PERFORMANCE EVALUATION OF DIVERSIFIED SVM KERNEL FUNCTIONS FOR BREAST TUMOR EARLY PROGNOSIS

Khondker Jahid Reza[1], Sabira Khatun[1], Mohd F. Jamlos[2], Md. Moslemuddin Fakir[3] and Sheikh Shanawaz Mostafa[4]
[1]School of Computer and Communication Engineering, Universiti Malaysia Perlis, Pauh Putra Main Campus, Arau, Perlis, Malaysia
[2]Advanced Communication Engineering Centre (ACE), School of Computer and Communication Engineering, Universiti Malaysia Perlis, Pauh Putra Main Campus, Arau, Perlis, Malaysia
[3]Institute of Engineering Mathematics, Universiti Malaysia Perlis (UniMAP), Pauh Putra Main Campus, Arau, Perlis, Malaysia
[4]Department of Biomedical Engineering, Khulna University of Engineering and Technology, Khulna, Bangladesh
E-Mail: jahid_rifat@yahoo.com

ABSTRACT

Ultra wide-band (UWB) microwave technology is a promising candidate to detect the early breast cancer. This paper aims to depict pattern recognition performance of support vector machine (SVM) for confocal UWB breast tumor imaging dataset. A novel feature extraction technique is also introduced in this paper for the signal classification perfectly and promptly. SVM classifier functions the comparative study between SVM kernel functions includes linear function, radial basis function, polynomial and multi layer perceptions are investigated and verified for pattern recognition performance with the help of receiver operating characteristic (ROC) graph and confusion matrix. The main motto of this paper is to identify the tumor in its smallest dimension from available works including their data using the proposed feature extraction. In total, thirteen different sizes of benign tumors are being considered where the smallest and largest tumor sizes utilized are 1mm and 9 mm, respectively.

Keywords: benign tumor, breast cancer, support vector machine, ultra wide-band.

INTRODUCTION

Breast cancer is indisputably considered as the second deadly cancer among women [1]. Usually, breast cancer causes due to unusual growth of either lobules (milk producing tissue) or ducts (connecting tissue between lobule and nipple). Most breast tumor tissue is basically two types: benign and malignant. Benign tissue is not the cancerous; it may remains in the body life-long with the negligible growth or by the period of time it may turns to deadly cancer malignant tumor [2, 3]. In a recent survey from American cancer society, it is found that 89% of women could survive if the cancer is detected within 5 years. This rate is lower as 82% and 77% if it is detected after 10 and 15 years, respectively [4]. Joy *et al*., [5] and Tabar *et al*., [6] showed in their researches that, early detection of tumor is important for the long term survival in the way of reducing the risk.

At present, Microwave UWB imaging is one of the most promising techniques in the arena of breast cancer detection. Many researchers have utilized this technology to detect breast cancer in early stage [7-11]. Very high detection capability and low power consumption encourage the researchers to utilize this technique.

At present pattern recognition methods are useful term in case of signal processing. A good number of techniques are also available. Artificial Neural Network (ANN) is already well justified for breast cancer pattern recognition and detection [11-12] that literally showed tremendous successful detection and size identifier capability with 95.8% accuracy. Some crucial factors i.e., training time, feature extraction etc., which influence the system performance was not well considered. A huge feature value also increased the ANN training duration and reduces the training performance. Particle Swam Algorithm can also be a better option [13]. Machine learning based artificial intelligence is employed in fine needle aspirate of breast for early breast cancer detection. Support Vector Machine (SVM) is already utilized to identify tumor either from mammographic image [14] or mammogram data [15]. A revolutionary technology of gene signature for breast cancer detection is described in [16]. This study gives an idea that SVM can classify the breast cancer with 34% better accuracy than conventional 70-gene signature. The basic idea of signal processing using SVM for breast cancer detection was described in [16]. Several studies conducted using Wisconsin breast cancer database of Machine Learning Repository from the University of California. The data set are mainly demonstrated to differentiate between malignant and benign tumors including 97.51% classification accuracy, 97.49% sensitivity and 97.52% specificity are achieved [17].

