Evaluation of Breast Cancer Tumor Classification with Unconstrained Functional Networks Classifier

5 authors, including:

Emad Elsebakhi
Sidra Medicine
39 PUBLICATIONS **606** CITATIONS

SEE PROFILE

Kanaan Faisal
King Fahd University of Petroleum and Minerals
30 PUBLICATIONS **649** CITATIONS

SEE PROFILE

Tarek Helmy
King Fahd University of Petroleum and Minerals
99 PUBLICATIONS **802** CITATIONS

SEE PROFILE

Farag Azzedin
King Fahd University of Petroleum and Minerals
70 PUBLICATIONS **1,232** CITATIONS

SEE PROFILE

Evaluation of Breast Cancer Tumor Classification with Unconstrained Functional Networks Classifier

Emad A. El-Sebakhy,* *Kanaan Abed Faisal*, T. Helmy, F. Azzedin, and A. Al-Suhaim

Department of Information & Computer Science

College of Computer Science & Engineering

King Fahd University of Petroleum & Minerals

Dhahran 31261, Saudi Arabia

dodi05, faisal, helmy, fazzedin, suhaim@ccse.kfupm.edu.sa

Abstract

This paper proposes functional networks as an unconstrained classifier scheme for multivariate data to diagnose the breast cancer tumor. The performance of this new technique is measured using two well known databases under the minimum description length criterion, the results are compared with the most common existing classifiers in both computer science and statistics literatures. This new classifier shown reliable and efficient results with better correct classification rate, and much less computational time.

Keywords*: Pattern Classification; Functional Networks; Breast cancer detection; Minimum Description Length.*

1. Introduction

Breast cancer has become a major cause of death among women in developed countries [2]. The most effective way to reduce breast cancer deaths is detect it earlier. However, earlier treatment requires the ability to detect breast cancer in early stages. Early diagnosis requires an accurate and reliable diagnosis procedure that allows physicians to distinguish benign breast tumors from malignant ones. Thus, finding an accurate and effective diagnosis method is very important. Biopsy is the best way to accurately determine whether the tumor is benign or malignant. However, it is invasive and expensive, and positive findings at biopsy for cancer are low, between 10% and 31% [1, 10, 16].
Recently, several researchers have used statistical and artificial intelligence techniques to successfully ”*predict*” breast cancer [14, 15]. Basically, the objective of these prediction techniques is to assign patients to either a ”*benign*”

group that does not have breast cancer or a ”*malignant*” group that has strong evidence of having breast cancer. Thus, breast cancer diagnostic problems are basically in the scope of the widely discussed classification problems [13]. The Breast Cancer Diagnosis (BCD) problem has attracted many researchers in computational intelligence, data mining, and statistics fields [6]. Artificial neural networks (ANNs) [9] and support vector machines [5, 11] have been recently proposed as a very effective method for pattern recognition, machine learning and data mining. Generally, liner/nonlinear discriminant analysis(LDA) and multinomial logistic regression (MLR) are the most commonly used data mining techniques to construct classification models. ANN and radial basis function (RBF), probabilistic neural network (PNN), and SVM became efficient alternatives in modeling classification problems due to their capability to capture complex nonlinear relationships among variables.

Recently, functional networks [3, 4] and [7, 8] have been introduced as a very effective scheme for the statistical pattern recognition problems and nonlinear complex prediction. In this paper, we evaluated breast cancer masses in pathologically proven tumors using functional networks based on analysis of speckles for classifying breast tumors.

The rest of the paper is organized as follows: Section 2 introduces the meaning of breast cancer and its existence world wide. Functional networks are introduced in Section 3 with general methodology, including their selection of initial architecture and learning methods. The maximum likelihood functional networks classifier methodology is explained in Section 4. Implementation and a comparative study are investigated using the two Wisconsin breast cancer databases are presented in Section 5. Section 6 contains both conclusion and future work.

