Breast Cancer Diagnosis on Three Different Datasets Using Multi-Classifiers

Gouda I. Salama[1], M.B.Abdelhalim[2], and Magdy Abd-elghany Zeid[3]

[1, 2, 3]Computer Science, Arab Academy for Science Technology & Maritime Transport, Cairo, Egypt

dr_gouda80@yahoo.com[1], mbakr@ieee.org[2] and Magdy_zeid83@yahoo.com[3]

Abstract—**This paper presents a comparison among the different classifiers decision tree (J48), Multi-Layer Perception (MLP), Naive Bayes (NB), Sequential Minimal Optimization (SMO), and Instance Based for K-Nearest neighbor (IBK) on three different databases of breast cancer (Wisconsin Breast Cancer (WBC), Wisconsin Diagnosis Breast Cancer (WDBC) and Wisconsin Prognosis Breast Cancer (WPBC)) by using classification accuracy and confusion matrix based on 10-fold cross validation method. Also, we introduce a fusion at classification level between these classifiers to get the most suitable multi-classifier approach for each data set. The experimental results show that in the classification using fusion of MLP and J48 with the PCA is superior to the other classifiers using WBC data set. The PCA is used in WBC dataset as a features reduction transformation method in which combines a set of correlated features. The selected attributes are: Uniformity of Cell Size, Mitoses, Clump thickness, Bare Nuclei, Single Epithelial cell size, Marginal adhesion, Bland Chromatin and Class. In WDBC data set the results show that the classification using SMO only or using fusion of SMO and MLP or SMO and IBK is superior to the other classifiers. In WPBC data set the results show that the classification using fusion of MLP, J48, SMO and IBK is superior to the other classifiers. All experiments are conducted in WEKA data mining tool.**

Keywords-Breast Cancer; Breast Cancer datase;, Classification; Performance; Feature selection.

I. INTRODUCTION

The second leading cause of death among women is breast cancer, as it comes directly after lung cancer [1]. Data mining approaches in medical domains is increasing rapidly due to the improvement effectiveness of these approaches to classification and prediction systems, especially in helping medical practitioners in their decision making. In addition to its importance in finding ways to improve patient outcomes, reduce the cost of medicine, and help in enhancing clinical studies. Although there was a great deal of public education and scientific research, Breast cancer considered the most common invasive cancer in women, with more than one million cases and nearly 600,000 deaths occurring worldwide annually [2]. Breast cancer is one of the most common cancers among Egyptian women; as it represents 18.3 % of the total general of cancer cases in Egypt and a percentage of 37.3 % of breast cancer is considered treatable disease. Early diagnosis helps to save thousands of disease victims. The age of breast cancer affection in Egypt and Arab countries is prior ten years compared to foreign countries as the disease targets women in the age of 30 in Arab countries, while affecting women above 45 years in European countries. Breast cancer comes in the top of cancer list in Egypt by 42 cases per 100 thousand of the population. However 80% of the cases of breast cancer in Egypt are of the benign kind [3]. The industrialized nations such as the United States, Australia, and countries in Western Europe witnessed the highest incidence rates. In many countries, breast cancer incidences increased during the 20th century, largely reflecting global changes in reproductive patterns and regional increases in mammography [4]. Because of social and cultural considerations, breast cancer ranks highest among women's health concerns. It is the most frequently diagnosed cancer in women. After thyroid cancer, melanoma, and lymphoma, breast cancer comes fourth in cancer incidences in women between 20 to 29 years.

Data mining and machine learning depend on classification which is the most essential and important task. Many experiments are performed on medical datasets using multiple classifiers and feature selection techniques. A good amount of research on breast cancer datasets is found in literature. Many of them show good classification accuracy.

In [5], the performance criterion of supervised learning classifiers such as Naïve Bayes, SVM-RBF kernel, RBF neural networks, Decision trees (J48) and simple CART are compared, to find the best classifier in breast cancer datasets (WBC and Breast tissue). The experimental result shows that SVM-RBF kernel is more accurate than other classifiers; it scores accuracy of 96.84% in WBC and 99.00% in Breast tissue. In [6], the performance of C4.5, Naïve Bayes, Support Vector Machine (SVM) and K- Nearest Neighbor (K-NN) are compared to find the best classifier in WBC. SVM proves to be the most accurate classifier with accuracy of 96.99%. In [7], the performance of decision tree classifier (CART) with or without feature selection in breast cancer datasets Breast Cancer, WBC and WDBC. CART achieves accuracy of 69.23% in Breast Cancer dataset without using feature selection, 94.84% in WBC dataset and 92.97% in WDBC dataset. When using CART with feature selection (PrincipalComponentsAttributeEval), it scores accuracy of 70.63% in Breast Cancer dataset, 96.99 in WBC dataset and 92.09 in WDBC dataset. When CART is used with feature selection (ChiSquaredAtrributeEval), it scores accuracy of 69.23% in Breast Cancer dataset, 94.56 in WBC dataset and 92.61 in WDBC dataset. In [8], the performance of C4.5 decision tree method obtained 94.74% accuracy by using 10-fold cross validation with WDBC dataset. In [9], the neural network classifier is used on WPBC dataset. It achieves accuracy of 70.725%. In [10], a hybrid method is proposed to enhance the classification accuracy of WDBC dataset (95.96) with 10 fold cross validation. In [11], the performance of linear discreet analysis method obtained 96.8% accuracy with WDBC dataset. In [12], the accuracy obtained 95.06% with neuron- fuzzy techniques when using WDBC dataset. In [13],

an accuracy of 95.57% was obtained with the application of supervised fuzzy clustering technique with WDBC dataset.