An improvement about 100% of accuracy using radial basis function (rbf) and about 83.55% utilizing the sigmoid kernel function is achievable and described in [18]. Another study shows that, among SVM kernel functions, polynomial kernel functions provide higher accuracy of detection about 97.5% [19]. But SVM become overwhelmed due to large amount of input data.

This paper is organized as follows. The next section presents the system model elaboration including data acquisition, signal pre-processing; followed by feature extraction. In section III, SVM Kernel Functions are defined breast phantom preparation and device specification. After that, signal processing tool, followed by System performance then results, and finally the conclusion.

VOL. 9, NO. 3, MARCH 2014 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

METHODOLOGY

The graphical presentation of the proposed system model is shown in the Figure-1.

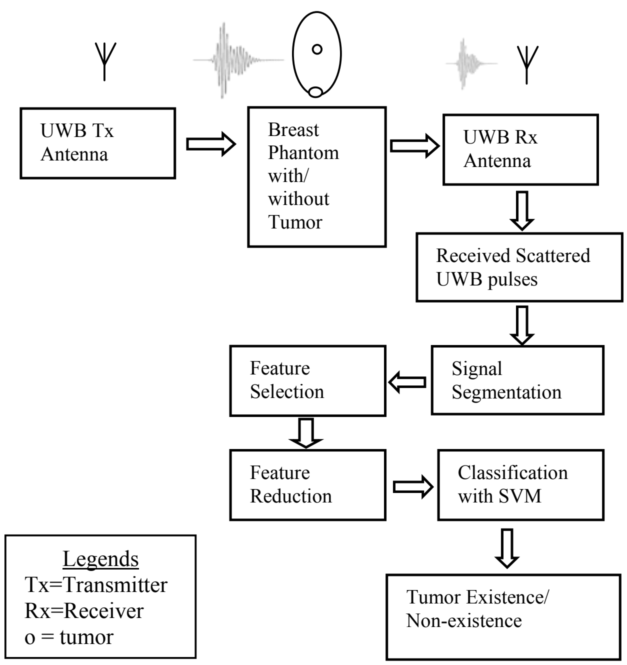


Figure-1. Proposed system model data acquisition.

Data acquisition

The UWB transceiver from Time domain Co. [20] is used for experiment with resonant frequency 4.7GHz and forward scattered pulses were received at the receiver end for signal processing. The descriptions of dataset and breast phantom are similar to Saleh *et al*., [21]. Used tumor sizes for the experiment are 1mm, 1.6 mm, 2 mm, 2.7 mm, 3 mm, 3.8 mm, 4 mm, 4.2 mm, 5 mm, 5.3 mm, 6 mm, 6.5 mm, 7 mm and in size. For each of these tumor sizes, 22 times UWB signals are transmitted through the breast phantom. On the other hand, 36 times received UWB signals were examined without any tumor inside the breast phantom. Each of the above cases, in average of 4500 data points was generated. Among these huge data points, 67% and 33% data sets kept for training and testing purposes, respectively.

Signal pre-processing

In general, each pattern recognition methodology for signal processing is almost the same which is shown in Figure-1. The UWB received signals go through the segmentation processing. Numerous analog to digital conversion techniques are available, Fast Fourier Transform (FFT), Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT) etc. are well known terms. For our research, we have used DCT, so that after receiving UWB pulses from receiver the analog UWB pulses are divided into discrete values. None of the de-noising method was applied in this context.

