International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 17 (2015) pp 37717-37723

Breast Cancer Detection Using Relevance Vector Machine

B.M.Gayathri
Research Scholar, Manonmaniam Sunderanar University,
S.D.N.B Vaishnav College for Women, Chromepet, Chennai-600044
gayathri_bm2003@yahoo.co.in

Dr.C.P.Sumathi,
Associate Professor & Head, Department of Computer Science
S.D.N.B Vaishnav College for Women, Chromepet, Chennai-600044
drcpsumathi@gmail.com

Abstract
Artificial Intelligence is used in the fields like medical diagnosis, robotics, Aviation, Music. The study of AI is divided into many branches. Some of the applications in which AI is used are Pattern recognition, Artificial Creativity, Image Processing, and Machine Learning. This paper deals with one of the application of AI that is Machine Learning. It has strong ties to artificial Intelligence. There are many techniques that come under machine learning for pattern recognition. One of the technique is Relevance Vector Machine (RVM).The aim of this article is to apply RVM to classify breast cancer and get better results with fewer numbers of attributes.

Keywords: Relevance vector machine, Support Vector Machine, Principal component analysis, Linear Discriminant analysis.

Introduction

The most deadly cancer among women is Breast cancer. It is nothing but the cancer that develops from the tissue of breast. Some of the signs of Breast cancer are Lump in the breast, red scaly patch of skin, shortness of breath, bone pain. Detecting the tumor in the earlier stage is still very difficult. One of the ways to reduce the death rate is to diagnose and remove the part that is affected by cancer. The researchers are developing many applications for diagnosing cancer in effective way. They use mammogram images, Fine Needle Aspiration (FNA) for detecting tumor. This work deals with one of the machine learning algorithm, Relevance Vector Machine (RVM), for detecting tumor with fewer attributes. In this model only four variables are used for diagnosing and these variables are selected using one of the variable selection techniques called Linear Discriminant Analysis (LDA). In this paper breast cancer classification using Support vector machine (SVM) is done and also the accuracy difference between Support Vector Machine and Relevance Vector Machine is also discussed. This paper is organized in the following manner. Section 2 describes the information about Data Preprocessing and feature selection process techniques used. Techniques, methodology and implementation used for detecting breast cancer is described in section 3.Section 4 consist of discussion and conclusion.

Data preprocessing and Feature Selection.

A. Data preprocessing

This is the first step to be performed before the classification is done. The dataset used in this work is taken from UCI repository, Wisconsin original dataset. The dataset consist of totally 10 attributes. The following table 1 shows the attributes and Figure 1 shows the samples of dataset.

Table 1: Attributes of dataset

S.no	Attributes/Features	Range of values
1.	Clump thickness	1-10
2.	Uniformity cell size	1-10
3.	Uniformity cell shape	1-10
4.	Marginal Adhesion	1-10
5.	Single Epithelian Cell size	1-10
6.	Bare Nuclei	1-10
7.	Bland Chromatin	1-10
8.	Normal Nucleoli	1-10
9.	Mitosis	1-10
10.	Class	2-benign 4-Malignant