---

*Corresponding author: dodi05@ccse.kfupm.edu.sa

2 Breast Cancer: An Overview

Breast cancer is the most common cancer disease among women, excluding nonmelanoma skin cancers. The information about the tumor from certain examinations and diagnostic tests are gathered using staging to determine how widespread the cancer is. The stage of a cancer is one of the most important factors in selecting treatment options, and it uses the *Tumour, Nodes and Metastasis* (TNM) system, which in turn determines treatment recommendations RWWD03. Breast cancer is a malignant tumor that has developed from cells of the breast. Breast cancer is cancer of breast tissue. Worldwide, it is the most common form of cancer in females, affecting approximately 10% of all women at some stage of their life in the Western world. Although significant efforts are made to achieve early detection and effective treatment, about 20% of all women with breast cancer will die from the disease. Although scientists know some of the risk factors (i.e. ageing, genetic risk factors, family history, menstrual periods, not having children, obesity) that increase a woman's chance of developing breast cancer, they do not yet know what causes most breast cancers or exactly how some of these risk factors cause cells to become cancerous. Research is under way to learn more and scientists are making great progress in understanding how certain changes in Deoxyribonucleic acid (DNA) can cause normal breast cells to become cancerous.

The information about the tumor from certain examinations and diagnostic tests are gathered using staging to determine how widespread the cancer is. The stage of a cancer is one of the most important factors in selecting treatment options, and it uses the *Tumour, Nodes and Metastasis* (TNM) system, which in turn determines treatment recommendations RWWD03. The TNM system is specific for each type of cancer, it is a standardized way in which the cancer care team describes the extent to which the cancer spread, where the letter T followed by a number from 0 to 4 describes the tumor's size and spread to the skin or chest wall under the breast, the letter N followed by a number from 0 to 3 indicates whether the cancer has spread to lymph nodes near the breast, and the letter M followed by a 0 or 1 indicates whether or not the cancer has spread to distant organs. Once a patient's T, N, and M categories have been determined, this information is combined in a process called stage grouping to determine a woman's disease stage. This is expressed in Roman numerals from Stage 0 (the least advanced stage) to Stage IV (the most advanced stage).

3. Functional Networks

Functional networks are a generalization of neural networks that combine both knowledge about the structure of the problem, to determine the architecture of the network, and data, to estimate the unknown functional neurons [3, 4] and [7, 8]. We call the node $X_j \in \mathbf{X}$, for all j as a *multiple node*, if it is an output of more than one neural functions. Otherwise, it is called a *simple node*.
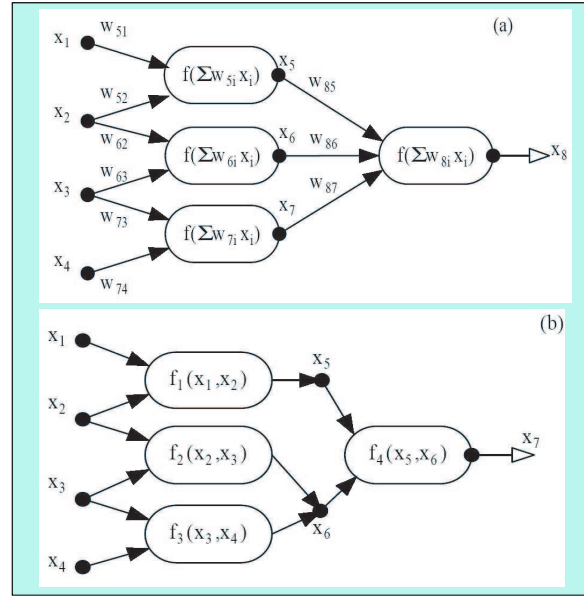

Figure 1. (a) A Neural Network, (b) A functional network.

As it is shown in Figure 1, a functional network consists of: a) several layers of storing units, one layer for containing the input data (x_i; $i = 1, \ldots, 4$), another for containing the output data (x_7) and none, one or several layers to store intermediate information (x_5 and x_6); b) one or several layers of processing units that evaluate a set of input values and delivers a set of output values (f_i) and c) a set of directed links. Functional networks extend neural networks by allowing neural functions f_i to be not only true multiargument and multivariate functions, but to be different and learnable, instead of fixed functions. In functional networks, the activation functions are unknown functions from a given family, *i.e.* polynomial, to be estimated during the learning process. In addition, functional networks allow connecting neuron outputs, forcing them to be coincident. Some differences between a neural network and a functional network are shown in Figure 1. Functional networks methodology can be more easily understood by organizing it into seven steps as it is shown in [7, 8].