Feature extraction

Processed segmented data are ready for feature extraction, where large amount of data points are reduced to some characteristic features. Tremendous feature reduction techniques are available; in our previous study Saleh *et al*., [21] used Principle Feature Analysis (PFA) extraction where 50-300 larger DCT values were considered. So far it is still time consuming for computing complexity of a huge set of data. Another, feature extraction method described in [22] where, features are selected from EEG features for neonatal seizure detection by acquiring the RMS amplitude, line length, and number of maxima and minima values. So in this paper, a new feature is introduced in which, we have taken only four feature values i.e., maxima, minima, mean and standard deviation values from the data set. It enhanced the processing speed and reduced the computation complexity. The general equation used for mean or average, μ and standard deviation, σ are given in equation (1) and equation (2), respectively.

$$\mu = \frac{1}{N}\sum_{n=0}^{N-1}[X_n] \tag{1}$$

$$\sigma = \frac{1}{N}\sum_{n=0}^{N-1}[X_n - \mu]^2 \tag{2}$$

In these equations, N is the total number and X_n is the number of data points.

SVM kernel functions

SVM is a supervised learning algorithm proposed by Vapnik in 1992 [23, 24]. Predominantly, this machine learning algorithm is prospered for the solution of classification problems. With the discovery of e-insensitive loss function [18], SVM is currently extended to solve non-linear regression problems.

The basic concept of the SVM can be described as follows, considering (x_i, y_i) is a training set of data as if, i = 1, 2,..., n with $x_i \in R^n$ and $y_i \in \{-1,1\}^{'}$, . Then the equation of a decision surface in the form of a hyperplane can be defined as, $w^T x + b = 0$ (3)

where, *w* stands for adjustable vector and *b* is for bias i.e., The condition then need to be satisfied is,

(i) $w^T x_i + b \geq 0$ for $y_i = +1$ (4)

(ii) $w^T x + b < 0$ for $y_i = -1$ (5)

To find out the optimized value of *w* and *b* the condition (ii) needs to be satisfied. Then the *w* minimizes the cost function:

$$\phi(w) = \frac{1}{2} w^T w$$

Kernel functions are introduced in SVM for solving non-linear functions. In this work, five kernels are verified, as stated above, to check the classification performance. The kernel functions and their respected equations are given below:

Kernel function: $k(x_i, x_{j)} = \phi(x_i)^T . \phi(x_i)$

Linear kernel: $k(x_i, x_j) = x_i^T . x_i$

Polynomial order kernel: $k(x_i, x_j) = (\gamma x_i^T x_i + r)^n, \gamma > 0$

Radial basis function (RBF):

$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0.$

Multi layer perception or sigmoid kernel:

$k(x_i, x_j) = \tan\mathrm{h}(\gamma x_i^T x_i + r)$

Here, γ, r and n are kernel parameters [25].

2nd order Polynomial kernel is known as quadratic kernel function and in use in the paper to differentiate from 3rd order polynomial kernel.

RESULTS AND DISCUSSIONS

A basic two-by-two contingency table is represented in Table-1. A set {*P, N*} of positive and negative label class is mapped which is known as *True class*. Another set {*Y, N*} or *Hypothesized Class* is used to separate the actual class and the predicted class [26]. So that, four possible outcomes are found from Table-1. If provided instance is positive and it is classified as positive, then it is counted as *True Positive*, marked as light green color. However, if it is classified as Negative then is called as *False Negative* and marked as light orange colored cell. On the other hand, if the instance is negative and it is classified as negative then it is called as *True Negative* and marked as light green cell. Alternatively, if the instance is positive then it is counted as *False Positive* and marked as orange cell. Sensitivity and specificity are defined to clarify the system performance.