1 : clumpthickness 5

	clumpthick	uniformi	uniformity_	marginal_a	single_epit	Bare_nucle	Bland_chro	Normal_nu	Mitosis	Class
1	5.00	1.00	1.00	1.00	2.00	1.00	3.00	1.00	1.00	2.00
2	5.00	4.00	4.00	5.00	7.00	10.00	3.00	2.00	1.00	2.00
3	3.00	1.00	1.00	1.00	2.00	2.00	3.00	1.00	1.00	2.00
4	6.00	8.00	8.00	1.00	3.00	4.00	3.00	7.00	1.00	2.00
5	4.00	1.00	1.00	3.00	2.00	1.00	3.00	1.00	1.00	2.00
6	8.00	10.00	10.00	8.00	7.00	10.00	9.00	7.00	1.00	4.00
7	1.00	1.00	1.00	1.00	2.00	10.00	3.00	1.00	1.00	2.00
8	2.00	1.00	2.00	1.00	2.00	1.00	3.00	1.00	1.00	2.00
9	2.00	1.00	1.00	1.00	2.00	1.00	1.00	1.00	5.00	2.00
10	4.00	2.00	1.00	1.00	2.00	1.00	2.00	1.00	1.00	2.00
11	1.00	1.00	1.00	1.00	1.00	1.00	3.00	1.00	1.00	2.00
12	2.00	1.00	1.00	1.00	2.00	1.00	2.00	1.00	1.00	2.00
13	5.00	3.00	3.00	3.00	2.00	3.00	4.00	4.00	1.00	4.00
14	1.00	1.00	1.00	1.00	2.00	3.00	3.00	1.00	1.00	2.00
15	8.00	7.00	5.00	10.00	7.00	9.00	5.00	5.00	4.00	4.00
16	7.00	4.00	6.00	4.00	6.00	1.00	4.00	3.00	1.00	4.00
17	4.00	1.00	1.00	1.00	2.00	1.00	2.00	1.00	1.00	2.00
18	4.00	1.00	1.00	1.00	2.00	1.00	3.00	1.00	1.00	2.00
19	10.00	7.00	7.00	6.00	4.00	10.00	4.00	1.00	2.00	4.00
20	6.00	1.00	1.00	1.00	2.00	1.00	3.00	1.00	1.00	2.00
21	7.00	3.00	2.00	10.00	5.00	10.00	5.00	4.00	4.00	4.00

Figure 1: Dataset Sample view

B. Replace missing values

At first the dataset is divided randomly for training and testing and those data must be checked, whether the data contains outliers, missing values, inconsistent values etc. In this dataset

there was 16 missing values totally including training and testing data. There are different methods for filling missing values, they are a) ignoring the tuple if the class label is missing [10] b) finding out mean of attribute to fill the missing values for same class or for different class [10]. This method is used mostly for discrete values. c) Predicting the missing values by using any one of the learning algorithm, such as decision tree, cluster algorithm etc.

In this work, method "b" is used as the dataset has discrete values. Method "a" is not used because there is no missing class value. Technique c is not used, as the use of learning algorithms is complex than finding the mean or median value for the attributes.

C. Feature Selection

This is the second step to be done before performing the implementation process. This plays an important role in data mining. Performing this process improves the accuracy of the dataset and also reduces the time complexity. First normalize the data by finding Log_{10} for each and every attribute. Next, the relevant attributes are selected by using Fisher's Linear Discriminant analysis (LDA) which is used for selecting linear combination of features. All the nine attributes with its values are given as input to the function LDA and the result gives the features that can be used for classification. Out of nine attributes, four attributes that gave good result are chosen for classification. One more method called Principal Component Analysis (PCA) is used for selecting variable parameter, which is to be used in training RVM function. Generally, PCA converts set of correlated variable into linearly uncorrelated variables.

D. Principal Component Analysis

PCA is a statistical technique which can be used to reduce the dimension of data. The aim of PCA is to find a new set of attributes. For selecting the attributes the following criteria must be met. They are a) The attributes must have linear combinations, b)The attributes must be orthogonal to each other, c)select those attributes which have maximum variation in the data. The variable is selected by the result that is provided by PCA. The variable that has the largest variance is selected as variable parameter.

i. Variable Parameter Selection using Principal Component Analysis

By calculating PCA the original data is first transformed into covariance matrix and then its principal components (eigenvectors) are extracted. The principal components are ranked based on the amount of variation in the original data. The attributes that varies the most are selected and remaining are discarded. The main advantage of PCA is that, it is the only unsupervised method which does not use the class attribute. In this work the PCA is used to select the variable parameter. This parameter is used in RVM for classification. The data that gave the maximum variance is selected as a variable parameter.

ii. Steps for performing Principal component Analysis

1. Select the dataset which consist of N-dimensional of dataset samples by ignoring the class labels, because PCA do not need class label for classification as it is an unsupervised method.
2. Compute the mean for whole dataset i.e., mean should be calculated for each and every dimension in the dataset.
3. Calculate the Covariance matrix for the full dataset.
4. Calculate Eigen vectors and its corresponding Eigen values.
5. Sort the Eigen vectors in descending order based on Eigen Values.

