2009 International Conference on Computer Engineering and Applications
IPCSIT vol.2 (2011) © (2011) IACSIT Press, Singapore

An Intelligent Ensemble Based Systems for Breast Cancer Diagnosis

M. Arfan Jaffar[1 +], Zeshan Hayder[1], Ayyaz Hussain[1] and Anwar M. Mirza[1]
[1]Department of Computer Science, FAST National University of Computer and Emerging Sciences, Islamabad, PAKISTAN
[+] Corresponding author. Tel.: + (923335375365); fax: +(92051410 3846).
E-mail address: (arfan.jaffar@nu.edu.pk).

Abstract— For accurate classification, we have proposed a technique based on ensemble classifiers. In the feature space, the distribution of data can be different in different regions. In the proposed technique, we split the input data into different blocks and generate classifiers on each block. For combination, we have used different rules among the classifiers. We have compared the performance of our proposed ensemble classifiers with the individual classifiers. As a result, we have found that our method improve performance. The paper provides the implementation details along with the corresponding results for all the assessed classifiers. Several comparative studies have been carried out concerning diagnosis problem demonstrating the superiority of the proposed algorithm in terms of sensitivity, specificity and accuracy

Keywords- breast cancer, ensemble classifier, support vector machine

I. INTRODUCTION

Breast cancer is the most common cancer diagnosed among U.S. women. After lung cancer, breast cancer is the second leading cause of cancer death in women. An estimated 182,480 new cases of invasive breast cancer are expected to be diagnosed in American women during 2008. It affects one of every 10 women in Europe and one of every eight in the United States [1]. Early diagnosis and early treatment are the best ways to reduce deaths due to breast cancer. Early diagnosis requires an accurate and reliable diagnostic procedure that allows physicians to distinguish benign from malignant breast tumors, and finding such a procedure is an important goal. Current procedures for early detection and diagnosis of breast cancer include self-examination, mammography [2,3,4], and ultra-sonography (US) [5]. In 1995, Stavros et al [6] showed that the sensitivity of breast US for malignancy was 98.4%; the specificity, 67.8%; and the overall accuracy, 72.9%. These results were achieved by experienced radiologists.

In mathematics, a classifier is a mapping from a (discrete or continuous) feature space X to a discrete set of labels Y. Classifiers may either be fixed classifiers or learning classifiers, and learning classifiers may in turn be divided into supervised and unsupervised learning classifiers. The applications of classifiers are wide-ranging. They find use in medicine (drug trial analysis, MRI data analysis), finance (share analysis, index prediction), mobile phones (signal decoding, error correction), computer vision (face recognition, target tracking), voice recognition, data mining (supermarket purchasing analysis, retail customer analysis) and uncountable other areas. An example is a classifier that accepts a person's salary details, age, marital status, home address and credit history and classifies the person as acceptable/unacceptable to receive a new credit card or loan.

In the supervised learning, ensemble learning is not to find the best one hypothesis for explaining given training data. It is used to create a set of hypotheses and predict the class of new data by combining the set of hypotheses like majority voting [7]. These ensemble classifiers often have been experimentally shown to be more effective than single classifiers [1,6].

To obtain an estimate of the desired output, a set of independently trained classifiers (i.e. neural networks) whose predictions are combined in some way is called ensemble classifier. In the literature, it has shown that an ensemble classifier is generally more accurate than any single classifier in the ensemble. An effective combining scheme is simply average the predictions of the networks [7] .There are two popular approaches for ensemble i.e. Bagging and Boosting. Previous work has demonstrated that these two methods are very effective for decision trees [8] but there has been comparably little empirical testing with neural networks.
The paper is organized as follows. Section II contains a survey on previous research that is most closely related to the present work. This is followed by a detailed description of the proposed system follows (Section III). Implementation and relevant results are presented in (Section IV). Finally, Section V ends the paper with several conclusions drawn from the design and the work with the proposed system.

II. RELATED WORK

Dasarathy and Sheela's [10], started work on ensemble systems in which they partition the feature space using two or more classifiers. Hansen and Salamon [11] proved that the performance of a neural network can be improved by using an ensemble of similarly configured neural networks. Schapire proved that a strong classifier can be generated by combining weak classifiers through boosting [12]. The long list includes composite classifier systems [10], mixture of experts [13], [14], stacked generalization [15], consensus aggregation [16], combination of multiple classifiers [17], change-glasses approach to classifier selection [18], dynamic classifier selection [19], classifier fusion [14]–[16], committees of neural networks [17], voting pool of classifiers [18], classifier ensembles [17], [19], and pandemonium system of reflective agents [20], among many others. Classifier selection and classifier fusion are generally two approaches that are used for ensemble. In classifier selection, each classifier is trained to become an expert in some local area of the total feature space [21]. The combination of the classifiers is then based on the given feature vector: the classifier trained with data closest to the vicinity of the feature vector, in some distance metric sense, is given the highest credit. In classifier fusion all classifiers are trained over the entire feature space. In this case, the classifier combination process involves merging the individual (weaker) classifier designs to obtain a single (stronger) expert of superior performance. Bagging and boosting are the two approaches that lie in this category [21].

