ORIGINAL ARTICLE

Local linear wavelet neural network for breast cancer recognition

M. R. Senapati · A. K. Mohanty · S. Dash · P. K. Dash

Abstract Breast cancer is the major cause of cancer deaths in women today and it is the most common type of cancer in women. Many sophisticated algorithm have been proposed for classifying breast cancer data. This paper presents some experiments for classifying breast cancer tumor and proposes the use local linear wavelet neural network for breast cancer recognition by training its parameters using Recursive least square (RLS) approach to improve its performance. The difference of the local linear wavelet network with conventional wavelet neural network (WNN) is that the connection weights between hidden layer and output layer of conventional WNN are replaced by a local linear model. The result quality has been estimated and compared with other experiments. Results on extracted breast cancer data from University of Wisconsin Hospital Madison show that the proposed approach is very robust, effective and gives better classification.

1 Introduction

Breast cancer has become a major cause of death among women in developed countries [1, 2]. Over one ten in Europe and one in eight women in United States may be affected by breast cancer during their life time [3].

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% [4–6].

Much effort has been devoted over the past decade to the development and improvement of pattern classification models for breast cancer detection [7–9]. Several researchers have used statistical and artificial intelligence to successfully "predict" breast cancer. Basically, the objective of these prediction techniques is to assign patients to either a "benign" group that does not have breast cancer or to a "malignant" group that has strong evidence of breast cancer.

Recently, local linear wavelet neural networks [10, 11] have been introduced as a very effective scheme for statistical pattern recognition problem and non-linear complex predictions.

In this paper, a local linear wavelet neural network (LLWNN) extends the application of [10, 11] and is proposed for breast cancer detection, in which the connection

M. R. Senapati (✉) · A. K. Mohanty
Department of Computer Science and Engineering, Gandhi Engineering College, Biju Patnaik University of Technology, Rourkela, Orissa 769007, India
e-mail: manas_senapati@sify.com

A. K. Mohanty
e-mail: asw_moh@yahoo.com

S. Dash
Balasore College of Engineering and Technology, Balasore, Orissa 756060, India
e-mail: sonali.isan@gmail.com

P. K. Dash
S 'O' A University, Bhubaneswar, Orissa 751030, India
e-mail: dashpk13@yahoo.com

weights between the hidden layer units and the output units are replaced by a local linear model, and the parameters of the network are updated using RLS. The usually used learning algorithm for WNN is gradient descent method. But its disadvantages are slow convergence speed and easy stay at local minimum. Because of this RLS with adaptive diversity learning is proposed for training the LLWNN. Simulation results for breast cancer Pattern Classification problem were compared with some other classification techniques, i.e., training RBFNN by RLS, RBFNN by Kalman filter, and RBFNN by Back propagation. The details about Kalman filter can be obtained from [12, 13]. The result thus derived shows the effectiveness of the proposed method. The main contributions of this paper are (1) the LLWNN providing a more parsimonious interpolation in high-dimension spaces when modeling samples are sparse; (2) a novel training algorithm for LLWNN was proposed. The paper is organized as follows. The LLWNN is introduced in Sect. 2. A RLS learning algorithm for training LLWNN is described in Sect. 3. A short discussion as well as experimental results obtained on Pattern Classification for Wisconsin Breast Cancer (WBC) problem is given in Sect. 4. Finally, concluding remarks are derived in the last section, i.e., Sect. 5.

2 Local linear wavelet neural network

In terms of wavelet transformation theory, wavelets in the following form:

$$\psi(x) = \prod_{i=1}^{n} \psi(x_i) \tag{1}$$

$$X = (x_1, x_2, \ldots, x_n),$$
$$X = (x_1, x_2, \ldots, x_n),$$
$$b_i = (b_{i1}, b_{i2}, \ldots, b_{in}),$$

are a family of functions generated from one single function $\psi(x)$ by the operation of dilation and translation. $\psi(x)$, which is localized in both the time space and the frequency space, is called a mother wavelet and the parameters a_i and b_i are named the scale and translation parameters, respectively. The x represents inputs to the WNN model.