Create a $(N \times K)$ matrix by selecting K Eigen vectors with largest Eigen values. By using this equation $Y = W^t \times x$, $(N \times K)$matrix can be transformed into the new subspace, where x=$N \times 1 - \text{dimensional vector}$ which represents one of the sample and Y is transformed new subspace, W=Transformation matrix. The values that are used while calculating PCA is shown in the following table 2 and table 3. These values are used internally for calculation

Table2: Sample of Principal component Coefficients

Attributes	Clthick	UcSize	UC Shape	Marg.adhesion
Clthick	-0.3032	0.3566	0.3550	0.3024
Ucsize	0.2072	0.0319	0.0358	0.0260
UCShape	0.6885	0.0249	0.0728	0.5380
Marginal Adhesion	0.0151	-0.2777	0.2445	0.4750
Norm nucleoli	-0.4736	-0.4011	0.5233	0.2606

Table3: Sample of Scores (observations, columns to components)

Clump thickness	Uniformity cell size	Uniformity cell shape	Marginal Adhesion
-1.942	-0.1193	0.4639	-0.0223
0.6452	-0.2188	-0.6526	0.1249
-2.0555	-0.0594	0.0173	0.0839
0.7656	-0.2158	0.4319	-1.6217
1.8336	-0.0343	-0.1273	0.2975
4.7156	1.3263	-0.9303	-0.5583
-1.5861	-0.5212	-0.4275	1.0020
-2.1184	0.0704	-0.1823	-0.1187
-2.0722	2.0074	0.4899	0.6414
-2.0656	0.0649	0.3537	-0.1148

The following figures 2 and 3 show the result of Principal component analysis.