The rest of this paper is organized as follows: In section 2, a proposed breast cancer diagnosis model is shown. Section 3 presents the preprocessing phase including details of all used datasets. Feature extraction and selection is discussed in section 4. Training and classification phases are discussed in details in section 5. Performance evaluations criteria is discussed in section 6. Section 7 reports the experimental results and evaluation of the classification techniques. Finally, Section 8 introduces the conclusion of this paper.

II. PROPOSED BREAST CANCER DIAGNOSIS MODEL

Fig. 1 depicts the functional block diagram of the Proposed Breast Cancer Diagnosis Model. It consists of two phases namely: training and testing phases. The training phase includes four steps: acquisition, preprocess, feature extraction and feature selection, whereas the testing phase includes the same four steps in the training phase in addition to the classification step. In acquisition step, the sensor data are subject to a feature extraction and selection process for determining the input vector for the subsequent classifier. This makes a decision regarding the class associated with this pattern vector. Based on either feature selection or feature extraction, Dimensionality reduction is accomplished. In the preprocessing step, the image is prepared and filtered to clear the noise and improve the quality of the images. On the other hand, feature extraction considers the whole information content and maps the useful information content into a lower dimensional feature space. Feature selection is based on omitting those features from the available measurements which do not contribute to class separability. That is, redundant and irrelevant features are ignored. In the Classification step different classifiers are applied to get the best result of diagnosing and prognosing the tumor.

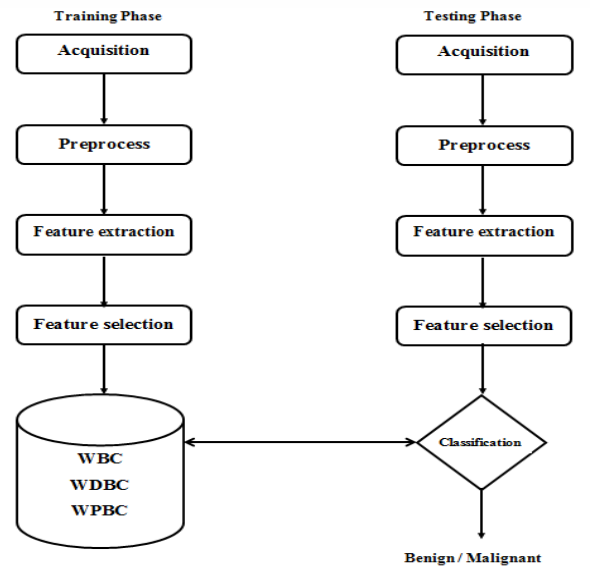

Figure 1. Proposed Breast Cancer Diagnosis Model

III. DATASET DESCRIPTION

The Wisconsin Breast Cancer datasets from the UCI Machine Learning Repository is used [14], to distinguish malignant (cancerous) from benign (non-cancerous) samples. A brief description of these datasets is presented in table 1. Each dataset consists of some classification patterns or instances with a set of numerical features or attributes.

TABLE 1 DESCRIPTION OF THE BREAST CANCER DATASETS

Dataset	No. of Attributes	No. of Instances	No. of Classes
Wisconsin Breast Cancer (Original)	11	699	2
Wisconsin Diagnosis Breast Cancer(WDBC)	32	569	2
Wisconsin Prognosis Breast Cancer(WPBC)	34	198	2

IV. FEATURE EXTRACTION AND SELECTION

An important step in breast cancer diagnosis model is Feature extraction. The Optimum feature set should have effective and discriminating features, while mostly reduce the redundancy of features space to avoid "curse of dimensionality" problem. The "curse of dimensionality" suggests that the sampling density of the training data is too low to promise a meaningful estimation of a high dimensional classification function with the available finite number of training data.

A. Wisconsin Breast Cancer Dataset [14]

The details of the attributes found in this dataset listed in table 2.

TABLE 2 WISCONSIN BREAST CANCER DATASET ATTRIBUTES

	Attribute	Domain
1	Sample code number	id number
2	Clump Thickness	1 - 10
3	Uniformity of Cell Size	1 - 10
4	Uniformity of Cell Shape	1 - 10
5	Marginal Adhesion	1 – 10
6	Single Epithelial Cell Size	1 – 10
7	Bare Nuclei	1 – 10
8	Bland Chromatin	1 – 10
9	Normal Nucleoli	1 – 10
10	Mitoses	1 - 10
11	Class	2 for benign, 4 for malignant

In the Clump thickness benign cells tend to be grouped in monolayers, while cancerous cells are often grouped in multilayer. While in the Uniformity of cell size/shape the cancer cells tend to vary in size and shape. That is why these parameters are valuable in determining whether the cells are cancerous or not. In the case of Marginal adhesion the normal cells tend to stick together, where cancer cells tend to lose this ability. So loss of adhesion is a sign of malignancy. In the Single epithelial cell size the size is related to the uniformity mentioned above. Epithelial cells that are significantly

enlarged may be a malignant cell. The Bare nuclei is a term used for nuclei that is not surrounded by cytoplasm (the rest of the cell). Those are typically seen in benign tumors. The Bland Chromatin describes a uniform "texture" of the nucleus seen in benign cells. In cancer cells the chromatin tends to be coarser. The Normal nucleoli are small structures seen in the nucleus. In normal cells the nucleolus is usually very small if visible. In cancer cells the nucleoli become more prominent, and sometimes there are more of them. Finally, Mitoses is nuclear division plus cytokines and produce two identical daughter cells during prophase. It is the process in which the cell divides and replicates. Pathologists can determine the grade of cancer by counting the number of mitoses.