III. PROPOSED METHOD

We have used Support Vector Machines (SVM) in our experiments for classification. SVM can be used as binary classifier. We have data in pair form, i.e. {(x1, y1), (x2, y2)… (xi, yi)… (xn, yn) }. For binary classification, Yi represents a Class Label; it's either 1 or -1 and Xi represents the data point. It's an n-dimensional vector. Vector is of normalized values i.e. [0 1]. Its main task is to place a linear boundary between the two different classes. For binary SVM classification we are using the method by Boser, Cortes and Vapnik which solves the following primal problem [22]

$$\min_{w,b,E} \frac{1}{2} W^T W + C \sum_{i=1}^{l} \varepsilon_i \qquad (1)$$

Subject to $\quad y_i\left(W^T \phi(x_i) - b\right) \geq 1 - \varepsilon_i \qquad \varepsilon_i \geq 0, i = 1, 2, ..., n$

Ensemble Based Systems

Instead of using results of a single classifier, we first obtain results from differently trained classifiers and then make decision as to which result would be best. This is very much similar to our decisions in ordinary life where we seek multiple opinions before making a decision [21]. Strategy in these experiments using ensemble based systems is to create a number of classifiers and combining their outputs using some combination rules such that it significantly improves the generalization performance as compared to single classifier systems. In our experiments, we are using SVM for ensembles and then combining their results using some algebraic combination rules. In our proposed method, first of all we divide data using K-fold splitting and then we are generating different overlapping datasets to train a combination of classifiers. SVM classifier with RBF Kernel has been used for these experiments. Figure (1) shows the abstract level representation of framework for K-fold data splitting used for these kinds of experiments. Initially whole training data is divided into K sub datasets randomly. Figure (1) shows continuous five data sub blocks but these are actually selected randomly. For training, each classifier is trained with different but overlapping datasets. The output of these classifiers is the probabilities with which input belongs to N output

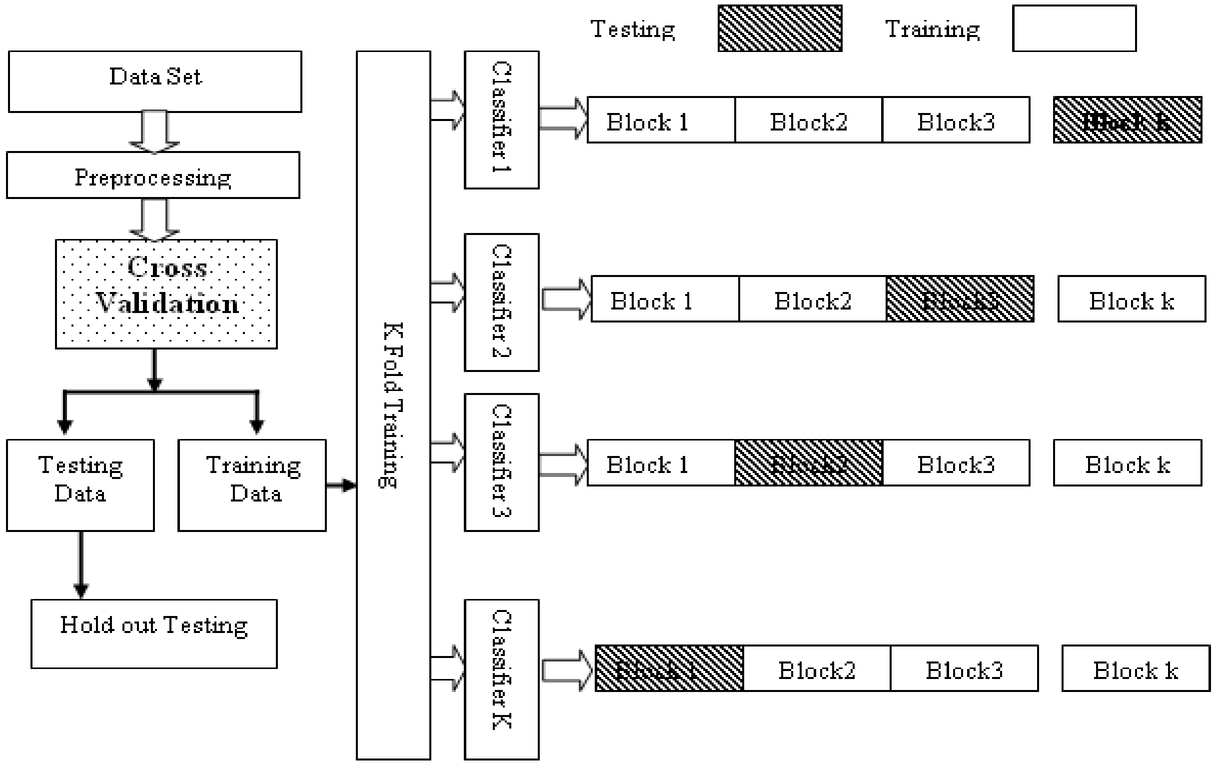
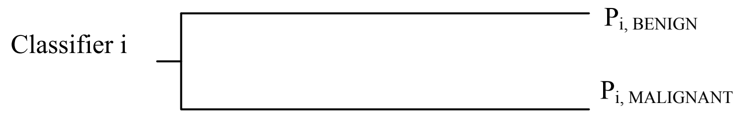