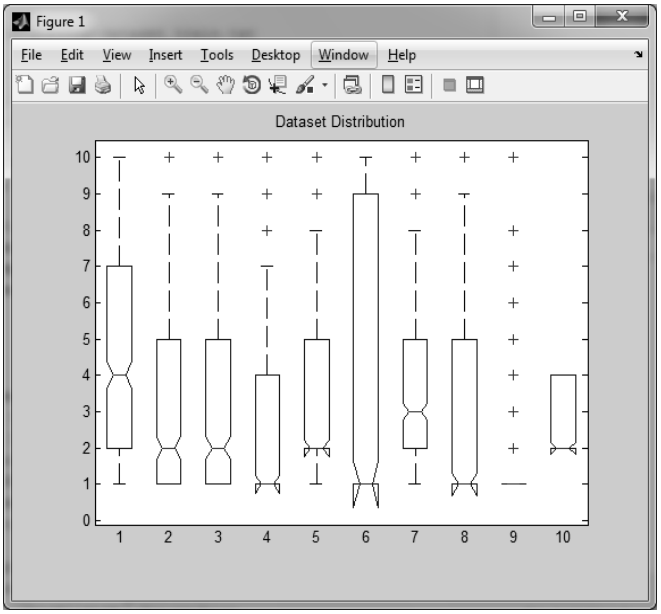

Fig 2: Distribution of Dataset

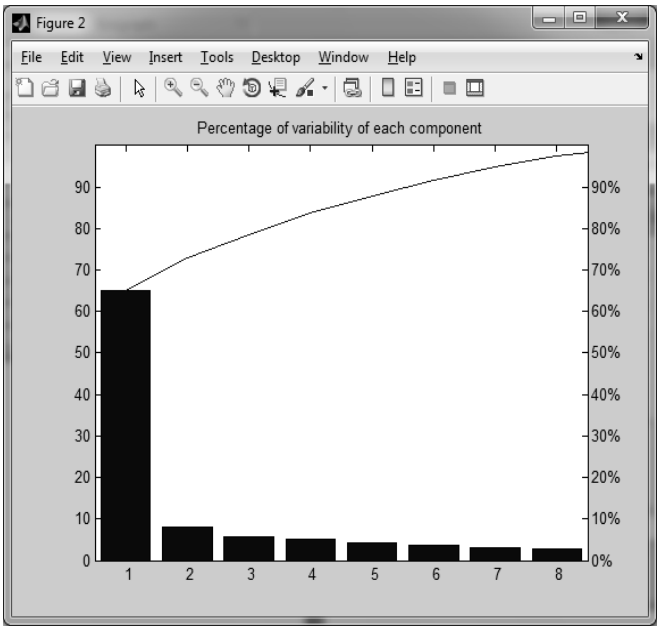

Fig 3: Plot for Variability of each plot in percentage

The plot of figure 4 shows the attributes with highest percentage of variability. The first attribute have highest percentage of variability. Hence in the RVM function, the variable that shows the highest variability is used as variable parameter. The following figure 4 shows the 3D plot for PCA coefficients and scores for each attribute.

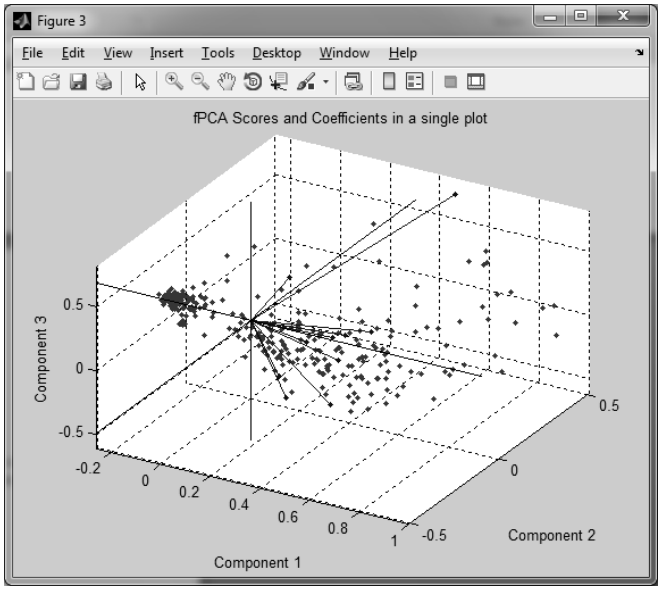

Fig 4: 3D plot for PCA Coefficients and scores for each variable.

E. Linear Discriminant Analysis

The main idea of LDA is to separate classes by finding projection to a line. The line that projects good shows that the classes are well separated. The following figure 5 shows the proper class separation.

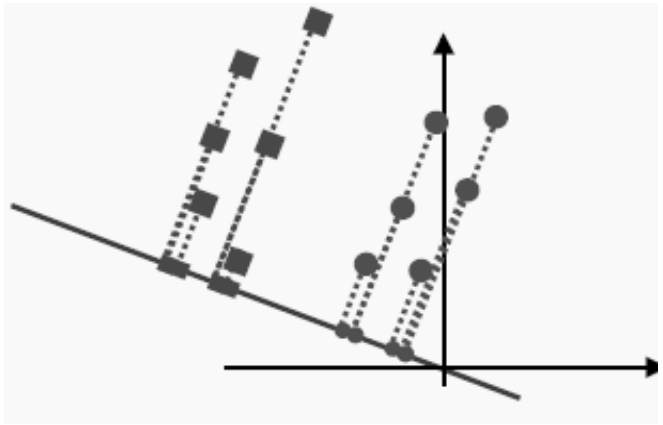

Fig 5: Projected line shows that classes are well separated

LDA is mainly used for dimensionality reduction.The reduction is calculated with the result that shows the maximum and minimum outcome of variances of class within the training dataset. The Linear Discriminant can be calculated by using the formula $D_i = b_0 + \sum_{k=1}^{p} b_k X_k$ where X =raw scores of each predictor, b= constant, p=target. The scoring function of LDA is $C_i = c_1 Z_1 + c_2 Z_2 + \cdots + c_p Z_p$ where c_i = discriminant function coefficient, Z=score on each predictor, p=target. By using the above function, different combinations of Discriminant equations (D_i) can be created, so that different combination of groups can be created

for example, discriminant function D_1 may create first set of group.
From the result of first group, second group is created with some differences, next third set of group is created which distinguishes from second set of group and so on. The resultant scores are stored so that it can be used for classification. The main aim of using LDA is to select the attributes that gives the best classification result and those attributes can be used for detecting breast cancer. This reduces the classification time and cost. One of the limitation of LDA is, categorical value cannot be used for classification.

i. Steps for performing LDA for feature selection

1. Load Dataset
2. Data Preprocessing. (Ex: Replacing Missing values)
3. Perform Linear Discriminant Analysis.
4. From the stored resultant score, select the set of attributes that gives maximum accuracy.