4. Constrained-Functional Networks Classifier

We applied functional networks with maximum likelihood technique, [8] to investigate the breast cancer tumor. Generally, we assume that the initial functional network model is written as

$$\pi_{ik} = h_k(\mathbf{x}_i, \boldsymbol{\Theta}_k) = \mathbf{p}\left(g_k(\mathbf{x}_i, \boldsymbol{\Theta}_k)\right), \qquad (1)$$

for all $i = 1, \ldots, n$, where $\mathbf{x}_i \equiv (x_{i1}, \ldots, x_{ip})$ is the ith observation. The vector $\mathbf{\Theta}_k = [\theta_{1k}, \ldots, \theta_{pk}]^T$ is the vector of the functional network parameters needed to be learned. The functions $h_k(\mathbf{x}_i, \mathbf{\Theta}_k)$ is to be estimated based on the available breast cancer data, bearing in mind the probability restrictions on $h_k(\mathbf{x}_i, \mathbf{\Theta}_k)$. The functions $g_k(\mathbf{x}_i, \mathbf{\Theta}_k)$ are unknown, but unrestricted functions to be learned from the data, and the function $\mathbf{p}(.)$ must satisfy the probability conditions. Since functional networks do not make any assumptions about the function $g_k(\mathbf{x}_i, \mathbf{\Theta}_k)$, then it can be known or unknown in form and/or in the parameters. It could be linear or non-linear. However, since we approximate $g_k(\mathbf{x}_i, \mathbf{\Theta}_k)$ by a linearly independent family, which depends on parameters a_{si}, [8]. Figure 2 shows the architecture of the corresponding functional network model (1).
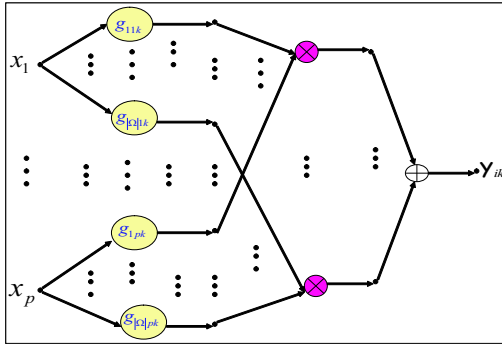

Figure 2. Unconstrained-functional networks classifier topology with p Predictors.

Follow both learning techniques and classification criterions in [8] and select the best network model according to the following criterions:

1. *Correct Classification Rate (CCR) and Average Squared Classification Error (ASCE)*:

$$CCR = \frac{\sum_{k=0}^{c-1} CC_k}{n}; \; ASCE = \frac{\sum_{k=0}^{c-1} [n_k - CC_k]^2}{n}, \quad (2)$$

where n_k is the number of observations in class k, and C_{kk} is the number of correctly classified observations in the class k. The best functional network is the one with both highest CCR and smallest ASCE. We construct the confusion matrix, which is a $c \times c$ matrix, its diagonal contains the number of correctly classified observations, CC_k, and the off-diagonal elements are the number of misclassified observations, mc_k, for $k = 0, \ldots, c - 1$.

2. *Computational cost* (Time of execution): It is the time needed to execute the classifier till obtaining the best model in both calibration and validation. The less computation cost is the better classifier.

3. *The Minimum Description Length* (MDL) criterion: As explained in [8], and then the best model is the one with the smallest MDL value. The form of the description length for the classification problem using the functional network is defined as

$$L(\mathbf{\Theta}_k) = \frac{m \, log(n \, k)}{2} + \frac{n \, k}{2} \, log \left(\frac{1}{n \, k} \sum_{i=1}^{n} \varepsilon_i^2(\mathbf{\Theta}_k) \right),$$

for all $k = 0, \ldots, c - 1$, where m and k are the number of elements in the family and the category levels, respectively. We note that the principle $L(\mathbf{\Theta}_k)$ is the code length of the estimated parameters $\mathbf{\Theta}_k, \; \forall k = 0, 1, 2, \ldots, c - 1$. We note that the description length has two terms:

(a) The first term $\frac{m \, log(n \, k)}{2}$ is a penalty for including too many parameters in the functional network model.

(b) The second term $\frac{n \, k}{2} log \left(\frac{1}{n \, k} \sum_{i=1}^{n} \varepsilon_i^2(\mathbf{\Theta}_k) \right)$ measures the quality of the functional network model fitted to the training set.

Therefore, the best model is the model with the smallest value of its description length. MDL is the best model performance.