In the standard form of WNN, the output of a WNN is given by

$$f(x) = \sum_{i=1}^{M} \omega_i \psi_i(x) = \sum_{i=1}^{M} \omega_i |a_i|^{-1/2} \psi\left(\frac{x - b_i}{a_i}\right) \tag{2}$$

where ψ_i is the wavelet activation function of ith unit of the hidden layer, and ω_i is the weight connecting the ith unit of the hidden layer to the output layer unit. Note that for the

n-dimensional input space, the multivariate wavelet basis function can be calculated by the tensor product of n single wavelet basis functions as follows

$$\psi(x) = \prod_{i=1}^{n} \psi(x_i) \tag{3}$$

Obviously, the localization of the ith units of the hidden layer is determined by the scale parameter a_i and the translation parameter b_i. According to the previous researches, the two parameters can either be predetermined based upon the wavelet transformation theory or be determined by a training algorithm. Note that the above WNN is a kind of basis function neural network in the sense of that the wavelets consists of the basis functions. Note that an intrinsic feature of the basis function networks is the localized activation of the hidden layer units, so that the connection weights associated with the units can be viewed as locally accurate piecewise constant models whose validity for a given input is indicated by the activation functions. Compared with the multilayer perceptron neural network, this local capacity provides some advantages such as the learning efficiency and the structure transparency. However, the problem of basis function networks is also led by it. Due to the crudeness of the local approximation, a large number of basis function units have to be employed to approximate a given system. A shortcoming of the WNN is that for higher dimensional problems, many hidden layer units are needed.

Local linear wavelet network in fact is a modification of WNN. The architecture of the proposed LLWNN is shown in Fig. 1. Its output in the output layer is given by

$$y = \sum_{i=1}^{M} (\omega_{i0} + \omega_{i1}x_1 + \cdots + \omega_{in}x_n)\psi_i(x)$$
$$= \sum_{i=1}^{M} (\omega_{i0} + \omega_{i1}x_1 + \cdots + \omega_{in}x_n)|a_i|^{-1/2}\psi\left(\frac{x - b_i}{a_i}\right) \tag{4}$$

where $X = [x_1, x_2, \ldots, x_n]$ instead of the straight forward weight ω_i (piecewise constant model), a linear model

$$v_i = \omega_{i0} + \omega_{i1}x_1 + \cdots + \omega_{in}x_n \tag{5}$$

is introduced. The activities of the linear models v_i $(I = 1,2,\ldots, M)$ are determined by the associated locally active wavelet functions $\psi_i(x)$ $(I = 1,2,\ldots, M)$ thus, v_i is only locally significant. The motivations for introducing the local linear models into a WNN are as follows: (1) local linear models have been studied in some neuro fuzzy systems and shown good performances [14, 15] and (2) local linear models should provide a more parsimonious interpolation in high-dimension spaces when modeling samples are sparse. The scale and translation parameters and local linear model parameters are randomly initialized at the beginning and are optimized by recursive least square algorithm discussed in the following section.

Fig. 1 A local linear wavelet neural network
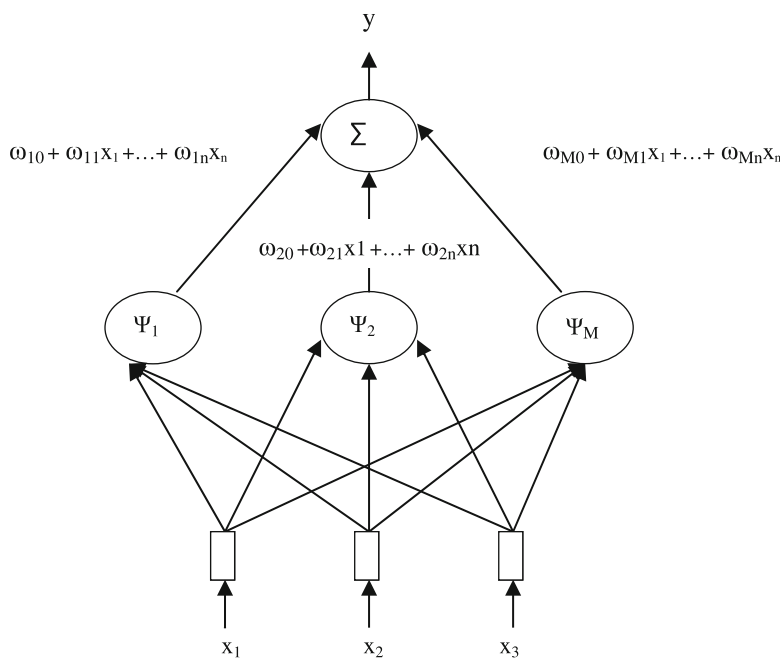
3 Recursive least square