The following tabular column shows the different set of attributes that is selected for classification and its accuracy.

Table 3: Combination of attributes selected for analysis and its accuracy

S.no	Attributes	Accuracy
1.	Clumpthickness, Uniformity cell size, Uniformity cell shape, Marginal Adhesion	**95.0%**
2.	Clumpthickness, Uniformity cell size, Uniformity cell shape, Single epithelian cell size.	94.3%
3.	Clumpthickness, Uniformity cell size, Uniformity cell shape, Bland Chromatin	94.3%
4.	Clumpthickness, Uniformity cell size, Uniformity cell shape, Normnucli	94.7%
5.	Clumpthickness, Uniformity cell size, Uniformity cell shape, Mitosis	94.3%
6.	Clumpthickness, Uniformity cell size, Uniformity cell shape, Bare Nuclei	94.3%
7.	Bare nuclei, Marginal adhesion, Single epithelian cell size, Normal nucleoli	92.7%
8.	Bland chromatin, Marginal adhesion, Single epithelian cell size, Normal Nucleoli	92.7%
9.	Bland Chromatin, Marginal adhesion, Single epithelian cell size, mitosis.	91.3%

The features that gave the best accuracy are selected for classifying breast cancer by using machine learning algorithms SVM and RVM.
The following figure 6 shows the sample output of LDA.

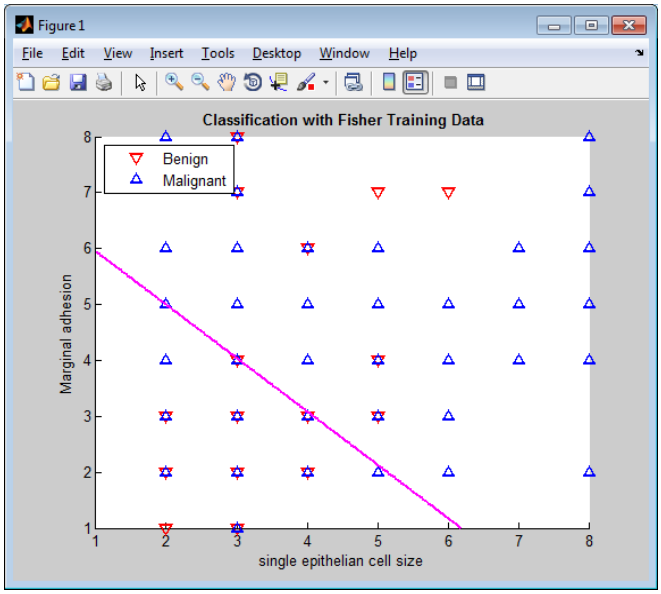

Fig 6: Linear Discriminant Analysis sample output

Techniques used for detecting breast cancer

A. Support vector Machine

SVM was introduced by Vapnik Chervonenkis. The use of SVM is to classify and find an optimal separating hyperplane. This generates a margin between two classes of data and non linear kernel function is used to perform this task [3]. SVM prevents overfitting. The maximum margin separator is found out by using the optimization equations such as $w.x$+b > +1, $w.x$+b< -1 for positive and negative cases where x=real vector, w= normal vector (not normalized) to the hyperplane. This test is done while training linear SVM. While testing the following equation must be satisfied i.e. $w.x$+b > 0 and $w.x$+b <0. Figure 7 is the sample of hyperplane which shows two different classes and support vectors.