5. Implementation and Comparative Study

We apply the constrained-functional networks explained in section 4 to Wisconsin Diagnostics Breast Cancer (WDBC) databases and compare its performances to the most common classification methods in both computer science and statistics literatures discussed in Section 1. The databases can be downloaded from the machine learning repository database at university of California, Irvine [1]. All the computations are implemented using MATLAB V6 under Pentium IV personal computer with a clock speed of 2.4 GHz. As in commonly done, we normalize the input variables to make sure that they are independent of measurement units. Thus, the predictors are normalized to interval of $(0, 1)$ using the formula:

$$x_i^{new} = \frac{x_i^{old} - x_{1:n}}{x_{n:n} - x_{1:n}}$$

where $x_{i:n}$ is the ith order statistics of $x_1, x_2, \ldots, x_n$. We use the *stratified Sampling* technique to make sure that we get the same proportion from each group as in the original data, as recommended by [12]. We randomly hold a total

of ($k = $ round $\left[\frac{n}{5}\right]$) or ($k = $ round $\left[\frac{n}{10}\right]$) observations, with $k_l = $ round $\left[k * \frac{n_l}{n}\right]$ observations from the class l, where n_l is the number of observations of the given data set in group l, for $l = 0, 1, \ldots, c - 1$. Therefore, to evaluate the performance of each classifier on a real-application, we use either *5-fold* or *10-fold* cross validation. Thus, we fold the given data into 5 or 10 parts, and we use 0.8 or 0.7 of the data for learning the classification model (building) and 0.2 or 0.3 for external validation (testing). Both description and relevant work on each data set under study are represented below:

1. **Wisconsin Diagnostics Breast Cancer** (WDBC): This database is created by William H. Wolberg at University of Wisconsin [18]. This database contains 569 observations among which 357 are *benign* cases and 212 are *malignant* cases. We note that in this database that for each observation, there are 30 featured variables. These features are computed from digital images of fine needle aspirates (FNA) of breast masses. These features describe the characteristics of the cell nuclei in the image. The author of this database considered 10 real-valued features for each cell nucleus: (i) *radius* (mean of distances from center to points on perimeter); (ii) *texture* (standard deviation of gray-scale values); (iii) *perimeter*; (iv) *area*; (v) *smoothness* (local variation in radius lengths); (vi) *compactness* ($\frac{perimeter^2}{area-1}$); (vii) *concavity* (severity of concave portions of the contour); (viii) *concave points* (number of concave portions of the contour); (ix) *symmetry*; and (x) *fractal dimension* (coastline approximation $- 1$). They computed the mean, standard error, and worst mean (the mean of the three largest values) of each feature. This process resulted in 30 feature variables for each image. For instance, field 1 is the *Mean Radius*, field 11 is the *Standard Error of Radius*, and field 21 is the *Worst Radius*.

2. **Wisconsin Breast Cancer** (WBC): The source of this database is the University of Wisconsin Hospital, Madison. There are 683 data points in this database of which 444 are *benign* and 239 are *malignant*. WBC is a nine-dimensional data set with the following features: (i) *Clump thickness*; (ii) *Uniformity of cell size*; (iii) *Uniformity of cell shape*; (iv) *Marginal adhesion*; (v) *Single epithelial cell size*; (vi) *Bare nucli*; (vii) *Bland chromatin*; (viii) *Normal nucleoli*; and (ix) *Mitoses*.

Several researchers studied WDBC database and proved that the best three attributes are *mean texture*, *worst mean area*, and *worst mean smoothness* [7]. Based on these three features they create separating hyperplane that uses multi-surface method-tree (MSM-T) to construct a decision tree. They reported that the best correct classification rate is 97.5%. This estimate was obtained using a repeated 10-fold cross-validation method, their CCR values are 93.5% and 95.9%, respectively.

We are utilizing both *internal* and *external* Validation Techniques as it is shown in [8]. We repeat the estimation and validation processes for $N = 1000$ times, then compute all the quality measures explained in Section 4 for all classifiers. Next, we summarize the results by computing the average, the standard deviation, and the coefficient of variation of each quality measure over these 1000 runs. In addition, we draw two graphs: One for the mean of CCR versus its standard deviation over the 1000 runs, and the other for the mean of the *ASCE* versus mean of *MDL*. These graphs help us to decide which classifier is better in its performance. In both plots, each classifier is represented by a symbol. In the graph of the average of CCR versus its standard deviation, a good classifier should appear in the upper left corner of the graph. In the graph of average of MDL versus average of ASCE, a good classifier should appear in the bottom left of the plot. In addition, corresponding to these graphs, we summarize the results in Tables. In these Tables, the highest CCR's are given in boldface [7].