B. Wisconsin Diagnosis Breast Cancer (WDBC) [14]

The details of the attributes found in WDBC dataset: ID number, Diagnosis (M = malignant, B = benign) and ten real-valued features are computed for each cell nucleus: Radius, Texture, Perimeter, Area, Smoothness, Compactness, Concavity, Concave points, Symmetry and Fractal dimension [15]. These features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image [16]. When the radius of an individual nucleus is measured by averaging the length of the radial line segments defined by the centroid of the snake and the individual snake points. The total distance between consecutive snake points constitutes the nuclear perimeter. The total distance between consecutive snake points constitutes the nuclear perimeter. The area is measured by counting the number of pixels on the interior of the snake and adding one-half of the pixels on the perimeter. The perimeter and area are combined to give a measure of the compactness of the cell nuclei using the formula $\text{perimeter}^2/\text{area}$. Smoothness is quantified by measuring the difference between the length of a radial line and the mean length of the lines surrounding it. This is similar to the curvature energy computation in the snakes. Concavity captured by measuring the size of the indentation (concavities) in the boundary of the cell nucleus. Chords between non-adjacent snake points are drawn and measure the extent to which the actual boundary of the nucleus lies on the inside of each chord. Concave Points: This feature is Similar to concavity but counted only the number of boundary point lying on the concave regions of the boundary. In order to measure symmetry, the major axis, or longest chord through the center, is found. Then the length difference between lines perpendicular to the major axis to the nuclear boundary in both directions is measured. The fractal dimension of a nuclear boundary is approximated using the "coastline approximation" described by Mandelbrot. The perimeter of the nucleus is measured using increasingly larger "rulers". As the ruler size increases, decreasing the precision of the measurement, the observed perimeter decreases. Plotting log of observed perimeter against log of ruler size and measuring the downward slope gives (the negative of) an approximation to the fractal dimension. With all the shape features, a higher value corresponds to a less regular contour and thus to a higher probability of malignancy. The texture of the cell nucleus is measured by finding the variance of the gray scale intensities in the component pixels.

C. Wisconsin Prognosis Breast Cancer (WPBC)[14]

The details of the attributes found in WPBC dataset: ID number, Outcome (R = recur, N = non-recur), Time (R => recurrence time, N => disease-free time), from 3 to 33 ten real-valued features are computed for each cell nucleus: Radius, Texture, Perimeter, Area, Smoothness, Compactness, Concavity, Concave points, Symmetry and Fractal dimension. The thirty four is Tumor size and the thirty five is the Lymph node status.

It's known from the previous lines that the diagnosis and prognosis has the same features yet the prognosis has two additional features as follows:

Tumor Size is the diameter of the excised tumor in centimeters. Tumor Size is divided into four classes: T-1 is from 0 - 2 centimeters. T-2 is from 2 - 5 cm. T-3 is greater than 5cm. T-4 is a tumor of any size that has broken through (ulcerated) the skin, or is attached to the chest wall.

Lymph node status is the number of positive auxiliary lymph nodes observed at time of surgery. The lymph nodes in the armpit (the axillary lymph nodes) are the first place breast cancer is likely to spread. As shown in fig. 2.

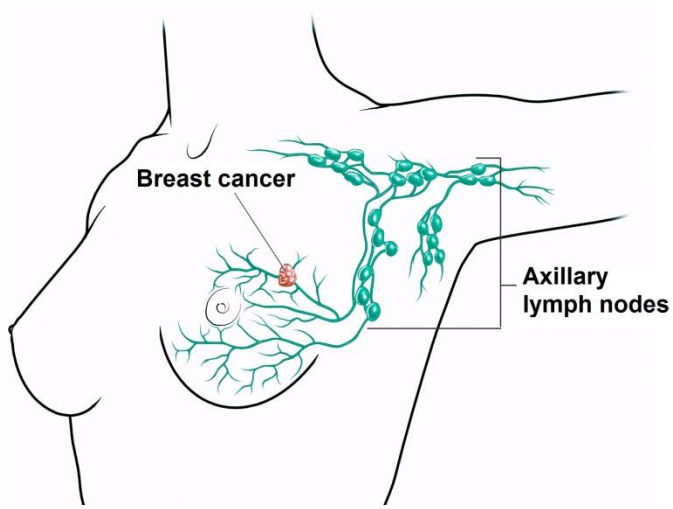

Figure 2. Axillary lymph nodes near a breast with cancer. *Illustration Copyright © 2011 Nucleus Medical Media, All rights reserved. www.nucleusinc.com*

The lymph nodes in the armpit (the axillary lymph nodes) are the first place breast cancer is likely to spread.

Lymph node status is highly related to prognosis. Lymph node-negative means the lymph nodes do not contain cancer. And Lymph node-positive means the lymph nodes contain cancer.

According to the attributes in WDBC and WPBC datasets, these attributes have 3 values with 3 columns in the data set.

- The *Mean* calculated as :

$$mean = \frac{\sum_{i=1}^{n} x_i}{n} \quad (1)$$

- The *Standard Error* calculated as:

$$S_{e=\delta\frac{s}{n}} \quad (2)$$

Where δ refers to *Standard error parameter*, s refers to *Standard deviation* and n refers to *sample size*

- *Worst mean* or *largest mean*.

