IJETMR
International Journal of Engineering Technologies and Management Research
A Knowledge Repository

ENHANCED SVM CLASSIFIER FOR BREAST CANCER DIAGNOSIS

Dr.M.Kavitha [*1], G.Lavanya [2], J.Janani [3], Balaji.J [4]
[*1] Professor, ECE, Mookambigai College of Engineering, Pudukottai, India
[2] UG Student, ECE, Mookambigai College of Engineering, Pudukottai, India

Abstract:
Breast cancer is the leading disease to cause death especially in women. In this paper, a frame work based algorithm for the classification of cancerous/non-cancerous data is developed using application of supervised machine learning. In feature selection, we derive basis set in the kernel space and then we extend the margin based feature selection algorithm. We are trying to explore several feature selection, extraction techniques and combine the optimal feature subsets with various learning classification methods such as KNN, PNN and Support Vector Machine (SVM) classifiers. The best classification performance for breast cancer diagnosis is attained equal to 99.17% between radius and compact features using SVM classifier. And also derive the features of a breast image in the WBCD dataset.

Keywords: *Breast Cancer; Feature Selection; SVM Classification.*

Cite This Article: Dr.M.Kavitha, G.Lavanya, J.Janani, and Balaji.J. (2018). "ENHANCED SVM CLASSIFIER FOR BREAST CANCER DIAGNOSIS." *International Journal of Engineering Technologies and Management Research,* 5(3), 67-74. DOI: 10.5281/zenodo.1207413.

1. Introduction

Breast cancer (malignant breast neoplasm) is cancer originating from breast tissue. So, there is a need for a reliable and an objective classification tool for detecting and classifying the breast cancer cases namely benign and malignant. Machine learning and neural networks help us out by providing a better classification tool that is reliable and objective too.

A neural network consists of an interconnected group of artificial neurons, [9] and it processes information using a connectionist approach to computation. Modern neural networks are non-linear statistical data modeling tools.

In that, Machine learning is an emerging approach to classifying and computing results which help technicians to take decisions in an environment of uncertainty and imprecision. The need of neural network is that unlike the traditional, hard computing ,it is aimed at an accommodation with the pervasive imprecision of the real world. Thus, the guiding principle of neural network is to exploit the tolerance for imprecision, uncertainty and partial truth to achieve tractability. The NN models can be classified according to various criteria, such as their learning methods, architecture, implementation, operation and so on. The scope of the project is to model problems

ISSN: 2454-1907
DOI: 10.5281/zenodo.1207413

with desire input output data sets, so the resulting network must have adjustable parameters that are updated by supervised learning rule. Under the category of supervised learning, perception is one kind of classifiers performing the classification in the two dimensional space. There is need of an algorithm which is capable of classify the candidates, which have the closest similarities. Support vector machine is such an algorithm which has direct bearing on machine intelligence.

Support vector machine [24] can be used as a best classification tool for classifying any kind of dataset even a nonlinear one. In this project the dataset regarding the cancerous and non-cancerous cases of breast cancer are been taken as input for training and classified in the feature space using a hyper plane and its classification performance is also calculated.

2. Materials and Methods

In the past, the feature selection was done by filter approach [7]. It doesn't account the bias of induction algorithm. In earlier classification was done by KNN and PNN. In proposed system a novel approach for feature selection is, namely wrapper approach and for classification is SVM.Wrapper approach [6] use induction algorithm and handle complex datasets and it estimates by LOOCV. In SVM produce minimum error rates and reduce the misclassifications.

2.1. Feature Selection in Kernel Space

Step1: Constructing a basis set by either kernel GP (or) kernel PCA.
Step2: Calculating weight by kernel RELIEF.
Step3: Ranking features by weight
Step4: Select features based on the rank.
Step: 1 Creating subsets:

Algorithm 1:
FSKGP (FEATURE SELECTION IN KERNEL GRAM SCHMIDT PROCESS)
Input: data x (i) (i = 1..N)
Output: an orthogonal set of basis vectors
```
 for i = 1 to N do
   v (i) = -(x (i))
    for j = 1 to i − 1 do
      v (i) ← v (i) − h-(x (i)), v (j) iv (j)
end for
Normalize: v (i) ←  v (i)/ ||v (i) ||
end for
```
Output: basis set {$v^{(i)}$}
Algorithm 2:

FSKSPCA (FEATURE SELECTION IN KERNEL SPACE PRINCIPAL COMPONENT ANALYSIS):
Input: training data xi, label yi
Output: selected features in the kernel space
1: Constructing a basis set by either Kernel GP or Kernel PCA

 ISSN: 2454-1907
DOI: 10.5281/zenodo.1207413

2: Calculating wi by Kernel Relief
3: Ranking implicit features by wi, select features based on the rank
4: Projecting the data into the learned subspace

2.2. Classification Models

2.2.1. KNN (K-Nearest Neighbor)

KNN classifier is one of the simplest and oldest methods for performing general, nonparametric classification [4]. In this model, the distances between the test sample and all the other samples in the training set is first measured. Then, the class of the test sample is assigned according to a simple majority vote over the labels of its K nearest neighbors.

2.2.2. PNN (Probabilistic Neural Network)

The Probabilistic Neural Network: PNN was proposed by Specht in 1988 [18]. It is designed to improve the performance of conventional neural networks in which long computation times are required. PNN replaces the sigmoid activation function often used in neural networks with a statistically derived exponential function. The PNN is an extension of what is probably the simplest possible classifier i.e., find the training sample closest to the test sample and assign it the same class. A single PNN is capable of handling multiclass problem. This is opposite to the so-called one-against-the rest or one-per class approach taken by some classifiers, such as the SVM, which decompose a multiclass classification problem into dichotomies and each chotomizer has to separate a single class from all others [19].

2.3. Svm Representation

2.3.1. Linear SVM

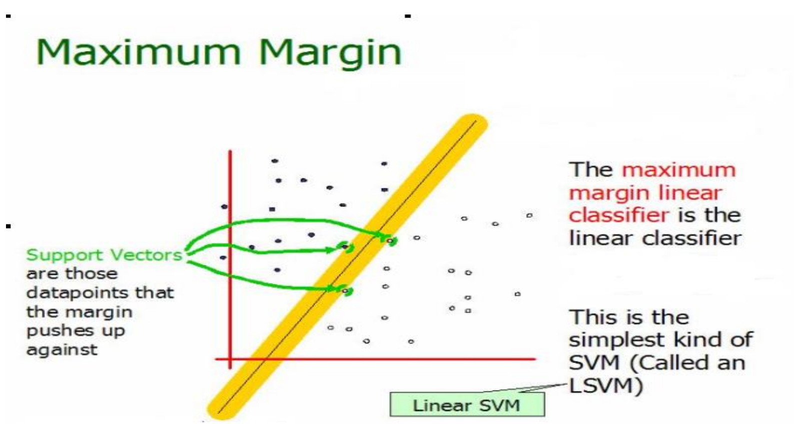

Figure 1: Linear SVM

Expression for Maximum margin is given as

$$\text{margin} \equiv \underset{\mathbf{x}\in D}{\arg\min}\, d(\mathbf{x}) = \underset{\mathbf{x}\in D}{\arg\min} \frac{|\mathbf{x}\cdot\mathbf{w}+b|}{\sqrt{\sum_{i=1}^{d} w_i^2}}$$

 ISSN: 2454-1907
DOI: 10.5281/zenodo.1207413

2.3.2. Representation of Hyper Plane

The goals of SVM are separating the data with hyper plane and extend this to non-linear boundaries using kernel trick [15] .For calculating the SVM we see that the goal is to correctly classify all the data. For mathematical calculations we have,

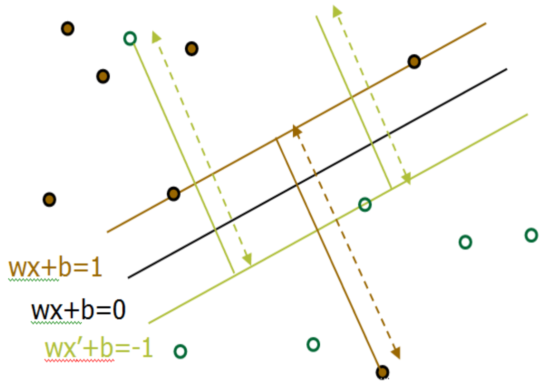

Figure 2: Representation of hyper plane

1. If Yi=+1; wxi+b>=1
2. If Yi=-1; wxi+b<=1
3. For all i; yi (wi +b)>=1

X →vector point
W →weight
Maximum margin
M=2/||w||

2.3.3. Representation of Support Vectors

The solution involves constructing a dual problem and where a Langlier's multiplier αi is associated. We need to find w and b such that Φ (w) =½ |w'||w| is minimized
And for all {(xi, yi)}: yi (w * xi + b) ≥ 1
Now solving: we get that
w =Σαi * xi
b= yk- w *xk for any xk such that αk≠ 0
Now the classifying function will have the following form:
F(x) = Σαi yi xi * x + b