The recursive least square (RLS) is a parameter identification technique. In RLS algorithm, there are two variables involved in the recursions (those with time index $n - 1$): $\hat{w}(i - 1)$, $P(i - 1)$. We must provide initial values for these variables in order to start the recursions:. $w(0)$

If we have some apriori information about the parameters $\hat{w}$, this information will be used to initialize the algorithm. Otherwise, the typical initialization is $w(0) = 0$. $P(0)$

$$P(i) = \left[\sum_{n=1}^{i} \lambda^{i-1} \psi(n)\psi(n)^T \right]^{-1}$$

the exact initialization of the recursions uses a small initial segment of the data $\psi(i_1); \psi(i_1 + 1)\ldots, \psi(0)$ to compute

$$P(0) = \left[\sum_{n=1}^{0} \lambda^{-1} \psi(n)\psi(n)^T \right]^{-1}$$

All the necessary equations to form the RLS algorithm are

$$k(i) = P(i-1)\varphi T(i)/\lambda + P(i-1)\varphi T(i) \tag{6}$$

$$w(j) = w_j(i-1) + k(i)\left[d_j(i) - w_j(i-1)\varphi T(i)\right] \tag{7}$$

$$P(i) = 1/\lambda[P(i-1) - k(i)\varphi(i)P(i-1)] \tag{8}$$

where λ is real number between 0 and 1, $P(0) = a - 1\, I$, and a is a small positive number and $w_j(0) = 0$.

4 Implementation and comparative study

We apply the local linear wavelet neural network explained in Sect. 2 to Wisconsin Breast Cancer (WBC) databases and compare its performances to the most common classification methods in both computer science and statistics literatures. The databases can be downloaded from University of Wisconsin Hospital, Madison[1]

All the computations are implemented using MATLAB V6.5[2] under Pentium IV personal computer with a clock speed of 2.4 GHz and the equations were written using Math Type or Microsoft Equation Editor. 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:

$$x_i^{new} = \frac{x_i^{old} - x_{1:n}}{x_{n:n} - x_{1:n}} \tag{9}$$

where $x_{1:n}$ 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 in the original data. We randomly hold a total of ($k = $ round $\frac{n}{5}$) or ($k = $ round $\frac{n}{10}$) observations, with $k_1 = $ round $\frac{n}{10}$ observations from the classl, where nl 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). The new technique LLWNN-RLS is compared with a wide range of classifiers to evaluate its performance with respect to correct classification rate and the time it takes to get trained. As defined in [16, 17], the best network model was selected based on the following criterions.

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

$$CCR = \frac{\sum_{k=0}^{c-1} CC_k}{n}; \quad ASCE = \frac{\sum_{k=0}^{c-1} [n_k - cc_k]^2}{n} \tag{10}$$

where n_k is the number of observations in class k, and CCR 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, CCR, and the off-diagonal elements are the number of misclassified observations, mck, for $k = 0, c - 1$.

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 [16, 17], 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

$$L(Q_k) = \frac{m\log(nk)}{2} + \frac{nk}{2}\log\left(\frac{1}{nk}\sum_{i=1}^{n} e_i^2(Q_k)\right) \tag{11}$$

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 $L(Q_k)$ is the code length of the estimated parameters Θ_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 $\frac{n_k}{2}\log\left(\frac{1}{n_k}\sum_{i=1}^{n} \varepsilon_i^2(\Theta_k)\right)$ 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. Both description and relevant work on the data set under study are represented below:

4.1 Wisconsin breast cancer (WBC):

The data set were obtained from University of Wisconsin Hospital; Madison WBC is a nine-dimensional data set with the following features:

(1) Clump thickness; (2) Uniformity of cell size;(3) Uniformity of cell shape; (4) Marginal adhesion;(5) Single epithelial cell size; (6) Bare nuclei; (7) Bland chromatin; (8) Normal nucleoli; and (9) Mitoses. For our classification purpose, 400 exemplars were used for training and the 299 exemplars for testing from a total of 699 exemplars. Several researchers studied WBC database and proved that the best three attributes are mean texture, worst mean area and worst mean smoothness [18].