Table-1. Confusion matrix. [26]

		True class		
		P	***N***	
Hypothesized Class	***P***	*True Positive*	*False Positive*	*Sensitivity*
	N	*False Negative*	*True Negative*	*Specificity*
	Total	Positive	Negative	*Accuracy*

$$true\ positve\ rate \approx \frac{Positives\ correctly\ classified}{Total\ positives} \quad (6)$$

$$false\ positve\ rate \approx \frac{Negatives\ incorrectly\ classified}{Total\ negatives} \quad (7)$$

Both can be defined as,

$$Sensitivity = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive(FP)} \quad (8)$$

$$Specificity = \frac{TrueNegative\ (TN)}{True\ Negative\ (TN) + False\ Positive\ (FP)} \quad (9)$$

$$Accuracy = \frac{TruePositi\ ve\ (TP) + True\ Negative\ (TN)}{Postive + Negative} \quad (10)$$

It is stated earlier that, 12 different sizes of tumor are used and tumor sizes are already defined in section II (A). We have trained and tested the SVM without tumor condition and with tumor condition. However, UWB pulses are passed through the breast phantom 22 times for each size of tumor. These propagation iterations are increased to 36 in case of without tumor condition. In the first case, 8 different tumor sizes (i.e., 0.16mm, 0.2mm, 0.27mm, 0.3mm, 0.38mm, 0.4mm, 0.42mm, 0.53mm) data sets are kept for training purposes. So in total 200 times (8×22times=176 and 24 times without tumor condition) of data set are used to train the SVM and kept the rest of the tumor sizes data set for the testing purpose. The number of iteration for testing purpose is 100 times (4×22=88 and 12 times without tumor condition). Then the proposed feature extraction i.e., four characteristic features is applied on the raw data set. Among these 200 cases, 24 cases out of 200 cases are found true negative and rest of the 176 cases are found true positive. The result is perfectly matched with the given data set. So the training accuracy reaches up to 100%.

VOL. 9, NO. 3, MARCH 2014 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

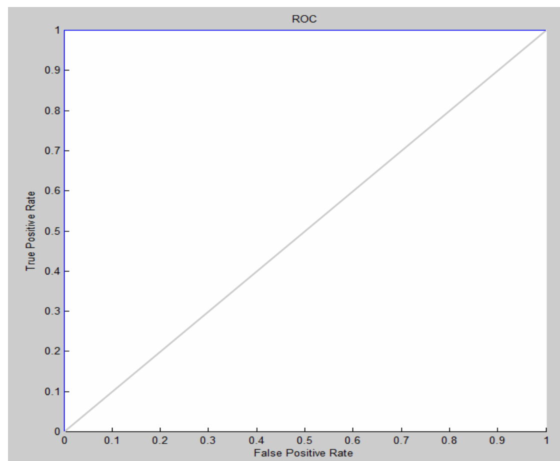
Table-2. Confusion matrix of linear, quadratic, polynomial order 3 and Rbf kernel functions.

		Training			**Testing**		
Output Class	*P*	176	0	100%	88	0	100%
		88%	0%	0%	88%	0%	0%
	N	0	24	100%	0	12	100%
		0%	12%	0%	0%	12%	0%
	Total	100%	100%	100%	100%	100%	100%
		0%	0%	0%	0%	0%	0%
		P	*N*		*P*	*N*	
		Input class			Input class		

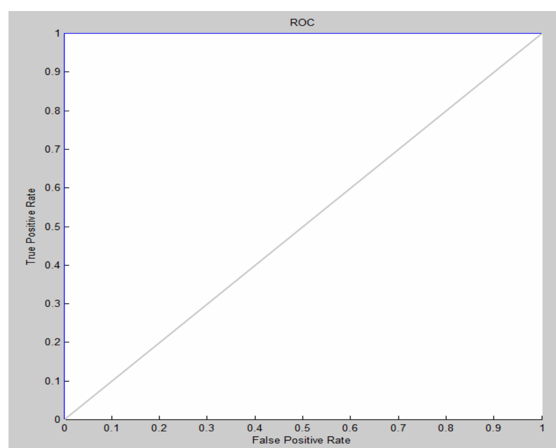
On the other hand, in testing phase, 100 sets of data are feed into the system where 88 are recognized as true positive. There are another 12 sets of data without tumor condition and it is found as true negative. So the accuracy for testing phase also rises up to 100%. The corresponding confusion matrix is represented in Table-2 for linear, quadratic, polynomial order 3 and rbf kernel functions. Two columns are being used to separate the training and testing outcomes. For the four kernel functions, training and testing data set are separately feed in to the SVM.