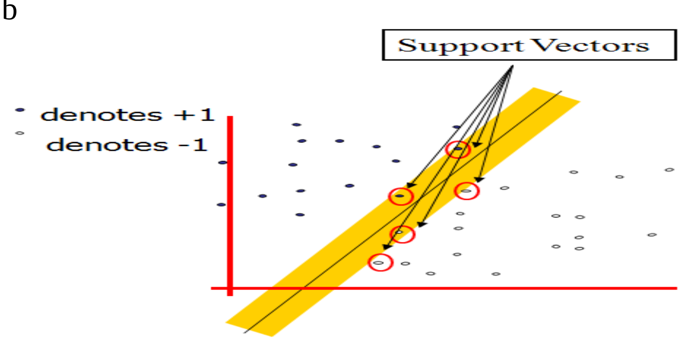

Figure 3: Representation of Support Vectors

ISSN: 2454-1907
DOI: 10.5281/zenodo.1207413

3. Results and Discussions

For the feature selection process, whole medical data is separated into two halves for training and testing. This process has been done automatically by SVM classifier. After that SVM classification, we can obtain the classification performance result in percentage in figure 4, 5, 6 and 7.

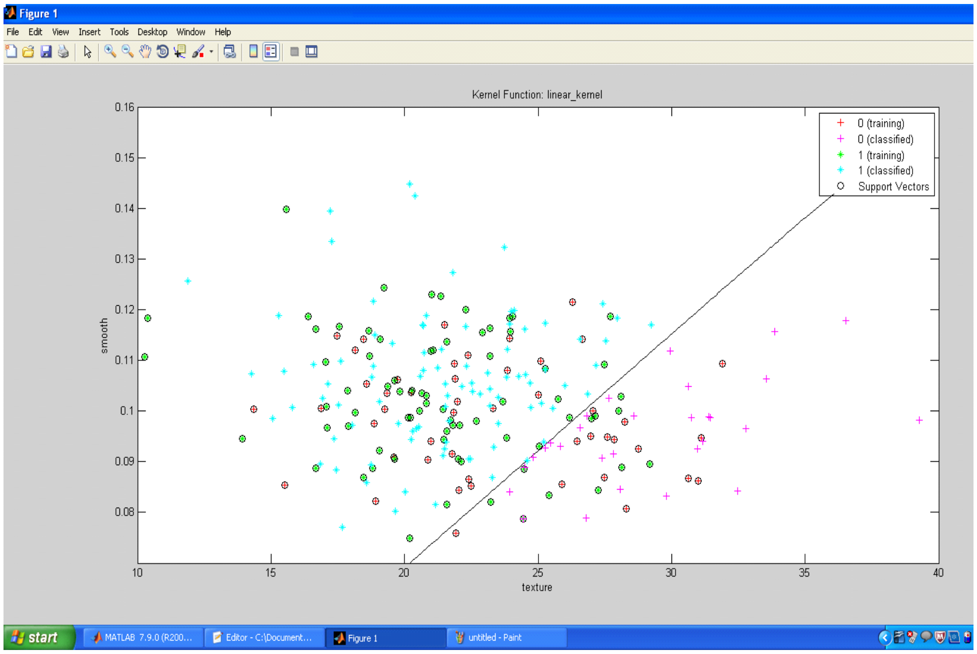

Figure 4: Texture Vs Smooth

CLASSIFICATION PERFORMANCE (CP) = **52.89%**

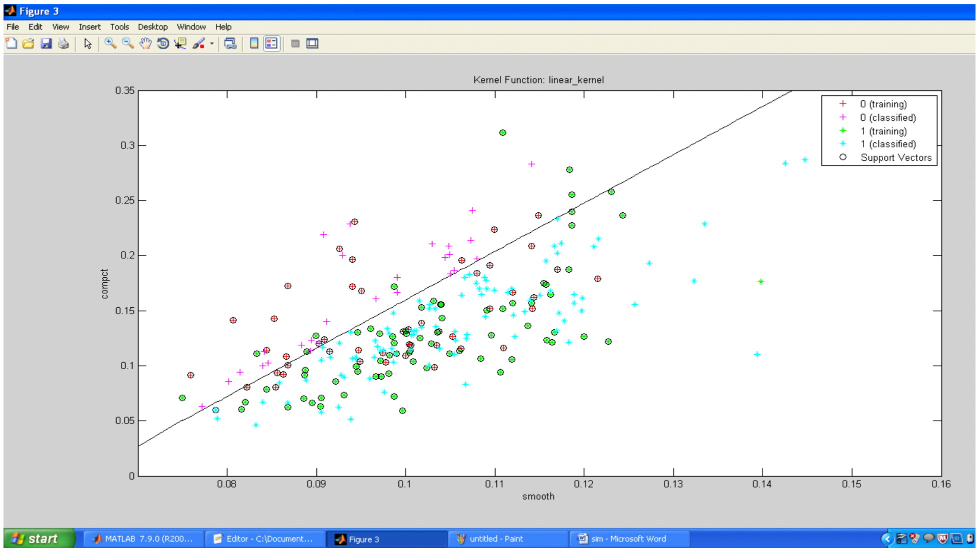

Figure 5: Smooth Vs Compact

CLASSIFICATION PERFORMANCE (CP) = **59.50%**

ISSN: 2454-1907
DOI: 10.5281/zenodo.1207413

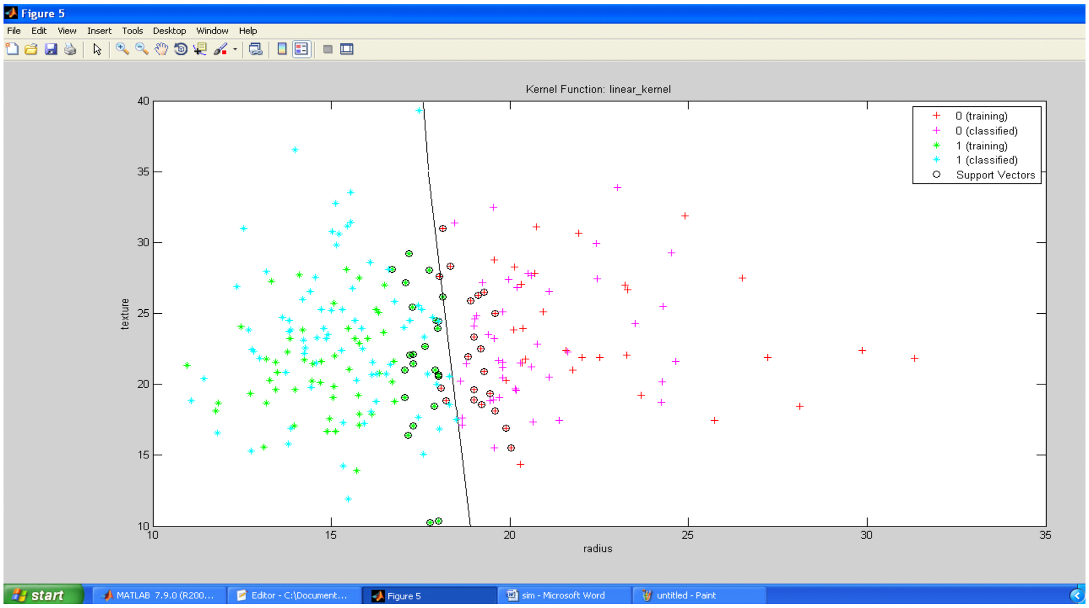

Figure 6: Texture Vs Radius

CP=96.69%

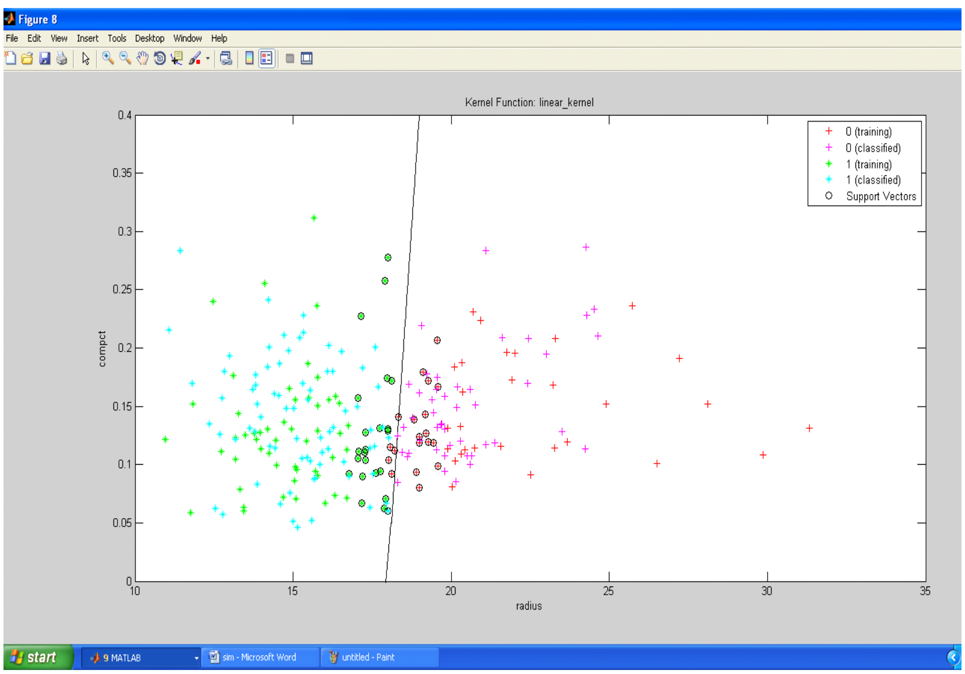

Figure 7: Radius Vs Compact

CP=99.17%

4. Conclusions and Recommendations

We have done classification of many features using SVM and also obtain classification performance percentage. Compare all the features compact and radius has maximum

 ISSN: 2454-1907
DOI: 10.5281/zenodo.1207413

classification performance. So, these two features are considered as efficient features that have more relevant information about breast cancer. Feature selection process has been done.