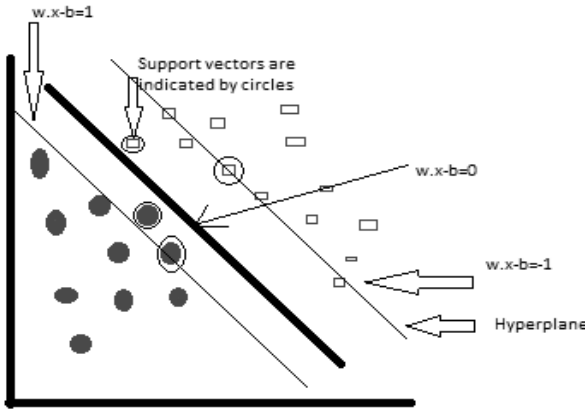

Fig 7: Example of hyper plane and samples of SVM that shows two trained classes. Circle indicates Support vectors.

i. Implementation and Methodology using SVM

Initially the dataset is randomly divided into two parts, such as testing and training. LDA is used to reduce the attributes. The dataset used for classification is Wisconsin original dataset. It consists of totally nine attributes out of which four variables are selected using the feature selection methodology. Reducing the variables reduces the complexity of training and testing. The dataset had 699 records out of which 16 records had missing values and it is replaced by one of the data preprocessing method called data cleaning by which the missing data are replaced by finding mean of that attribute. The data are classified by using SVM classifier. The performance is evaluated and it gave an accuracy of up to 94%. The sensitivity and specificity were 92% and 93% respectively.

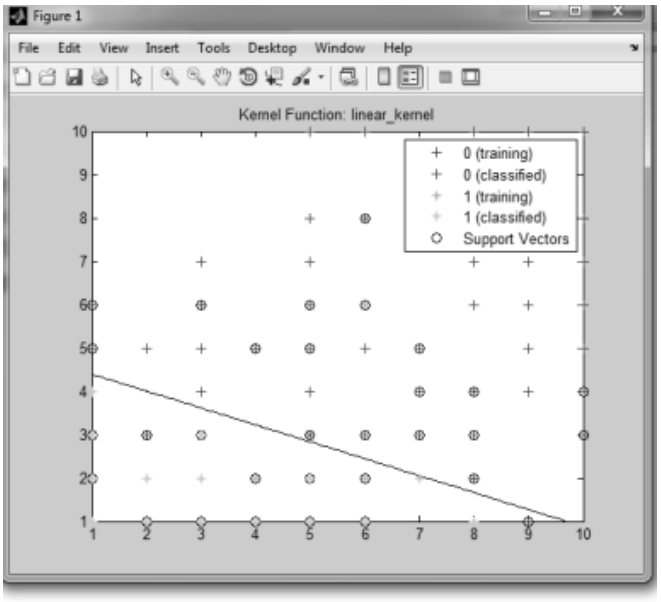

Fig 8: Sample output of SVM Classifier

C. Relevance Vector Machine

RVM is a fully probabilistic model which is based on Bayesian Maximum A Priori (MAP). The main aim is to separate the input data into different classes by predicting posterior probabilities of their class membership [6]. The Functional form of RVM is

$K(x, x') = \sum_{j=1}^{N} \frac{1}{\alpha} \varphi(x, x_j)\varphi(x', x_j)$ where φ is the kernel function. Variable x refers to input vectors of training set and α represents variable parameter. N shows the target. Some of the most common kernel functions that can be used are Fisher's Kernel, Linear kernel, RBF kernel etc.

i. Previous Works

RVM is used by many researchers for their research work. They have used RVM not only in medical field but also in other fields like weather forecasting (Guoqian sun et al., 2014), Electro discharge Machining process (Kanhu charan nayak et al., 2013) etc. In medical field RVM is used for detecting breast cancer, ovarian cancer, optical cancer etc. RVM is also used for the assessment of neonate pain intensity of a patient using digital images. (Behnood Gholami et al.2010).

Shovan K.Majumder (et al., 2005) has used RVM for optical cancer diagnosis. The algorithm is developed by using vivo auto fluorescence spectral data.The dataset was collected from Government Cancer Hospital, Indore, India. The author has shown the comparison between the performance of SVM and RVM, in which RVM shown good classification result than SVM.