Feature selection is an important step in building a classification model. It is advantageous to limit the number of input attributes in a classifier in order to have good predictive and less computationally intensive models [17]. Chi-square test and Principal Component Analysis are the two feature selection techniques proposed in this paper.

Chi-square is a statistical test commonly used for testing independence and goodness of fit. Testing independence determines whether two or more observations across two populations are dependent on each other (that is, whether one variable helps to estimate the other). Testing for goodness of fit determines if an observed frequency distribution matches a theoretical frequency distribution.

Principal Component Analysis (PCA) is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables.

V. TRAINING AND CLASSIFICATION

The following classifications are used here because they are used by most researchers for their popularity.

Naive Bayes (NB) classifier is a probabilistic classifier based on the Bayes theorem. Rather than predictions, the Naïve Bayes classifier produces probability estimates. For each class value they estimate the probability that a given instance belongs to that class. Requiring a small amount of training data to estimate the parameters necessary for classification is the advantage of the Naive Bayes classifier. It assumes that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is called class conditional independence [18].

The Multilayer Perceptron (MLP), a feed-forward back-propagation network, is the most frequently used neural network technique in pattern recognition [19] [20]. Briefly, MLPs are supervised learning classifiers that consist of an input layer, an output layer, and one or more hidden layers that extract useful information during learning and assign modifiable weighting coefficients to components of the input layers. Fig. 3 depicts the previous words. In the first (forward) pass, weights assigned to the input units and the nodes in the hidden layers and between the nodes in the hidden layer and the output, determine the output. The output is compared with the target output. An error signal is then back propagated and the connection weights are adjusted correspondingly. During training, MLPs construct a multidimensional space, defined by the activation of the hidden nodes, so that the three classes (malignant, benign and normal tissue) are as separable as possible. The separating surface adapts to the data.

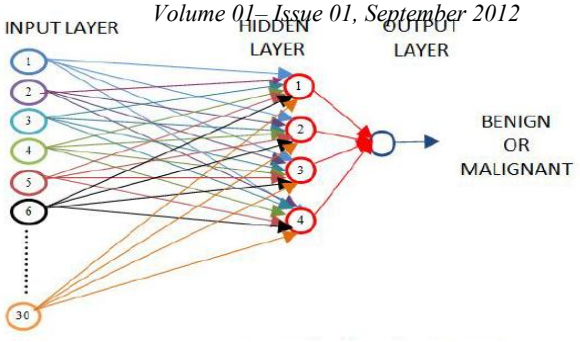

Figure 3. General architecture MLP

Support Vector Machine (SVM) is introduced by Vapnik et al. [21] it is a very powerful method that has been applied in a wide variety of applications. The basic concept in SVM is the hyper plane classifier, or linear separability. Two basic ideas are applied to achieve linear separability, SVM: margin maximization and kernels that is, mapping input space to a higher-dimension space (or feature space).

SVM projects the input data into a kernel space. Then it builds a linear model in this kernel space. A classification SVM model attempts to separate the target classes with the widest possible margin. A regression SVM model tries to find a continuous function such that maximum number of data points lie within an epsilon-wide tube around it. Different types of kernels and different kernel parameter choices can produce a variety of decision boundaries (classification) or function approximators (regression). In WEKA this classifier is called SMO.

K-Nearest Neighbor (KNN) classification [6] classifies instances based on their similarity. It is one of the most popular algorithms for pattern recognition. It is a type of Lazy learning where the function is only approximated locally and all computation is deferred until classification. An object is classified by a majority of its neighbors. K is always a positive integer. The neighbors are selected from a set of objects for which the correct classification is known. In WEKA this classifier is called IBK

Decision tree J48 implements Quinlan's C4.5 algorithm [22] for generating a pruned or un pruned C4.5 tree. C4.5 is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by J48 can be used for classification. J48 builds decision trees from a set of labeled training data using the concept of information entropy. It uses the fact that each attribute of the data can be used to make a decision by splitting the data into smaller subsets.

J48 examines the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. To make the decision, the attribute with the highest normalized information gain is used. Then the algorithm recurs on the smaller subsets. The splitting procedure stops if all instances in a subset belong to the same class. Then a leaf node is created in the decision tree telling to choose that class. But it can also happen that none of the features give any information gain. In this case J48 creates a decision node higher up in the tree using the expected value of the class.

J48 can handle both continuous and discrete attributes, training data with missing attribute values and attributes with differing costs. Further it provides an option for pruning trees after creation.

Fusion of classifiers is combining multiple classifiers to get the best accuracy. It is a set of classifiers whose individual predictions are combined in some way to classify new examples. Integration should improve predictive accuracy. In WEKA the class for combining classifiers is called Vote. Different combinations of probability estimates for classification are available.

VI. PERFORMANCE EVALUATION CRITERIA

Confusion matrix is a visualization tool which is commonly used to present the accuracy of the classifiers in classification [18]. It is used to show the relationships between outcomes and predicted classes.

The level of effectiveness of the classification model is calculated with the number of correct and incorrect classification in each possible value of the variable being classified in the confusion matrix. [23]

The entries in the confusion matrix have the following meaning in the context of our study:

- *a* is the number of correct predictions that an instance is negative,
- *b* is the number of incorrect predictions that an instance is positive,
- *c* is the number of incorrect of predictions that an instance negative, and
- *d* is the number of correct predictions that an instance is positive.

TABLE 3 CONFUSION MATRIX

		Predicted	
		Negative	*Positive*
Actual	*Negative*	a	b
	Positive	c	d

The accuracy (AC): is the proportion of the total number of predictions that were correct. It is determined using equation 3.

$$AC = \frac{a+d}{a+b+c+d} \quad (3)$$

VII. EXPERIMENTAL RESULTS

To evaluate the proposed model, three experiments were performed.