We are utilizing external validation techniques as it is shown in [16, 17]. We repeat the estimation and validation processes for $N = 1,000$ times, then compute all the quality measures explained in Sect. 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 1,000 runs.

In addition, we draw two graphs to reach the conclusion: one for the mean of CCR versus its standard deviation 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 mean of CCR versus its standard deviation, a good classifier should appear in the lower right corner of the graph. In the graph of mean ASCE verses mean MDL, 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.

For the sake of simplicity and space, we did the implementations for two predictors and three predictors to check the performance of the LLWNN-RLS classifier against other classifiers.

4.2 Result and discussion

From Tables 1 & 2 and Figs. 2 & 3 of average of CCR versus its standard deviation, we observe, for example, the following:

1. The two classifiers: RBFNN and RBFNN-Kalman filter neural networks are the worst performance.

Table 1 WBC data: the external validation results with 2 predictors

Classification method	No. parameter		MDL		Time in sec		CCR		ASCE	
	Mean	St. Dev	Mean	St. Dev	Mean	St. Dev	Mean	St. Dev	Mean	St. Dev
RBFNN_BP	2.000	0.000	−764.531	35.888	5.705	0.294	0.885	0.015	1.163	0.395
RBFNN-Kalman	2.000	0.000	−786.421	35.456	2.345	0.021	0.889	0.041	1.012	0.412
RBFNN-RLS	4.000	1.095	−826.403	34.387	2.728	0.614	0.892	0.013	0.981	0.238
LLWNN-RLS	5.000	2.115	−888.275	32.886	1.652	0.352	**0.897**	0.012	0.881	0.187

Table 2 WBC data: the external validation results with 3 predictors

Classification method	No. parameter		MDL		Time in sec		CCR		ASCE	
	Mean	St. dev	Mean	St. dev	Mean	St. dev	Mean	St. dev	Mean	St. dev
RBFNN-BP	2.000	0.000	−938.910	99.239	2.367	0.048	0.946	0.020	0.296	0.229
RBFNN-Kalman	2.000	0.000	−942.850	123.359	1.459	0.038	0.964	0.030	0.345	0.312
RBFNN-RLS	6.000	1.403	−1250.802	542.58	16.22	3.032	0.968	0.010	0.104	0.075
LLWNN-RLS	7.000	2.325	−1353.776	656.82	2.31	0.655	**0.972**	0.010	0.010	0.009

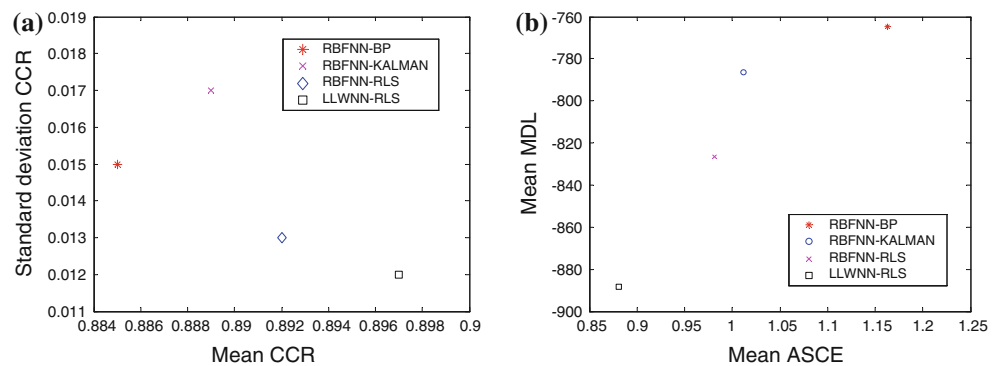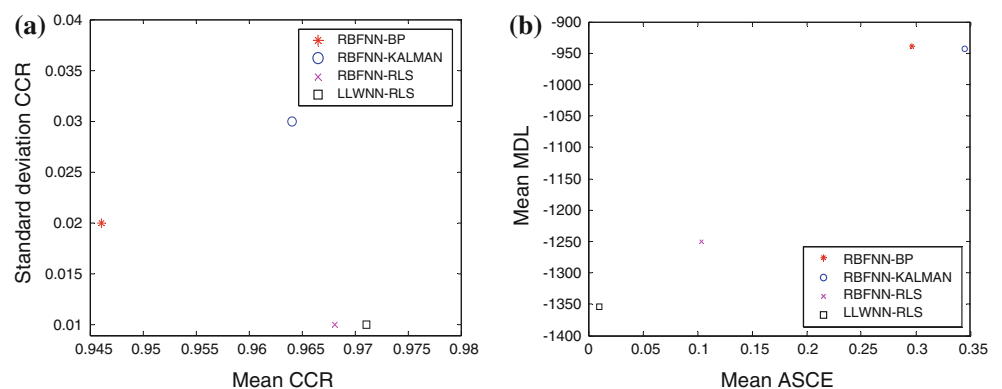**Fig. 2 a** Mean CCR verses std CCR for 2 predictors **b** Mean ASCE verses mean MDL for two predictors