Figure 1: Architecture of an ensemble based system

classes. We are using K-1 blocks for training and using one remaining block for testing. We are using here K classifier combinations. All of these classifiers are binary-class classifiers. By using this we are having K probabilities of classifiers and a total of K*N probabilities. We are using different rules on these K*N probabilities to combine the results of these K classifiers. To combine classifier, we have used different combination rules. For a given class, the output produced by a classifier is used as the degree of support given to that class, and it is usually accepted as an estimate of the posterior probability for that class. For combining classifier, we have used different rules as shown below.



Product Rule:
$$C_x = MAX\left(\prod_{i=1}^{5} P_{i,BENIGN}, \prod_{i=1}^{5} P_{i,MALIGNANT}\right) \quad (2)$$

Median Rule:
$$C_x = MAX\left(MED\left(P_{i,BENIGN}\right), MED\left(P_{i,MALIGNANT}\right)\right) \quad (3)$$

Sum Rule:
$$C_x = MAX\left(\sum_{i=1}^{5} P_{i,BENIGN}, \sum_{i=1}^{5} P_{i,MALIGNANT}\right) \quad (4)$$

Mean Rule:
$$C_x = MAX\left(\frac{1}{5}\sum_{i=1}^{5} P_{i,BENIGN}, \frac{1}{5}\sum_{i=1}^{5} P_{i,MALIGNANT}\right) \quad (5)$$

IV. EXPERIMENTAL RESULTS AND DISCUSSION

We have implemented the proposed system by using the MATLAB environment on PENTIUM IV, processor speed 1.7 GHz. The WDBC dataset [23] is the results of the efforts made at the University of Wisconsin Hospital for the diagnosis and prognosis of breast tumors solely based on FNA test. This test involves fluid extraction from a breast mass using a small-gauge needle and then visual inspection of the fluid under a microscope.

Single Classifier System

We started our experiments using single classifier system. For these experiments, we have used SVM with RBF Kernel for classification. These are holdout experiments and Cross validation indices are used to divide 80% data into training and 20% data into testing subset. We generate results by repeating experiments 100 times and taking the average value of results. We are again and again repeating the same process multiple times to cancel out the random effect. Training and testing results are shown in Table (1). Through these experiments, we observed that although our training results are better but testing results are not that good. We come to know that it's a problem with classification as our testing data is not generalized. A set of classifiers may reduce error in testing, so we switch to multiple classifier systems.

Data	**1st classifier**	**2nd classifier**	**3rd classifier**	**4th classifier**	**5th classifier**
Breast cancer	96.893	97.713	97.533	96.974	97.255

Table 1: K-Fold results on breast cancer dataset

Data	**1st classifier**	**2nd classifier**	**3rd classifier**	**4th classifier**	**5th classifier**
Breast cancer	99.051	99.343	98.759	98.905	98.978

Table 2: Hold out results on breast cancer dataset

Multiple Classifiers System

In order to reduce testing error, we used combination of classifiers. For these kinds of experiments we used the approach described in section 3.1. Training results are obtained using 4 out of 5 blocks and testing is performed on remaining one block. For each classifier, we are performing testing on separate block so we cannot combine these results. In order to make testing results combined, we pass testing blocks into all classifiers and generate output of all trained classifiers. And then we combined results by using simple algebra rules described. Following table (3) shows the training and testing results obtained after combining results of classifiers by combination rules. And Table (4) shows the results and comparison with the standard classifiers