A receiver operating characteristic (ROC) graph is used to visualize, organize and selecting the classifiers based on their performance. It is basically used to tradeoff between true positive rates (tp rate) and false positive rate (fp rate) of classifiers which are defined in equation x and y. Relevant ROC graphs for training and testing are shown Figures 2 (a) and (b).

In Figure-2, the horizontal axis is marked as *False Positive Rate* and vertical axis is marked as *True Positive rate*. The blue coloured line is ROC curve and grey coloured line is a reference curve. It is visible from both the Figures 2(a) and (b), ROC curve has a higher left point (0, 1) both in training and testing which represents such a classifier commits no false positive errors as well as no negative errors which means the perfect classification [26]. It can be defined in another way, the point where the ROC curve goes flat thence *false positive rate* is 0 (zero). It mentions the 100% pattern recognition accuracy is achieved.



(a)



(b)

Figure-2. (a) Training and (b) Testing ROC graphs of linear, quadratic, polynomial order 3 and Rbf kernel functions.

The worst performance is noticed for multilayer perception or sigmoid kernel function among these five kernel functions. Sensitivity and specificity percentage for this type of kernel function is indexed in Table-3. In total

VOL. 9, NO. 3, MARCH 2014 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

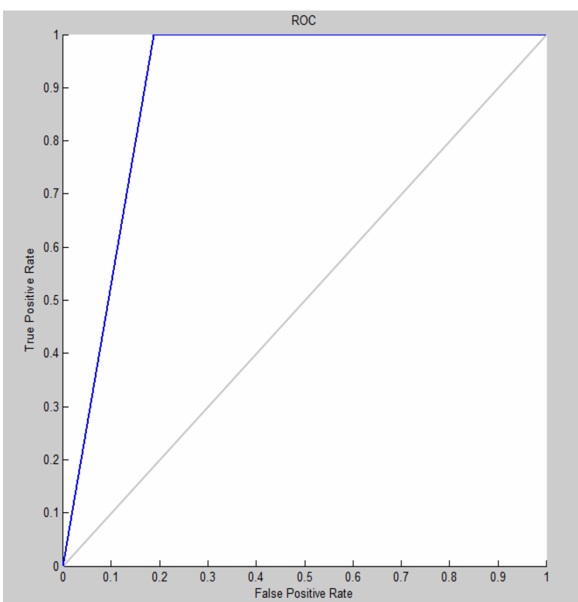
www.arpnjournals.com

of 200 training cases, *True Negative* is found in 24 cases; but *False Negative* is about 16.5% that reduces the performance. On the other hand, 143 out of 200 cases are successfully identified the existence of a tumor. This is equivalent to 71.5% success as *True Positive* without any *False Positive*. That's why, the total training accuracy reduces to 83.5% or incorrect classification performance is found 16.5%. However, in testing phase, *True Negative* and *True Positive* success rate are recorded as 12% and 42%, respectively.

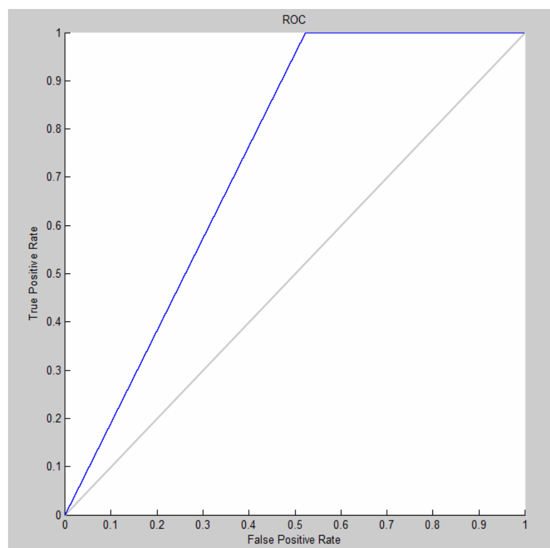
Table-3. Confusion matrix of multi layer perception kernel functions.