In this work, FEATURE SELECTION in SVM is done by training and testing the WBC data and also finds the CLASSIFICATION PERFORMANCE for breast cancer diagnosis. Experimental results along with classification percentage shows that while comparing all features, the features that have highest classification performance percentage such as radius and compact are identified as the best features used for breast cancer diagnosis.

This method of classification through machine learning algorithm say support vector machine algorithm is a reliable and efficient methodology and this can be used for any kind classification provided the reliable data sets are available. Not only in medical field, this kind of classification through support vector machine can be done in all developing and unpredictable fields like Stock Market exchange, Weather Conditions, Predicting Natural Calamities, Automobile MPG predictions and so on. The best classification performance for breast cancer diagnosis is attained equal to 99.17% between radius and compact features using SVM classifier

Acknowledgements

The completion of this paper has been made possible because of the involvement of many individuals at various stages. We would like to thank all of them in this regard. We would like to express our sincere thanks to our department professors for extending full co-operation and support. Finally we express our gratitude to each one of our faculty members, friends and family members for their timely help and constant support.

References

[1] Andrei V. Anghelescu, Ilya and B. Muchnik," Combinatorial PCA and SVM Methods for Feature Selection in Learning Classifications (Applications to Text Categorization)", KIMAS 2003 BOSTON, USA
[2] Burges C., "A tutorial on support vector machines for pattern Recognition", In "Data Mining and Knowledge Discovery". Kluwer Academic Publishers, Boston, 1998.
[3] Bin Cao, Dou Shen, [et al]," Feature Selection in a Kernel Space", Peking University, Beijing, China, 2006.
[4] Catapano I, Crocco L, Scapaticci. R [et al]," Recent results on a novel microwave breast cancer imaging approach based on magnetic nanoparticle as contrast agent", University of Naples Federico II,2012.
[5] Francis arena M.D., Element Barone M.D and Thomas Dicicco ,'' Use of Digital Infrared Imaging in Enhanced Breast Cancer Detection and Monitoring of the Clinical Response to Treatment", arena oncology associates, great neck, NY, c.m. barone, M.D. P. C., NY, NY, infrared sciences corp., smithtown, NY.
[6] Fear Elise C, Member, IEEE, Xu Li, Student Member, IEEE, Susan C. Hagness, member, IEEE, and Maria A. Stuchly, Fellow, IEEE "Confocal Microwave Imaging for Breast Cancer Detection: Localization of Tumors in Three Dimensions", IEEE transactions on biomedical engineering, vol. 49, no. 8, august 2002.
[7] Freund R., Osuna E, and Girosi F., "Support Vector Machines: Training and Applications", A.I. Memo No. 1602, Artificial Intelligence Laboratory, MIT, 1997.