For the sake of simplicity and space, we did the implementations for two predictors (feature variables and three predictor. We did the implementations and analysis with ten predictor variables (high-dimensional data) to check the performance of the functional network classifiers against other proposed classifiers.

1. **WDBC database with two predictor variables**: As we can see from the scatter plot of the data in Figure 3, there is some overlap between the two groups. The correlation coefficient between the two variables is 0.3295.
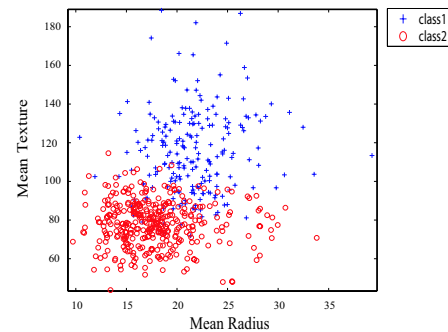

Figure 3. WDBC Data: The scatter plot of Mean Texture and Mean Radius

The quality measures and the corresponding scatter plots are shown in Table 1 and Figure 4, respectively.

Table 1. WDBC Data: The External Validation Results with 2 Predictors

Classification Method	No. Parameter		MDL		Time of Exec.		CCR		ASCE	
	mean	StDev	mean	StDev	mean	StDev	mean	StDev	mean	StDev
Logistic Regression	3.000	0.000	-674.541	35.332	0.275	0.069	0.889	0.017	1.044	0.317
Linear Disc Anal.	3.000	0.000	-759.904	41.170	0.002	0.009	0.884	0.017	0.981	0.335
KNN	3.000	0.000	-728.713	39.696	0.122	0.018	0.870	0.019	1.247	0.334
SVM	3.000	0.000	-561.948	39.586	0.375	0.038	0.759	0.033	5.148	1.316
PNN	2.000	0.000	-752.677	42.508	4.461	0.145	0.880	0.019	1.661	0.537
RBFN	2.000	0.000	-764.531	35.888	5.705	0.294	0.885	0.015	1.163	0.395
FFN	2.000	0.000	-640.306	177.133	34.893	11.790	0.773	0.171	10.147	15.113
FNBF-MLE	4.000	1.095	-826.403	34.387	2.728	0.614	**0.892**	0.013	0.981	0.238

Table 2. WDBC Data: The Internal Validation Results with 2 Predictors

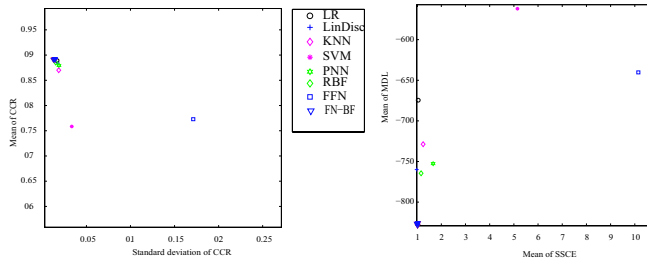| Classifier	COR-OBS1	COR-OBS1	MISS-OBS1	MISS-OBS1	CCR
LogReg	196.000	344.000	16.000	13.000	0.949
LinearDiscAnal	212.000	300.000	0.000	57.000	0.900
KNN	192.000	344.000	20.000	13.000	0.942
SVM	210.000	354.000	2.000	3.000	**0.991**
PNN	162.000	352.000	50.000	5.000	0.903
RBFN	201.000	350.000	11.000	7.000	0.968
FFN	209.000	354.000	3.000	3.000	**0.990**
FNBF-MLE	202.000	351.000	10.000	6.000	0.986

Figure 4. (a) Average of CCR versus σ_{CCR}, (b) Average of MDL versus Average of ASCE

Figure 5. WDBC Data: Scatter plot matrix

For the internal validation purpose, we summarize the output in Table 2.