A. *Experiment (1) using Wisconsin Breast Cancer (WBC) dataset:*

Fig. 4 shows the comparison of accuracies for the five classifiers (NB, MLP, J48, SMO and IBK) based on 10-fold cross validation as a test method. The accuracy of SMO (96.9957%) is the best classifier and the accuracy obtained by NB is better than that produced by MLP, J48 and IBK.

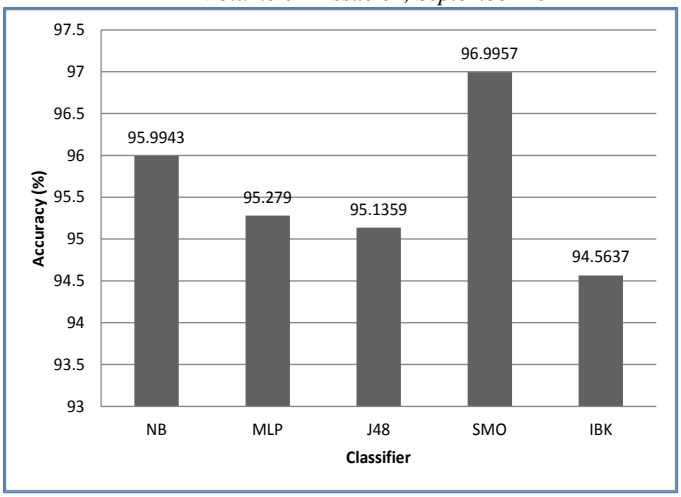

Figure 4. Single classifier in WBC

Fig. 5 shows the result of combining SMO and each of the other classifiers.it can be noticed that the fusion between SMO and MLP, SMO and IBK and between SMO and NB gives the same accuracy 96.9957%.

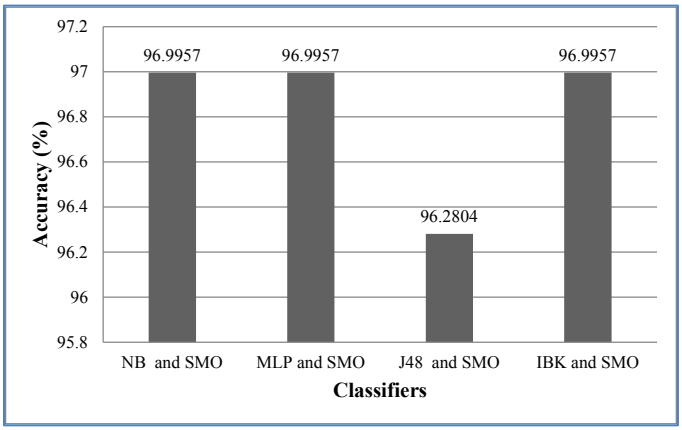

Figure 5. Fusion of two classifiers in WBC

Fig. 6 shows the result of fusion between the three classifiers SMO+IBk+NB, SMO+IBk+MLP and SMO+IBk+J48. It can be noticed that the recognition accuracy increase to 97.1388%.

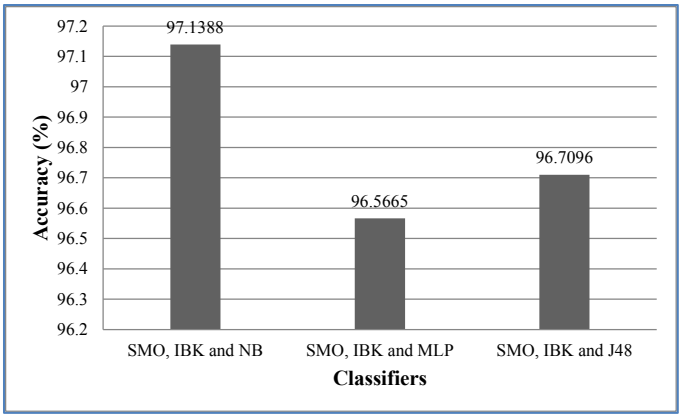

Figure 6. Fusion of three classifiers in WBC

Fig. 7 shows that the fusion between the four classifiers SMO, IBK, NB and J48 achieves accuracy (97.2818%).This

fusion is better than single classifiers, fusion of 2 classifiers, fusion of 3 classifiers and fusion of 5 classifiers.

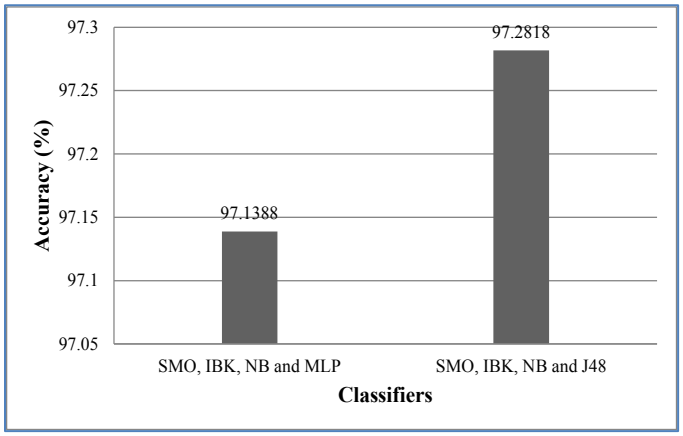

Figure 7. Fusion of four classifiers in WBC

When using features selection on WBC dataset with J48 and MLP classifiers, the best accuracy (97.568%) is got with PCA as a select attribute as shown in Fig. 8.