	Product Rule	**Sum Rule**	**Mean Rule**	**Median Rule**
Accuracy	99.635	99.635	99.635	**99.781**
Sensitivity	99.438	99.438	99.438	**99.662**
Specificity	100	100	100	**100**

Table 3: Results of proposed ensemble based classifier on breast cancer dataset

Classifier	**Proposed**	NB	RBE	ADA-7	Bagging	RBC	AntMiner+	C4.5	1NN	Logit	SVM
Accuracy	**99.78**	93.33	93.50	95.96	92.98	94.03	96.40	94.69	96.40	96.53	92.81

Table 4: Comparison of proposed algorithm with previous techniques on breast cancer data set

REFERENCES

[1] Bothorel S, Meunier BB, Muller SA. Fuzzy logic based approach for semilogical analysis of microcalcification in mammographic images. Int J Intell Syst 1997; 12:819–848.

[2] Dhawan AP, Chitre Y, Kaiser-Bnoasso C, Moskowitz M. Analysis of mammographic microcalcifications using gray-level image structure features. IEEE Trans Med Imaging 1996; 15:246–259.

[3] Petrick N, Chan HP, Sahiner B, Wei D. An adaptive density-weighted contrast enhancement filter for mammographic breast mass diction. IEEE Trans Med Imaging 1996; 15:59–67.

[4] Sahiner B, Chan HP, Petrick N, et al. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. IEEE Trans Med Imaging 1996; 15:598–610.

[5] Shankar PM, Reid JM, Ortega H, Piccoli CW, Goldberg BB. Use of non-Rayleigh statistics for identification of tumors in ultrasonic B-scans of the breast. IEEE Trans Med Imaging 1993; 12:687–692.

[6] Stavros AT, Thickman D, Rapp CL, Dennis MA, Parker SH, Sisney GA. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. Radiology 1995; 196:123–134.

[7] Leo Breiman, Bagging predictors, Machine Learning, v.24 n.2, p.123-140, Aug. 1996

[8] Drucker, H. Cortes, C. 1996. Boosting decision trees In Touretsky, D., Mozer, M. Hasselmo, M., Advances in Neural Information Processing Systems, 8, 479-485 Cambridge, MA. MIT Press

[9] Efron, B., & Tibshirani, R. (1993). An Introduction to the Bootstrap. Chapman and Hall, New York.

[10] B.V. Dasarathy and B.V. Sheela, “Composite classifier system design: Concepts and methodology,” Proceedings of the IEEE, vol. 67, no. 5, pp. 708–713, 1979.

[11] L.K. Hansen and P. Salamon, “Neural network ensembles,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 10, pp. 993–1001, 1990.

[12] R.E. Schapire, “The strength of weak learnability,” Machine Learning, vol. 5, no. 2, pp. 197–227, 1990.

[13] R.A. Jacobs, M.I. Jordan, S.J. Nowlan, and G.E. Hinton, “Adaptive mixtures of local experts,” Neural Computation, vol. 3, pp. 79–87, 1991.

[14] M.J. Jordan and R.A. Jacobs, “Hierarchical mixtures of experts and the EM algorithm,” Neural Computation, vol. 6, no. 2, pp. 181–214, 1994.

[15] D.H. Wolpert, “Stacked generalization,” Neural Networks, vol. 5, no. 2, pp. 241–259, 1992.

[16] J.A. Benediktsson and P.H. Swain, “Consensus theoretic classification methods,” IEEE Trans. on Systems, Man and Cybernetics, vol. 22, no. 4, pp. 688–704, 1992.

[17] L. Xu, A. Krzyzak, and C.Y. Suen, “Methods of combining multiple classifiers and their applications to handwriting recognition,” IEEE Transactions on Systems, Man and Cybernetics, vol. 22, no. 3, pp. 418–435, 1992.

[18] K. Woods, W.P.J. Kegelmeyer, and K. Bowyer, “Combination of multiple classifiers using local accuracy estimates,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 4, pp. 405–410, 1997.

[19] L.I. Kuncheva, “Change-glasses’ approach in pattern recognition,” Pattern Recognition Letters, vol. 14, no. 8, pp. 619–623, 1993.

[20] I. Bloch, “Information combination operators for data fusion: A comparative review with classification,” IEEE Transactions on Systems, Man,

[21] Robi Polikar, Ensemble Based Systems in decision making, IEEE circuits and magazine, 2006

[22] Burges C. A tutorial on support vector machines for pattern recognition. http://www.kernel-machines.org/

[23] http://ftp.ics.uci.edu/pub/machine-learning-databases/breast-cancer-wisconsin/