International Journal of Artificial Intelligence and Neural Networks– IJAINN
Volume 5 : Issue 1 [ISSN : 2250-3749]
Publication Date: 30 April, 2015

Mammographic Mass Classification by Using a New Naïve Bayesian Classifier

Murat KARABATAK

Abstract— **Mammography is considered as the most effective method for breast cancer screening. It is effective, but it suffers from the low positive predictive value of breast biopsy resulting from mammogram interpretation leads to approximately 70% unnecessary biopsies with benign outcomes. Recently, several computer-aided diagnosis (CAD) systems have been proposed to reduce the high number of unnecessary breast biopsies. Thus, in this paper, we propose a decision support system for helping the physicians in their decision to perform a breast biopsy on a suspicious lesion seen in a mammogram or to perform a short term follow-up examination instead. To accomplish this aim, we used a weighted Bayesian classifier. Naïve Bayesian (NB) is known to be the simple classifier and there have been so many applications in the literature. We conduct several experiments to evaluate the performance of the weighted NB on mammographic mass classification database. The experiments were realized with 5-fold cross validation test. Moreover, various performance evaluation metrics such as sensitivity, specificity and accuracy were considered. According to the experiments, the weighted NB obtained the following evaluation values. The calculated sensitivity, specificity and the accuracy values are 85.64%, 86.75% and 86.18%. Moreover, a comparison with the existing methods in the literature was presented. As a result, performance evaluating metrics of weighted NB are better than NB and many other existing methods.**

***Keywords*— NB classifier, Weighted NB classifier, Mammographic Mass Classification, Performance evaluation tests.**

I. Introduction

Breast cancer, which is the second most common cancer type after lung cancer and the fifth most common cause of cancer death, is very common and serious cancer for women [1]. Mammography is a traditional method that has been used to detect the breast cancer [2]. Interpreting mammography necessitates highly skilled radiologists because in literature, radiologists show considerable variation in interpreting a mammography [3]. In addition, the low positive predictive value of breast biopsy resulting from mammogram interpretation leads to a vast of unnecessary biopsies with benign outcomes.

Murat KARABATAK
College of Technology / Firat University
Turkey

Breast calcifications are often seen in mammography and most of them are benign calcifications, but especially calcifications smaller than 1 mm are the most precise mammographic finding of early breast cancer. 70 % of in situ carcinomas manifest themselves only through micro calcifications. Therefore, identification of calcifications constitutes an important field of mammographic evaluation [4]. In recent years, computer assisted diagnostic (CAD) systems have been designed based on BI_RADS [5] standards. This helps to determine tissue deformation so that the doctor can follow the suspected area or perform a breast biopsy. Generally, BI_RADS features are gathered from different radiology centers for a suspected area seen in the mammography for different BI-RADS features such as shape and boundary of the mass provided by differentially trained physicians.

As it is known, the artificial intelligence and machine learning techniques have been applied in predicting the mammographic mass classification [6]. The objective of these identification techniques is to assign patients to either a 'benign' group that does not have breast cancer or a 'malignant' group who has strong evidence of having breast cancer [7-10]. Up to now, there have been many proposed techniques for classification of breast cancer patterns with high classification accuracies. In [11], a decision tree method (C 4.5) was used for breast cancer detection with 94.74% classification accuracy. In [12], a rule induction algorithm based on approximate classification method was applied to breast cancer detection problem. The obtained accuracy was 94.99%. In [13], linear discriminant analysis (LDA) and neural networks (NN) methods were proposed to classify the breast cancer. The accuracy of the proposed LDA+NN was 96.8%. In [14], a support vector machines classifier was used and the obtained classification accuracy was 97.2%. In [15] a classification scheme which was based on a feed forward neural network rule extraction algorithm was proposed. The reported accuracy was 98.10%. A neuro-fuzzy technique was proposed by Nauck and Kruse [16]. The accuracy was 95.06%. In [17], an AR+NN method was proposed to use in breast cancer diagnosis problem. The obtained classification accuracy was 97.4%. In [3], three different methods, optimized learning vector quantization (LVQ), big LVQ, and artificial immune recognition system (AIRS), were applied and the obtained accuracies were 96.7%, 96.8%, and 97.2%, respectively. In [18], a supervised fuzzy clustering technique was proposed for breast cancer detection. The accuracy of 95.57% was obtained. In [19], Ubeyli presented the mixture experts (ME) network structure for breast cancer diagnosis, the obtained total classification accuracy was 98.85%.

In this paper, a weighted NB classifier was proposed for classification of the breast cancer. The NB classifier is one of the simple yet powerful classification methods. But it suffers from the crisp classes assigned to the training data [20]. So, in this paper, we used the weighted version of the NB classifier. In optimization of the weights of the classifier, a grid search mechanism was adopted. Several experimental works were conducted on mammographic mass dataset. The obtained classification accuracy was 86.18%.