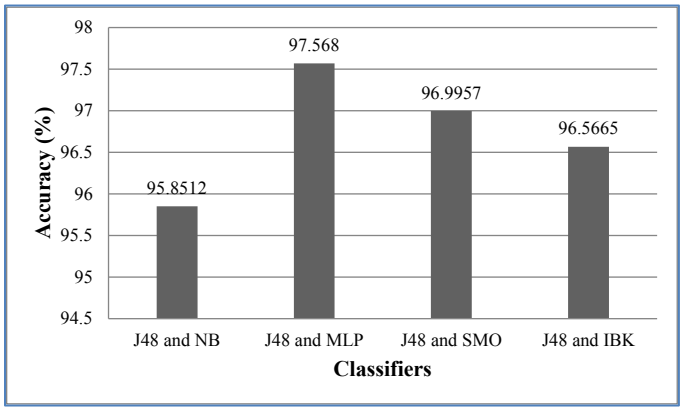

Figure 8. Fusion of two classifiers in WBC with PCA

Table 4 shows a comparison between classification accuracies of other papers and the recent proposed method for WBC dataset.

TABLE 4 COMPARISON OF EXISTING AND RECENT EXPERIMENTAL RESULTS

Method(Reference)	Classifier	Classification accuracy
[5]	SVM-RBF kernel	96.84%
[6]	SVM	96.99%
[7]	CART with feature selection (Chi-square)	94.56%
Proposed method	SMO+J48+NB+IBk	97.2818%

B. Experiment (2) using Wisconsin Diagnosis Breast Cancer (WDBC) dataset without feature selection:

Fig. 9 shows the comparison of accuracies for the five classifiers (NB, MLP, J48, SMO and IBK) based on cross validation of 10-fold as a test method. SMO is more accurate than other classifiers (97.7153%).

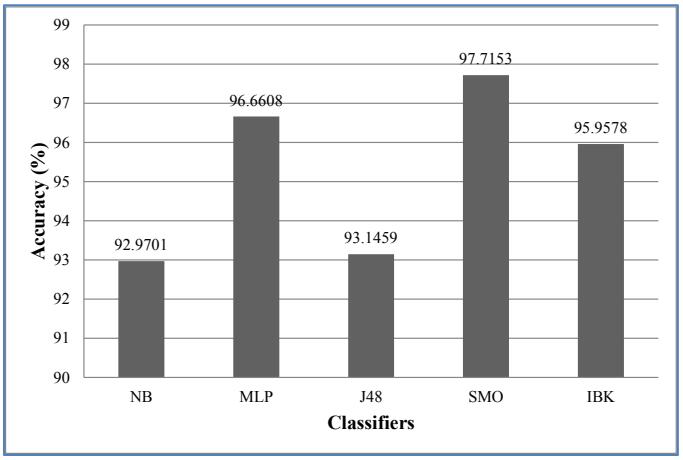

Figure 9. Single classifier in WDBC

Fig. 10 shows that fusion between SMO and each of other classifiers led to the following results: the fusion between SMO and MLP and the fusion between SMO and IBK gives the same highest accuracy as of SMO alone.

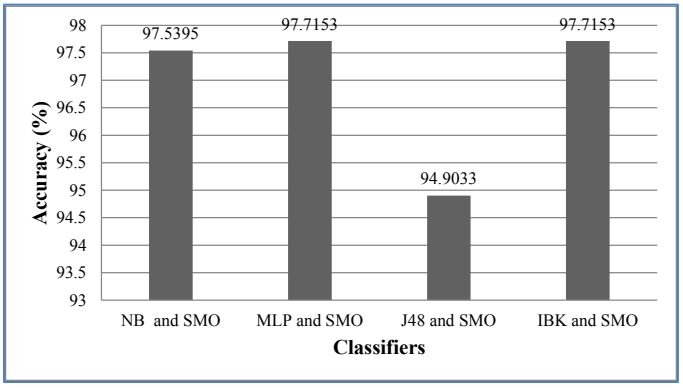

Figure 10. Fusion of two classifiers in WDBC

Fig. 11 shows that after we try to fuse SMO with each two of the other classifiers, the accuracy decreases.

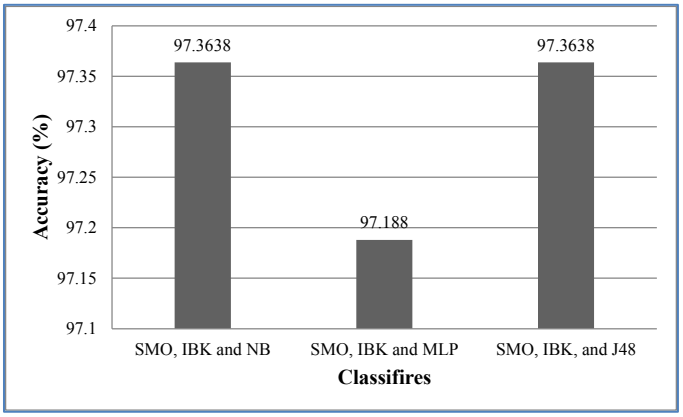

Figure 11. Fusion of three classifiers in WDBC

Fig. 12 shows that the fusion between SMO, IBK and NB with MLP increases the accuracy slightly but still lower than the highest accuracy in single classifiers and fusion of two classifiers.