2. **WDBC database with both three and ten predictor variables**: Figure 5 is a three-dimensional scatter plot of the data. As we can see, there is some overlap between the two groups. The pairwise correlation coefficients are: $r_{1,2} = 0.3435$, $r_{1,3} = 0.0775$, and $r_{2,3} = 0.2091$.

The reason for choosing the high-dimensional data is to check the performance of the functional network classifiers against other proposed classifiers, and check the measure of quality. The quality measures and the corresponding scatter plots are shown in Tables 1 and Figure 6, respectively.

From the above two tables 1, 3, and 2 with the corresponding figures of average of CCR versus its standard deviation, we observe, for example the following:

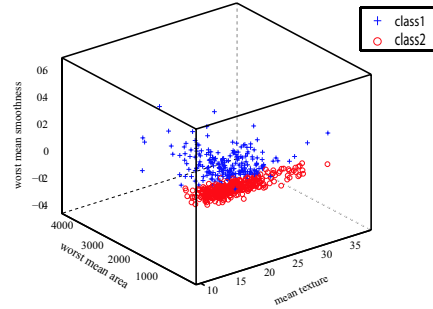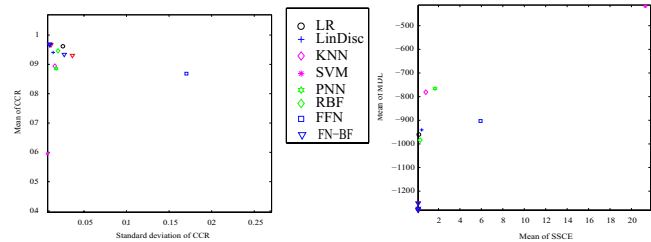1. The two classifiers: support vector machines, and

Figure 6. (a) Average of CCR versus σ_{CCR} (b) Average of MDL versus Average of ASCE

Table 3. WDBC Data: The External Validation Results with 3 Predictors

Classification Method	No. Parameter		MDL		Time of Exec.		CCR		ASCE	
	mean	StDev	mean	StDev	mean	StDev	mean	StDev	mean	StDev
Logistic Regression	4.000	0.000	-960.529	96.657	0.569	0.091	0.961	0.025	0.170	0.354
Linear Disc Anal.	4.000	0.000	-940.701	65.758	0.008	0.011	0.941	0.014	0.433	0.200
KNN	4.000	0.000	-781.178	41.149	0.085	0.014	0.894	0.016	0.817	0.245
SVM	4.000	0.000	-417.572	4.910	0.269	0.064	0.596	0.007	21.289	0.611
PNN	2.000	0.000	-766.211	40.919	1.612	0.679	0.886	0.017	1.679	0.562
RBFN	2.000	0.000	-983.910	99.239	2.367	0.048	0.946	0.020	0.296	0.229
FFN	2.000	0.000	-903.091	282.320	25.771	20.129	0.868	0.170	5.924	11.541
FNBF-MLE	6.000	1.403	-1250.802	542.581	16.223	3.032	**0.968**	0.010	0.104	0.075

Feedforward neural networks are the worst performance. Moreover, the performance of SVM becomes even worse with the high-dimensional predictors. The linear discriminant analysis classifier, and radial basis functions networks are the second worst average CCR.

2. Both SVM and neural networks classifiers work well for internal validation with small number of predictors. In the case of high-dimensional data, their performance is affected, and the CCR values becomes less. The Feedforward neural network has by far the highest execution time. Yet, it the has the second worst performance among all ten classifiers.

3. With the low-dimensional data, both Feedforward neural networks (FFNs) and support vector machines (SVMs) classifiers are the highest value of CCR. The functional networks classifier (MLE) is the second highest values of the average CCR in the low-dimensional data, but they are the highest in the high-dimensional data. Their performance is stable in both dimensions. The logistic regression, radial basis functions network and K-nearest neighbor classifiers perform more or less the same.

4. The functional networks classifier (MLE) is giving the highest values of the average CCR in the high dimensional data with less time of computations.

5. The functional networks classifier(MLE), logistic regression are giving the highest values of the average CCR. In addition, for the case of high-dimensional data, functional networks classifier (MLE) surpasses logistic regression.

6. All other classifiers perform more or less the same, with and logistic regression, linear discriminant analysis, K-nearest neighbor classifier probabilistic neural network, radial basis function, and functional networks classifiers (MLE).

