & Kruse, R. (1999). "Obtaining interpretable fuzzy classification rules from medical data". Artificial Intelligence in Medicine, 16, 149–169.

[11] Pena-Reyes, C., and M. Sipper. (1999)."A fuzzy-genetic approach to breast cancer diagnosis." Artificial Intelligence in Medicine, 17: pp. 131-155.

[12] Setiono, R. (2000). "Generating concise and accurate classification rules for breast cancer diagnosis". Artificial Intelligence in Medicine, 18(3), pp.205–217.

[13] Albrecht, A. A., Lappas, G., Vinterbo, S. A., Wong, C. K., & Ohno-Machado, L. (2002). "Two applications of the LSA machine". Proceedings of the 9th international conference on neural information processing, pp. 184–189.

[14] Abonyi, J., & Szeifert, F. (2003). "Supervised fuzzy clustering for the identification of fuzzy classifiers". Pattern Recognition Letters, 14(24), 2195–2207.

[15] Martinez-Munoz, G., & Suarez, A. (2005). "Switching class labels to generate classification ensembles". Pattern Recognition, 38, 1483-1494.

[16] Sahan, S., Polat, K., Kodaz, H., et al., "A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis". Comput. Biol. Med. 37(3):415-423, 2007.

[17] Polat, K., & Gunes, S. (2007). "Breast cancer diagnosis using least square support vector machine." Digital Signal Processing, 17(4), pp. 694–701.

[18] Guijarro-Berdias, B., Fontenla-Romero, O., Perez-Sanchez, B., & Fraguela, P. (2007). "A linear learning method for multilayer perceptrons using least squares". Lecture Notes in Computer Science, pp. 365–374.

[19] Akay, M. F. (2009). "Support vector machines combined with feature selection for breast cancer diagnosis". Expert Systems with Applications, 36(2), pp. 3240–3247.

[20] Karabatak, M., & Cevdet-Ince, M. (2009). "An expert system for detection of breast cancer based on association rules and neural network". Expert Systems with Applications, 36, pp. 3465–3469.

[21] Peng, L., Yang, B., & Jiang, J. (2009). "A novel feature selection approach for biomedical data classification". Journal of Biomedical Informatics, 179(1), pp. 809–819.

[22] Marcano-Cedeño, A., J. Quintanilla-Domínguez, and D. Andina. (2011). "WBCD breast cancer database classification applying artificial metaplasticity neural network." Expert Systems with Applications, 38, pp. 9573-9579.

[23] Planey, K.,. "Machine Learning Approaches to Breast Cancer Diagnosis and Treatment Response Prediction." (2011). Stanford Biomedical Informatics.

[24] Michalski, R.S., Mozetic, I., Hong, J., & Lavrafi, N. (1986). "The multi-purpose incremental learning system AQ15 and its testing application to three medical domains". Proceedings of the Fifth National Conference on Artificial Intelligence, pp. 1041-1045. Philadelphia, PA: Morgan Kaufmann.

[25] UC Irvine Machine Learning Repository Breast Cancer Wisconsin Database https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29. Source of the database: Dr. William H. Wolberg (physician), University of Wisconsin Hospitals, Madison, Wisconsin, USA.

[26] Wolberg. W.H., and O.L. Mangasarian: "Multisurface method of pattern separation for medical diagnosis applied to breast cytology", Proceedings of the National Academy of Science, U.S.A., Vol 87, December 1990, pp. 9193-9196.

[27] "A Course in Machine Learning", Hal Daume III, pp.149-155

[28] Orange Machine Learning software suite http://orange.biolab.si/.

[29] "Machine Learning in Action", Peter Harrington. Manning Shelter Island, Manning Publications Co. Shelter Island, NY. pp 9, 11-13, 37-82, 101-127.

[30] "Bayesian Reasoning and Machine Learning", David Barber. http://www.cs.ucl.ac.uk/staff/D.Barber/brml, pp. 203-213.