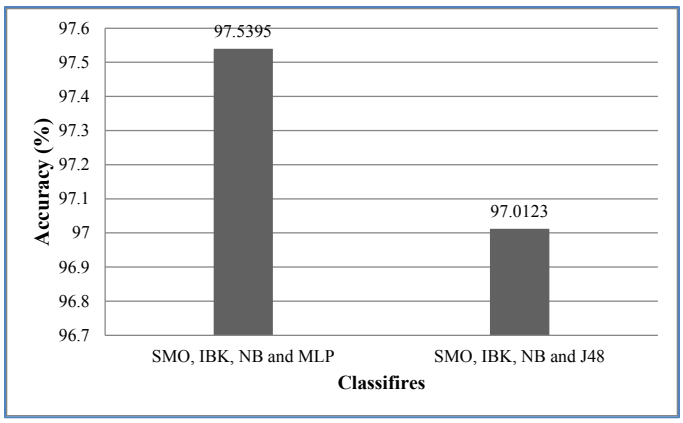

Figure 12. Fusion of four classifiers in WDBC

When using features selection on WDBC dataset with NB, MLP, J48, SMO and IBK classifiers, the results don't change.

Table 5 shows a comparison between classification accuracies of other papers and recent proposed method for WDBC dataset.

TABLE 5 COMPARISON OF EXISTING AND RECENT EXPERIMENTAL RESULTS

Method(Reference)	Classifiers	Classification accuracy
[7]	CART with feature selection (Chi-square)	92.61%
[8]	C4.5	94.74%
[10]	Hybrid Approach	95.96%
[11]	linear discreet analysis	96.8%
[12]	neuron-fuzzy	95.06%
[13]	supervised fuzzy clustering	95.57%
Proposed method	SMO	97.7153%

C. *Experiment (3) using Wisconsin Prognosis Breast Cancer (WPBC) dataset without feature selection:*

Fig. 13 shows the comparison of accuracies for the five classifiers (NB, MLP, J48, SMO and IBK) based on 10-fold cross validation as a test method. The accuracy of SMO and J48 is better than other classifiers and they are the same (76.2887%).

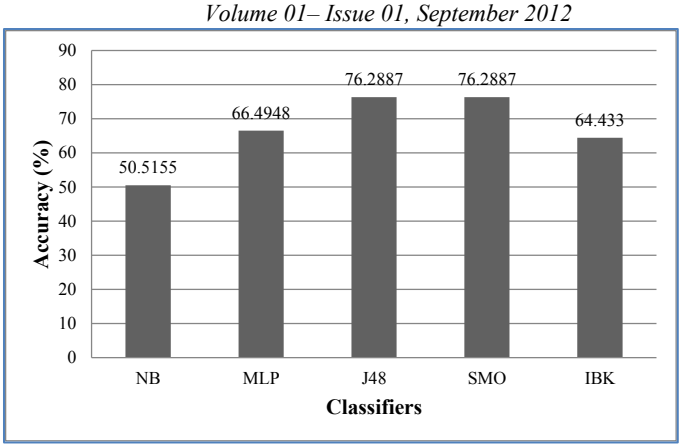

Figure 13. Single classifier in WPBC

Fig. 14 shows that the fusion between SMO and each of other classifiers led to the same accuracy (76.2887%).

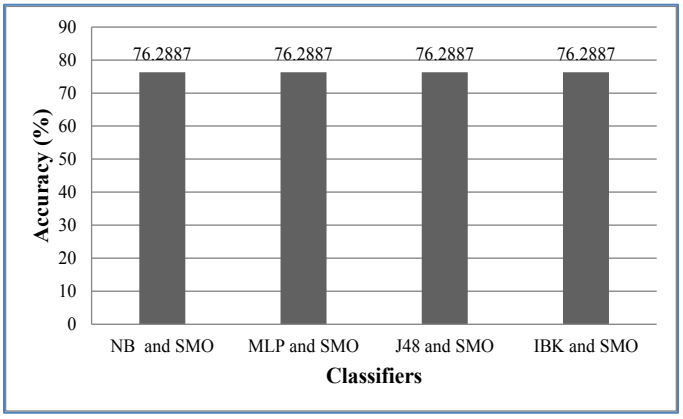

Figure 14. Fusion of two classifiers in WPBC

Fig. 15 shows that the fusion between SMO and J48 with other classifiers (NB, MLP and IBk), achieves the same accuracy (76.2887%).

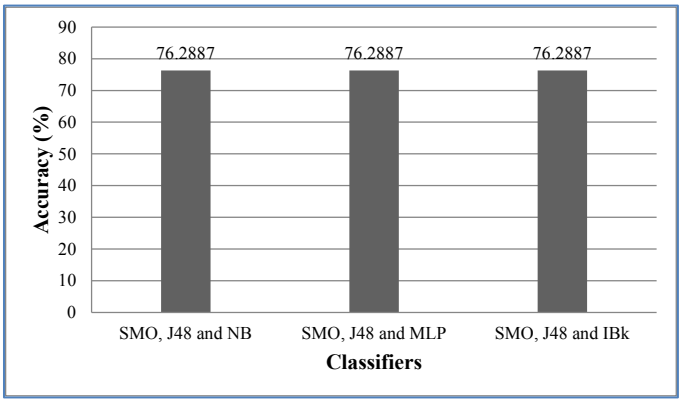

Figure 15: Fusion of three classifiers in WPBC

Fig. 16 shows that the fusion between SMO, J48, MLP and IBK is superior to the other classifiers. It achieves accuracy of (77.3196%).