		Training			Testing		
Output Class	*P*	143	33	42.1%	42	46	20.7%
		71.5%	16.5%	57.9%	42%	46%	79.3%
	N	0	24	100%	0	12	100%
		0%	12%	0%	0%	12%	0%
	Total	100%	81.3%	83.5%	100%	47.7%	54%
		0%	18.8%	16.5%	0%	52.3%	46%
		P	*N*		*P*	*N*	
		Input class			Input class		

However, the *False Negative* percentage of testing performance is 46%. So, the total accuracy of testing is 54% and incorrect classification percentage is reached to 46%. The ROC graphs for training and testing phase are presented in Figures 3(a) and (b), respectively. It is also matched with Table-3. In Figure-3(a) point (0.188, 1) indicate that 18.8% of positive data are classified as negative. And this false positive rate increase to 52.3% in testing phase cleared from (0.523, 1) point in Figure-3(b). In both cases sensitivity was 1 [26].



(a)



(b)

Figure-3. (a) Training and (b) Testing ROC graphs of multi layer perception kernel functions.

CONCLUSIONS

SVM based pattern recognition technique for early breast tumor detection is investigated in this paper. Comparative study among different kernel functions exhibit that, Linear, Quadratic, Polynomial order 3 and Radial Basis kernel functions can classify the tumor availability with 100% accuracy both in training and testing phases. Tumor sizes are varied from the smallest size of 1mm up to 7mm. Unlike other kernel functions, MLP kernel functions provide less accuracy about 83.5% in training and 54% in testing result. It can be concluded that, SVM is well capable of classifying the tumor availability using the proposed feature extraction and is recommended to solve the regression problem for further investigation. Also, it is kept for the future research to identify the more than 2 patterns from the set of data.

ACKNOWLEDGEMENT

This work is supported by the Long-term Research Grant Scheme, LRGS (Code no. LRGS/TD/2011/UKM/ICT/02/05/, Grant no.: 9012-00005) of Ministry of Higher Education, Malaysia. Authors would also like to express their gratitude to the R&D Department of Universiti Malaysia Perlis, Malaysia for financial (Graduate Assistance) support and research facilities.

REFERENCES

[1] D. Parkin. 1998. Epidemiology of cancer: global patterns and trends. Toxicology Letters. 5: 102-103.

[2] Abdullah N. Hisham and Cheng-Har Yip. 2004. Overview of breast cancer in Malaysian women: a problem with late diagnosis. Asian Journal of Surgery. 27(2).

[3] Available at: http://www.radiologymalaysia.org/breasthealth/about/FactsNStats.htm. Updated on: updated 11 August 2008.

[4] Breast cancer facts and figures, Atlanta: American Cancer Society, Inc, 2011-2012.

[5] Joy JE, Penhoet EE and Petitti DB. 2005. Saving women's lives: strategies for improving breast cancer detection and diagnosis. The National Academy Press.

[6] Tabar L, Yen MF, Vitak B, Chen HH, Smith RA and Duffy SW. 2003. Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. Lancet. 361: 1405-1410.

[7] Fear E. C., S. C. Hagness, P. M. Meaney, M. Okoniewski and M. A. Stuchly. 2002. Enhancing breast tumor detection with near-field imaging. IEEE Microwave Magazine. 3: 48-56.

[8] Bond E. J., X. Li, S. C. Hagness and B. D. Van Veen. 2003. Microwave imaging via space-time beam forming for early detection of breast cancer. IEEE Trans. Antennas Propag. 51: 1690-1705.