The remainder of this paper is organized as follows. In section 2, we briefly gave the theory of NB and its extension to weighted NB approach. Moreover, in section 2, the description of Mammographic Mass Data Set was given. In section 3, the experimental work and the performance evaluation of the proposed method was given. Finally, in section 4, we concluded the paper.

II. Preliminaries

A. *Bayesian Theorem*

Bayesian theorem is an important subject in probability theory and statistics. It shows a relationship between conditional probabilities and marginal probabilities for random variables [21]. Let P(A) be the prior probability which is the initial degree of belief in A and P(B) be the prior probability which is the initial degree of belief in B. P(A|B) is the conditional probability that the degree of belief in A, having taken B into account. Bayesian theorem is stated mathematically as the following simple form;

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \qquad (1)$$

A. *NB Classifier*

Let D be a training set of tuples and their associated class labels, and each tuple is represented by an n-D attribute vector X = (x1, x2, …, xn). Suppose there are m classes C1, C2, …, Cm. Classification is to derive the maximum posteriori, i.e., the maximal P(Ci|X). This can be derived from Bayes' theorem as following [22];

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \qquad (2)$$

Since P(X) is constant for all classes, only needs to be maximized

$$P(C_i|X) = P(X|C_i)P(C_i) \qquad (3)$$

A simplified assumption in NB is that the attributes are conditionally independent (i.e., no dependence relation between attributes). So, the class assignment of the test samples are based on the following equations;

$$P(X|C_i) = \prod_{k=1}^{n} P(X_k|C_i) \qquad (4)$$

$$\arg\max_{C_i} \{P(X|C_i)P(C_i)\} \qquad (5)$$

For example, if a new sample comes and it's posterior probability P(C2|X) is the highest among all the P(Ck|X) for all the k classes, it belongs to C2 class according to the NB rule.

B. *Weighted NB Classifier*

As it was mentioned earlier, the NB classifier has drawbacks and the weighted NB was proposed to overcome them. In addition, in NB, all attributes equally contribute in calculating the posterior probabilities which cannot always represent in several applications. To address this problem, this paper extends NB by weighting attribute value in posterior probability computation. Therefore, a weight parameter was added to the Eq. 4 and it is presented in the following equations;

$$P_w(X|C_i) = \prod_{k=1}^{n} (W_k + P(X_k|C_i)) \qquad (6)$$

$$\arg\max_{C_i} \{ P_w(X|C_i)P(C_i) \} \qquad (7)$$

In Weighted NB classifier, if the weights are supposed to be equal, than the W_k=0. An illustration of the weights in a product operation is given in fig. 1.

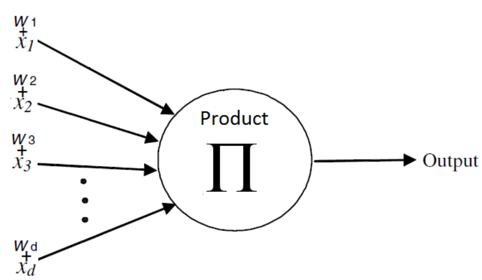

Figure 1. Weighting operation for Product function

C. *Mammographic Mass Database*

In this paper, the mammography data set from the UCI machine learning repository [23] was used. This data set collected from the Institute of Radiology of the University Erlangen-Nuremberg between 2003 and 2006. It can be used to predict the severity (benign or malignant) of a mammographic mass lesion from BI-RADS attributes and the patient's age. It contains a BI-RADS assessment, the patient's age and three BI-RADS attributes together with the ground truth (the severity field) for 516 benign and 445 malignant masses that have been identified on full field digital mammograms collected. Each instance has associated BI-RADS assessment ranging from 1 (definitely benign) to 5 (highly suggestive of malignancy) assigned in a double-review process by physicians. Assuming that all cases with BI-RADS assessments greater or equal a given value (varying from 1 to 5), are malignant and the other cases benign, sensitivities and associated specificities can be calculated. These can be an indication of how well a CAD system performs compared to the radiologists. Class Distribution: benign: 516; malignant: 445

6 Attributes in total (1 goal field, 1 non-predictive, 4 predictive attributes)

1. BI-RADS assessment: 1 to 5 (ordinal, non-predictive!)
2. Age: patient's age in years (integer)
3. Shape: mass shape: round=1 oval=2 lobular=3 irregular=4 (nominal)
4. Margin: mass margin: circumscribed=1 microlobulated=2 obscured=3 ill-defined=4 spiculated=5 (nominal)
5. Density: mass density high=1 iso=2 low=3 fat-containing=4 (ordinal)
6. Severity: benign=0 or malignant=1 (binominal, goal field!)