ISSN: 2454-1907
DOI: 10.5281/zenodo.1207413

[8] Image found on the web search for learning and generalization in SVM following links given in the book above
[9] Ireaneus. Y, Anna Rejani and Dr. S. Thamarai Selvi." EARLY DETECTION OF BREAST CANCER USING SVM CLASSIFIER TECHNIQUE", International Journal on Computer Science and Engineering Vol.1(3), 2009, 127-130.
[10] JANG. J.-S.R, SUN. C.-T and E.MIZUTANI. "Neuro-Fuzzy AND Soft Computing", A Computational Approach to Learning and machine Intelligence.
[11] Kung. S.Y," ON FEATURE SELECTION FOR GENOMIC SIGNAL PROCESSING AND DATA MINING", Princeton University, USA, 2007.
[12] Khamsa Djaroudib, Abdelmadjid Zidani and Abdelmalik Taleb Ahmed," Mass Abnormality Segmentation in Mammographic Images for Different Densities of Tissue", International Conference on Control, Engineering & Information Technology (CEIT'13) Proceedings Engineering & Technology - Vol.1, 2013.
[13] Mohammad Darzi, Ali Asghar Liaei,Mahdi Hosseini and Habibollah Asghari," Feature Selection for Breast Cancer Diagnosis: A Case-Based Wrapper Approach", World Academy of Science, Engineering and Technology 53 2011.
[14] Poh Chee Khun, Zhang Zhuo, Liang Zi Yang, Li Liyuan, and Liu Jiang," Feature Selection and Classification for Wireless Capsule Endoscopic Frames", Agency for Science, Technology and Research, Singapore,2009.
[15] Ramani, R. Valarmathy. S. and Suthanthira Vanitha. N," Breast Cancer Detection in Mammograms based on Clustering Techniques- A Survey", International Journal of Computer Applications (0975 – 8887) Volume 62– No.11, January 2013.
[16] Shai Avidan," Joint feature-basis subset selection", Mitsubishi Electric Research Lab, Broadway, Cambridge, 2004
[17] Sonia Narang, Harsh K Verma and Uday Sachdev," A Review of Breast Cancer Detection using ART Model of Neural Networks", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 10, October 2012.
[18] The Math Works, Inc., [Online]. Available: http://www.mathworks.com/
[19] Tinghua Wang," FS_KPARD: An Effective SVM Feature Selection Method", Sixth International Conference on Natural Computation (ICNC 2010), 2010.
[20] Vapnik.V. The Nature of Statistical Learning Theory. Springer, N.Y., 1995. ISBN0-387-94559-8.
[21] Wikipedia Online. http://en.wikipedia.org/wiki
[22] Xu. J.Q. and Yuan. Z.D,'' A Feature Selection Method Based on Minimizing Generalization Bounds of SVM via GA", Proceedings 01 the 2003 IEEE International Symposium on Intelligent Control Houston. Texas October 5-8. 2003.
[23] Yaman Aksu, George Kesidis, and David J Miller," SCALABLE, EFFICIENT, STEPWISE-OPTIMAL FEATURE ELIMINATION IN SUPPORT VECTOR MACHINES", Pennsylvania State University University Park, PA,2007.
[24] Yaohua Tang, Weimin Guo and Jinghuai Gao," Efficient Model Selection for Support Vector Machine with Gaussian Kernel Function", 2009.
[25] Zhenyu Chen and Jianping Li," Least Squares Support Feature Machine", Graduate University of Chinese Academy of Sciences, China, 2006.

*Corresponding author.
E-mail address: kavitha79ramar@ gmail.com