Fig. 3 a Mean CCR verses std CCR for 3 predictors **b** Mean ASCE verses MDL for 3 predictors

2. The LLWNN-RLS is giving the highest values of the average CCR in the high dimensional data with less time of computations.

3. The LLWNN-RLS, RBFNN-RLS are giving the highest values of the average CCR.

4. The LLWNN-RLS gives both smallest MDL and smallest ASCE. In addition, its execution time is much lower than other classifiers.

We draw the comparative study by utilizing the information provided in Table 3 and conclusion in

Table 3 Classification accuracy on the Wisconsin Breast Cancer (WBC) dataset

Algorithm	Result in percentage	Source
LLWNN-RLS (proposed)	97.2	Proposed
C4.5	94.7	[21]
LVQ	96.6	[21]
FDA	96.7	[21]
SSV	96.3	[21]
CART	93.5	[21]
LDA	96.0	[21]
SMLP	97.1	[20]
FNBF-MLE	96.8	[17]
HMLP (Elimination method)	96.0	[19]
MLP + BP	96.7	[21]
SVM lin, opt C	96.7	[21]
Rough Set Theory	85	[22]

Sect. 5 by utilizing the useful information shown in Tables 1 and 2.

Rule for classification of WBC data sets using LLWNN and RLS

$$\text{if}(oo(r, 1) \geq 0.06 \ \& \ oo(r, 1) \leq 0.4627) \ \& \ (oo(r, 2) \geq 0.0611 \ \& \ oo(r, 2) \leq 0.41) \quad \text{then Benign};$$
$$\text{if}(oo(r, 1) \geq -3.7718 \ \& \ oo(r, 1) \leq 0.0394) \ \& \ (oo(r, 2) \geq -3.2997 \ \& \ oo(r, 2) \leq 0.036) \quad \text{then Malignant};$$

4.3 Comparative study with existing methods

The proposed technique is compared with some of the existing techniques [17, 19–21]. The comparison is depicted in Table 3. The result of the comparison shows that the proposed technique gives better classification as compared to some of the existing techniques.

5 Conclusion

Even though mammography is one of the best techniques for breast tumor detection, but in some cases, despite their experience radiologists cannot detect tumors. There computer-aided methods like the one presented in this paper could assist medical staff and increase the accuracy of detection. Statistics show that only 20–30% of breast tumor cases are cancerous. In a false negative detection, if an actual tumor remains undetected that could lead to higher costs or even to the cost of a human life. Here is the trade off that motivated us to develop a classification system

In this paper, we presented a technique for breast tumor classification. The objective of this study is to examine the effectiveness of LLWNN for classifying breast cancer data. The technique was compared with different methods already developed. We showed empirically that the proposed approach has better performance, high quality and generalization than common existing approaches. As it is known that data mining techniques are more suitable to larger databases, we intend to use a larger data base, from medical science and/or business sector to evaluate the performance of the proposed technique. Also the technique needs to be evaluated using time series data to validate the findings.