III. Performance Evaluation

The experiments were evaluated with 5-fold cross validation test. In 5-fold cross validation dataset is randomly split into 5 exclusive subsets of approximately equal size and the holdout method is repeated 5 times. At each time, one of the 5 subsets is used as the test set and the other 4 subsets are put together to form a training set. The advantage of this method is that it is not important how the data is divided. Every data point appears in a test set only once, and appears in a training set 4 times. The performance of the proposed weight NB classification method was evaluated with sensitivity, specificity and accuracy tests [24]. Sensitivity, specificity and accuracy terms are commonly used statistics in pattern recognition applications [20]. Moreover, True positive (TP), true negative (TN), false negative (FN), and false positive (FP) terms are commonly used along with the description of sensitivity, specificity and accuracy, where TP is the number of true positives, which means that some cases with ‘positive’ class is correctly classified as positive; FN, the number of false negatives, which means that some cases with the ‘positive’ class is classified as negative; TN, the number of true negatives, which means that some cases with the ‘negative’ class is correctly classified as negative; and FP, the number of false positives, which means that some cases with the ‘negative’ class is classified as positive. Thus, sensitivity, specificity and accuracy are described in the following equations;

$$Sensitivity = \frac{TP}{TP + FN} \tag{8}$$

$$Specificity = \frac{TN}{TN + FP} \tag{9}$$

$$Acuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

IV. Experimental Results

As we declared earlier, Mammographic Mass database with 5 attributes and 961 records was used in the experimental works. It is worth to mention that 136 records were discarded because of the missing values. Moreover, 5-fold cross validation test was applied and average values were calculated for performance measurements. When 5-fold cross validation test was applied, the training dataset contained 660 samples for training set and the retained 165 samples were used in the test set.

The weights that were used in new NB approach were determined with a heuristic search algorithm. In other words, a grid search mechanism was employed in the likelihood domain for obtaining the maximum cost value. For each attribute, an optimum weight value was searched that gave the maximum cost value. Thus, the related weights were obtained for subsequent NB calculations. According to the experimental works the constructed confusion matrix was given in Table 1.

TABLE 1. CONFUSION MATRIX FOR OBTAINED RESULTS

		Actual	
		Malignant	Benign
Classified	Malignant	364(TP)	53 (FP)
	Benign	61(FN)	347(TN)

As one can see in Table 1, 364 malignant samples were classified as malignant. Thus, the true positive of the confusion matrix was 364. In addition, 53 benign samples were detected as malignant which indicated the false positives. There were 61 false negative samples and 347 benign samples were classified as benign which symbolized the true negatives. The calculated sensitivity, specificity and the accuracy values are 85.64 %, 86.75% and 86.18% respectively.

On the other hand, in Table 3 and Table 4, the correct classification rate for each fold was given. In table 3, NB approach was used and in table 4 new NB approach was used. Moreover, the number of the correctly classified samples and the number of miss-classified samples were given accordingly. It is evident in Table 4 that the best performances were obtained in the fourth folds where the calculated accuracy was 89.09 %. As we indicated in the previous paragraph the overall accuracy was 86.18 %.

TABLE 3 CLASSIFICATION ACCURACIES FOR EACH FOLD FOR NB APPROACH

Folds	Number of Training Data	Number of Test Data	Correct classified	Miss classified	Correct classification rate (%)
Fold 1	660	165	132	33	80.00
Fold 2	660	165	143	22	86.66
Fold 3	660	165	138	27	83.63
Fold 4	660	165	143	22	86.66
Fold 5	660	165	135	30	81.81
TOTAL		**825**	**691**	**134**	**83.75**

TABLE 4 CLASSIFICATION ACCURACIES FOR EACH FOLD FOR NEW NB APPROACH

Folds	Number of Training Data	Number of Test Data	Correct classified	Miss classified	Correct classification rate (%)
Fold 1	660	165	135	30	81.81
Fold 2	660	165	146	19	88.48
Fold 3	660	165	142	23	86.06
Fold 4	660	165	147	18	89.09
Fold 5	660	165	141	24	85.45
TOTAL		**825**	**711**	**114**	**86.18**

V. Conclusions

The NB classifier is considered as one of the simple yet powerful classification methods. But it has several drawback such as the crisp classes assigned to the training data. In order to overcome the drawbacks of the NB classifier, a weighted NB was proposed and its application on breast cancer detection was presented. Based on the conducted experimental works, the applied weighted NB obtained 85.64 % sensitivity, 86.75 % specificity and 86.18 % the accuracy values respectively. We also compared the proposed method with Naïve Bayesian method.