[9] Li X., S. K. Davis, S. C. Hagness, D. W. Weide and B. D. Veen. 2004. Microwave Imaging via Space-Time Beamforming: Experimental Investigation of Tumor Detection in Multilayer Breast Phantoms. IEEE Trans. Microwav. Theory Tech. 52: 1856-1865.

[10] Shannon C., E. Fear and M. Okoniewski. 2005. Dielectric-filled slotline bowtie antenna for breast cancer detection. Electronics Letters. 41: 388-390.

[11] S. A. Alshehri, S. Khatun, A. B. Jantan, R. S. A. Raja Abdullah, R. Mahmud, Z. Awang. 2011. 3D experimental detection and discrimination of malignant and benign breast tumor using NN-based UWB imaging system. Progress In Electromagnetics Research. 116: 221-237.

[12] S. A. AlShehri and S. Khatun. 2009. UWB Imaging For Breast Cancer Detection Using Neural Network. Progress In Electromagnetics Research C. 7: 79–93, .

[13] S. H. Zainud-Deen, W. M. Hassen, E. M. Ali, K. H. Awadalla and H. A. Sharshar. 2008. Breast Cancer Detection Using a Hybrid Finite Difference Frequency Domain and Particle Swarm Optimization Techniques. Progress in Electromagnetics Research B. 3: 35-46.

[14] Y. Ireaneus Anna Rejani and S. Thamarai Selvi. 2009. Early Detection of Breast Cancer Using SVM Classifier Technique. International Journal on Computer Science and Engineering. 1(3): 127-130.

[15] Walker H. Land, Jr., Lut Wong, Dan Mckee, Mark embrechts and Frances Anderson. 2004. Applying Support Vector Machines to Breast Cancer Diagnosis using Screen Film Mammogram Data. 17th IEEE Symposium on Computer -Based Medical Systems (CBMS'04). pp. 224-228.

[16] Wang Yi and Wan Fuyong. 2006. Breast Cancer Diagnosis via Support Vector Machines. 25th Chinese Control Conference. pp. 1853-1856.

[17] Santi Wulan Purnami, S.P. Rahayu and Abdullah Embong. 2008. Feature Selection and Classification of Breast Cancer Diagnosis Based on Support Vector Machines. International Symposium on Information Technology, ITSim. pp. 1-6.

[18] Shang Gao and Hongmei Li. 2012. Breast Cancer Diagnosis Based on Support Vector Machine. International Conference on Uncertainty Reasoning and Knowledge Engineering. pp. 240-243.

[19] Muhammad Hussain, Summrina Kanwal Wajid, Ali Elzaart and Mohammed Berbar. 2011. A Comparison of SVM Kernel Functions for Breast Cancer Detection. 8th International Conference Computer Graphics, Imaging and Visualization. pp. 145-150.

[20] Time Domain Corporation, Cummings Research Park, 330 Wynn Drive, Suite 300, Huntsville, AL 35805, USA.

[21] S. A. Alshehri, S. Khatun, A. B. Jantan, R. S. A. Raja Abdullah, R. Mahmood and Z. Awang. 2011. Experimental Breast Tumor Detection Using Nn-Based UWB Imaging. Progress in Electromagnetics Research. 111: 447-465.

[22] Temko, A., E. Thomas, W. Marnane, G. Lightbody, and G. Boylan. 2011. EEG-based neonatal seizure detection with support vector machines. Clinical Neurophysiology 122 (3): 464-473.

[23] V.N. Vapnik. 1995. The Nature of Statistical Learning Theory. Wiley, New Work.

[24] V.N Vapnik. 1998. Statistical Learning Theory. Wiley, New York, USA.

[25] Simon Haykin. 2001. Neural Networks: A Comprehensive Foundation. Pearson Education, Inc., 2nd Edition.

[26] Tom Fawcett. 2006. An Introduction to ROC analysis. Pattern Recognition Letters. 27: 861-874.