References

1. Cancer net home page of National Cancer Institute, (2002), http://biomed.nus.sg
2. Jemal A, Murray T, Ward E, Samuels A, Tiwary RC, Ghafoor A, Feuer EJ, Thun MJ (2005) Cancer statistics 2005, CA. Cancer J Clin 55:10–30
3. Bothorel S, Meunier BB, Muller SA (1997) Fuzzy logic based approach for semilogical analysis of microcalification in mammographic images. Int J Intell Syst 12:819–848
4. Bassett LW, Liu TH, Giuliano AE, Gold RH (1991) The prevalence of carcinoma in palpable versus impalpable, mammographically detected lesions. AJR 157:21–24
5. Gisvold JJ, Martin JK Jr (1984) Pre biopsy localization of non palpable breast lesions. AJR Am J Roentgenol 143(3):477–481
6. Rubin M, Horiuchi K, Joy N (1997) Use of fine needle aspiration for solid breast lesions is accurate and cost effective. Am J Surg 174(6):694–696

7. Downs J, Harrison RF, Cross SS (1998) A decision support tool for the diagnosis of breast cancer based upon Fuzzy ARTMAP. Neural Comput Appl 7(2):147–165

8. Cheng HD, Shan J, Ju W, Guo Y, Jhang L (2010) Automated breast cancer detection and classification using ultra sound images, A survey. Pattern Recogn 43:299–317

9. Huang Y-L, Wang K-L, Chen D-R (2006) Diagnosis of breast tumors with ultrasonic texture analysis using support vector machines. Neural Comput Appl 15(2):164–169

10. Chen Y, Yang B, Dong J (2006) Time series prediction using a local linear wavelet neural network. Neuro Comput 69:449–465

11. Chen Y, Dong J, Yang B, Zhang Y (2004) local linear wavelet neural network. Fifth world congress on intelligent control and automation (WCIA), Hangzhou, pp. 1954–1957

12. de Jesus Rubio J, Pacheco J (2009) An stable online clustering fuzzy neural network for nonlinear system identification. Neural Comput Appl 18(6):633–641

13. An Introduction to Kalman Filter by Welch and Bishop, Welch@cs.unc.edu, Available: http://ww.cs.unc.edu/~welch

14. Fischer B, Nelles O, Isermann R (1998) Adaptive predictive control of a heat exchanger based on a fuzzy model. Control Eng Pract 6:259–269

15. Foss B, Johansen TA (1993) On local and fuzzy modeling. In: Proceedings of 3rd international industrial fuzzy control and intelligent systems. IEEE, pp. 80–87. doi: 10.1109/IFIS.1993.324209

16. Cheng HD, Shan J, Ju W, Guo Y, Jhang L (2010) Automated breast cancer detection and classification using ultra sound images, a survey. Pattern Recogn 43:299–317

17. El-Sebakhy EA, Faisal KA, Helmy T, Azzedin F, Al-Suhaim A (2006) Evaluation of breast cancer tumor classification with unconstrained functional networks classifier. IEEE international conference on computer systems and applications, pp. 281–287. doi:10.1109/AICCSA.2006.205102

18. El-Sebakhy E (2004) Functional networks as a new frame work for the pattern classification problems, PhD Thesis, Cornell University, USA

19. Mat A, Harsa S, Salleh NM, Othman NH (2009) Neural network inputs selection for breast cancer cells classification. KES-IDT'09, studies in computational intelligence, Springer 199: 1–11, doi:10.1007/978-3-642-00909-9

20. Kordos M (2005) Search-based algorithms for multilayer perceptrons. PhD Thesis, The Silesian University of Technology, Gliwice, available at http://www.fizyka.umk.pl/kordos/pdf/mKordos-PhD.pdf

21. Ster B, Dobnikar A (1996) Neural network in medical diagnosis: comparison with other methods. EANN'96, pp. 427–430

22. Hassanien AE, Ali JMH. Feature extraction and rule classification algorithm of digital mammography based on rough set theory. Available at www.wseas.us/e-library/conferences/digest2003/papers/463-104.pdf