It is also worth to mention the drawbacks of the proposed. As we described in the experimental part of the paper, our algorithm used a grid search mechanism to find the optimum weight values. This search was computationally expensive and the initialization of the weights vector is crucial and application dependent. To overcome the drawback of the weighted NB approach, genetic algorithms and immune system will be investigated in the future works.

References

[1]. Chen, H. L., Yang, B., Wang, G., Wang, S. J., Liu, J., & Liu, D. Y. (2012). Support vector machine based diagnostic system for breast cancer using swarm intelligence. Journal of medical systems, 36(4), 2505-2519.

[2]. Pena-Reyes, C. A., and Sipper, M., A fuzzy-genetic approach to breast cancer diagnosis. Artif Intell Med 17(2):131–155, 1999.

[3]. Goodman, D., Boggess, L., and Watkins, A., Artificial immune system classification of multiple-class problems. 2002: In Proceedings of the Artificial Neural Networks in Engineering, 2002, pp.179–183.

[4]. M. Elter, R. Schulz-Wendtland and T. Wittenberg (2007), The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process. Medical Physics 34(11), pp. 4164-4172

[5]. American College of Radiology Breast imaging reporting and data system (BI-RADS), 4th edn Reston, VA: American College of Radiology, 2003

[6]. Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med 2005;353:1773–83

[7]. Sahan, S., 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.

[8]. Ubeyli, E. D., Implementing automated diagnostic systems for breast cancer detection. Expert Syst Appl 33(4):1054–1062, 2007.

[9]. Polat, K., and Gunes, S., Breast cancer diagnosis using least square support vector machine. Digital Signal Process 17(4):694–701, 2007.

[10]. Akay, M. F., Support vector machines combined with feature selection for breast cancer diagnosis. Expert Syst Appl 36 (2):3240–3247, 2009.

[11]. Quinlan, J., Improved use of continuous attributes in C4. 5. J Artif Intell Res 4:77–90, 1996.

[12]. Hamilton, H.J., et al., RIAC: a rule induction algorithm based on approximate classification. 1996, University of Regina: International conference on engineering applications of neural networks.

[13]. Ster, B. and A. Dobnikar, Neural networks in medical diagnosis: Comparison with other methods. 1996: In Proceedings of the international conference on engineering applications of neural networks. p. 427–430.

[14]. Bennett, K. and J. Blue, A support vector machine approach to decision trees. 1998, in Neural Networks Proceedings,. p. 2396–2401.

[15]. Setiono, R., Generating concise and accurate classification rules for breast cancer diagnosis. Artif Intell Med 18(3):205–219, 2000.

[16]. Nauck, D., and Kruse, R., Obtaining interpretable fuzzy classification rules from medical data. Artif Intell Med 16(2):149–169, 1999.

[17]. Karabatak, M., and Ince, M. C., An expert system for detection of breast cancer based on association rules and neural network. Expert Syst Appl 36(2, Part 2):3465–3469, 2009.

[18]. Abonyi, J., and Szeifert, F., Supervised fuzzy clustering for the identification of fuzzy classifiers. Pattern Recognit Lett 24 (14):2195–2207, 2003.

[19]. Übeyli, E. D., A mixture of experts network structure for breast cancer diagnosis. J Med Syst 29(5):569–579, 2005.

[20]. Alhammady, H. (2007, May). Weighted Naive Bayesian Classifier. In Computer Systems and Applications, 2007. AICCSA'07. IEEE/ACS International Conference on (pp. 437-441). IEEE.

[21]. Hanson, R., Stutz, J., & Cheeseman, P. (1991). Bayesian classification theory. NASA Ames Research Center, Artificial Intelligence Research Branch.

[22]. Duda, R. O., & Hart, P. E. (1973). Pattern classification and scene analysis (Vol. 3). New York: Wiley.

[23]. J. Newmann, S. Hettich, C. L. Blake, and C. J. Merz, "UCI repository of machine learning databases", Univeristy of California, Irvine, Dept. of Information and Computer Sciences, http://www.ics.edu/simmlearn/MLRepository.html,1998.

[24]. Sengur, A. (2008). An expert system based on principal component analysis, artificial immune system and fuzzy k-NN for diagnosis of valvular heart diseases. Computers in Biology and Medicine, 38(3), 329-338.

About Author:

Mr. Dr. Murat Karabatak received BS degree in 1999, Master degree in 2002 and PhD title in Electric and Electronics Engineering Department at Firat University in 2008. He worked for University of Sam Houston State (Texas-USA) as visitor researcher in 2014. Assist. Prof. Dr. Karabatak is vice president of Software Engineering Department now. His interest areas are Data Mining, Big Data, intelligent systems, image processing and e-learning.