7. The functional networks classifier(MLE) gives both smallest MDL and smallest ASCE. In addition, its execution time is much lower than its corresponding with both feedforward neural networks and support vector machines.

We draw our conclusion in section 6 by utilizing the useful information shown in both tables 1 and 3.

6. Conclusion and Future Work

We conclude that in the case of internal validation, the support vector machine and feeedforward neural networks are the worst performance. Both feeedforward neural networks and the constrained functional networks (FN-MLE) are the highest values of the average CCR. On the other hand, for the external validation, the results show that the linear discriminant analysis and probabilistic neural network classifiers give the lowest CCR value among all classifiers. The new functional networks classifier with $q = 2$ has the highest CCR values, both smallest MDL and smallest ASCE with execution time much lower than its corresponding with both feedforward neural networks and support vector machines.

We showed empirically that the proposed approach has better performance, high quality and generalization than common existing approaches, with much lower computational cost. However, more work is needed in evaluating the performance of the proposed method on other medical and other science or business databases.

Acknowledgments

References

[1] Bassett L. W, Liu T. H, Giuliano A. E., and Gold R. H., *The prevalence of carcinoma in palpable vs impalpable, mammographically detected lesions*. AJR; 157:2124, 1991.

[2] Bothorel S., Meunier B. B., and Muller S. A., *Fuzzy logic based approach for semilogical analysis of microcalcification in mammographic images*. International Journal of Intelligent System; 12(1112): 819848, 1997.

[3] Castillo E., Cobo A., Gómez-Nesterkín J., and E. Pruneda, *Introduction to Functional Networks With Applications, A Neural Based Paradigm*. Kluwer Academic Publishers:, New York, 1998.

[4] Castillo E., Cobo A., Gómez-Nesterkín R., and Hadi A. S., A general framework for functional networks. *Networks*, 35:70–82, 1999.

[5] J. S.-T. N. Cristianini, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, 2000.

[6] P. E. H. R. O. Duda and D. G. Stock, editors. *Pattern Classification, Second Edition*. John Wiley & sons, New york, 2001.

[7] E. El-Sebakhy, *Functional Networks as a New Framework for the Pattern Classifcation Problems*. Ph.D. Thesis, Cornell University, USA, 2004.

[8] E. El-Sebakhy, *The Maximum Likelihood Functional Networks as a Novel Approach for Pattern Classifcation Problems*. Journal of Neurocomputing. In Press, 2005.

[9] Furundzic D., Djordjevic, and Bekic A. J., *Neural Networks approach to early breast cancer detection*. Systems Architecture, 44:617-633, 1998.

[10] Gisvold J. J. and Martin J. K. J. r., *Prebiopsy localization of nonpalpable breast lesions*. AJR Am J Roentgenol; 143(3):477481, 1984.

[11] C. Hsu and C. Lin, A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, (13):415–425, 2002.

[12] D. R. Hush and B. G. Horne, "Progress in supervised neural networks: What's new since Lippmann?", IEEE Signal Processing Magazine", volume 1, pages=8-38, 1993.

[13] Johnson R. A. and Wichern D. W., Applied multivariate statistical analysis (5th ed.). *Upper Saddle River, NJ: Prentice-Hall*, 2002.

[14] Kovalerchuck B., Triantaphyllou E., Ruiz J. F., and Clayton J. *Fuzzy logic in computer-aided breast-cancer diagnosis: analysis of lobulation*. Artificial Intelligence in Medicine, 11, 7585, 1997.

[15] Pendharkar P. C., Rodger J. A., Yaverbaum G. J., Herman N., and Benner M., *Associations statistical, mathematical and neural approaches for mining breast cancer patterns*. Expert Systems with Applications, 17, 223232, 1999.

[16] Rubin M., Horiuchi K, and Joy N., *Use of fine needle aspiration for solid breast lesions is accurate and cost-effective*. Am J Surg; 174(6):694696, 1997.

[17] Stock D. G. and Yom-Tov E., *Computer Manual in MATLAB to Accompany Pattern Classification*. John Wiley & Sons, 2nd Edition, New York, 2004.

[18] Wolberg W. H. and Mangasarian O. L., *Multisurface method of pattern separation for medical diagnosis applied to breast cytology*, in: Proceedings of the USA National Academy of Sciences 87, 1990, pp. 91939196.

287