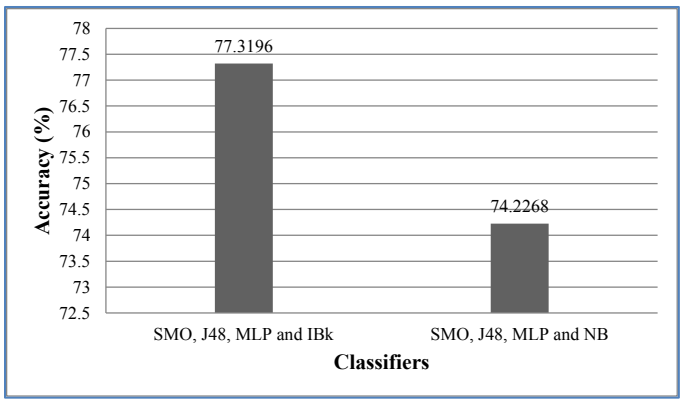

Figure 16. Fusion of four classifiers in WPBC

After using features selection with NB, MLP, J48, SMO and IBK classifiers on WDBC dataset, the outcomes didn't alter.

Table 6 shows a comparison between classification accuracies of other papers and recent proposed method for WPBC dataset.

TABLE 6 COMPARISON OF EXISTING AND RECENT EXPERIMENTAL RESULTS

Method(Reference)	Classifiers	Classification accuracy
[9]	ANN	70.725%.
Proposed method	SMO+J48-MLP+IBk	77.3196%

VIII. CONCLUSION

The experimental results in WBC dataset show that the fusion between MLP and J48 classifiers with features selection (PCA) is superior to the other classifiers. On the other hand WDBC dataset shows that using single classifiers (SMO) or using fusion of SMO and MLP or SMO and IBK is better than other classifiers. Finally, the fusion of MLP, J48, SMO and IBK is superior to the other classifiers in WPBC dataset.

IX. ACKNOWLEDGEMENT

Thank are due to Dr. William H. Wolberg at the University of Wisconsin for supporting us with the breast cancer dataset which we have used in our experiments.

X. REFERENCES

[1] U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2008 Incidence and Mortality Web-based Report. Atlanta (GA): Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute; 2012.

[2] Lyon IAfRoC: World Cancer Report. *International Agency for Research on Cancer Press* 2003:188-193.

[3] Elattar, Inas. "Breast Cancer: Magnitude of the Problem",Egyptian Society of Surgical Oncology Conference, Taba,Sinai, in Egypt (30 March – 1 April 2005).

[4] Daniel F. Roses (2005). Clinical Assessment of Breast Cancer and Benign Breast Disease, In: *Breast Cancer*: Vol. 2, Ch. 14, M. N. Harris [editor], Churchill Livingstone, Philadelphia.

[5] S. Aruna et al. (2011). Knowledge based analysis of various statistical tools in detecting breast cancer.

[6] Angeline Christobel. Y, Dr. Sivaprakasam (2011). An Empirical Comparison of Data Mining Classification Methods. International Journal of Computer Information Systems,Vol. 3, No. 2, 2011.

[7] D.Lavanya, Dr.K.Usha Rani,..," Analysis of feature selection with classification: Breast cancer datasets",Indian Journal of Computer Science and Engineering (IJCSE),October 2011.

[8] E.Osuna, R.Freund, and F. Girosi, "Training support vector machines: Application to face detection". Proceedings of computer vision and pattern recognition, Puerto Rico pp. 130–136.1997.

[9] Vaibhav Narayan Chunekar, Hemant P. Ambulgekar (2009). Approach of Neural Network to Diagnose Breast Cancer on three different Data Set. 2009 International Conference on Advances in Recent Technologies in Communication and Computing.

[10] D. Lavanya, "*Ensemble Decision Tree Classifier for Breast Cancer Data*," International Journal of Information Technology Convergence and Services, vol. 2, no. 1, pp. 17-24, Feb. 2012.

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

[12] T.Joachims, Transductive inference for text classification using support vector machines. Proceedings of international conference machine learning. Slovenia. 1999.

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

[14] Frank, A. & Asuncion, A. (2010). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[15] Street WN, Wolberg WH, Mangasarian OL. Nuclear feature extraction for breast tumor diagnosis. Proceedings IS&T/ SPIE International Symposium on Electronic Imaging 1993; 1905:861–70.

[16] William H. Wolberg, M.D., W. Nick Street, Ph.D., Dennis M. Heisey, Ph.D., Olvi L. Mangasarian, Ph.D. computerized breast cancer diagnosis and prognosis from fine needle aspirates, Western Surgical Association meeting in Palm Desert, California, November 14, 1994.

[17] Chen, Y., Abraham, A., Yang, B.(2006), Feature Selection and Classification using Flexible Neural Tree. Journal of Neurocomputing 70(1-3): 305–313.

[18] J. Han and M. Kamber,"Data Mining Concepts and Techniques", Morgan Kauffman Publishers, 2000.

[19] Duda, R.O., Hart, P.E.: "Pattern Classification and Scene Analysis", In: Wiley-Interscience Publication, New York (1973)

[20] Bishop, C.M.: "Neural Networks for Pattern Recognition". Oxford University Press,New York (1999).

[21] Vapnik, V.N., The Nature of Statistical Learning Theory, 1st ed., Springer-Verlag,New York, 1995.

[22] Ross Quinlan, (1993) C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA.

[23] Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. and Zanasi, A. (1998). *Discovering Data Mining: From Concept to Implementation*, Upper Saddle River, N.J., Prentice Hall.