Elie Tcheimegni (et al.,) has developed application using RVM using different clinical data, not only for detecting breast cancer but also for other general diseases. The author has also done a comparison between SVM and RVM and predicted that RVM shows better accuracy than SVM. The author has used all the attributes of Wisconsin dataset for prediction.

Dr.S.SanthoshBaboo (et. al.) proposed a technique for classifying cancer in living organisms using microarray gene expression data. They have used relevance vector machine for detecting cancer cells with the ANOVA technique. The performance was evaluated by using Lymphoma and Leukemia datasets.

Liyang Wei (et. al.) has used RVM for detecting clustered microcalcifications automatically. He has used 141 clinical mammograms for testing and compared with SVM and the result showed that RVM shows good accuracy than SVM and also it reduces computational complexity.

The proposed method is different from the previous works. In this work Wisconsin original data set is used and the system is designed as a user friendly environment.User can enter the details of cancer, which calculates and predicts the cancer whether it is benign or malignant. In this only four variables are used for prediction

ii. Implementation and Methodology using RVM.

The proposed model runs on Intel core i3 processor. The Wisconsin original dataset is used for testing and training the data. The data are classified by using RVM algorithm. The GUI is designed for user friendly environment. The user can input the data, and the given data is classified as Benign or Malignant. The result is displayed in the screen. The accuracy is checked by entering each record manually. Hence by finding out the True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN), the accuracy is calculated as (TP+TN)/ (TP+FP+FN+TN) which is equal to 96%. The sensitivity and specificity are 98% and 94% respectively.

Table 4: Classification using RVM classifier

Dataset	True Positives	False Positives	True Negatives	False Negatives
300	98%	0.02%	96%	0.04%

Table 5: Performance Evaluation of SVM & RVM models

Dataset	Algorithm	Sensitivity	Specificity	Accuracy
300	SVM	92%	93%	93%
	RVM	98%	98%	97%

Fig 9. Sample output of RVM classifier.

Discussion and Conclusion

Relevance vector machine is a powerful statistical machine learning tool used for classification. Many authors have used RVM not only in the field of Medicine, but also in the other fields like forecasting wind speed[X Liu et al.], Rainfall Prediction[P Samui et al.] etc. The main aim of presenting this paper is to show that the RVM can be used to detect breast cancer with fewer attributes. In this paper, SVM and RVM are successfully applied for diagnosing breast cancer with few attributes. Even though both models show good classification results the RVM gives better result than SVM. The following figure 10 shows the comparison of SVM and RVM proposed in this model.

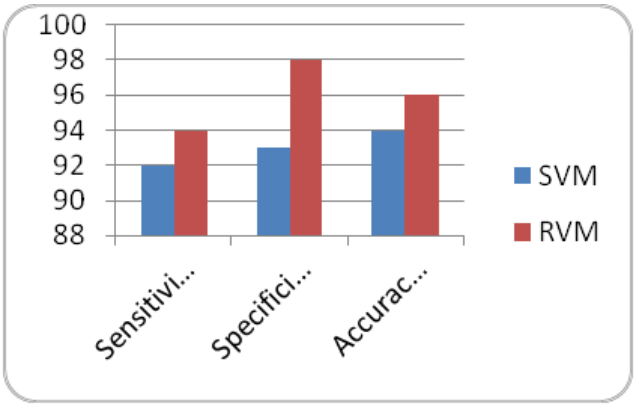

Fig 10: Comparison between SVM & RVM models

From the performance evaluation (table 5), we can see that RVM shows better accuracy than SVM and also figure 10 shows the same. The dataset used in this model is Wisconsin original dataset. In future, the same model can be trained in the real time dataset also.

References

[1] Bhanumathi, R., and Suresh, G.R., "Latest Advances in Computer-Aided Detection of Breast Cancer by Mammography", International Journal in IT and Engineering, Vol. 01 No. 6, 2013, ISSN: 2321–1776.

[2] Karpagachelvi, S., Dr.Arthanari, M,, and Sivakumar, M., "Classification of ECG Signals Using Extreme Learning Machine", Computer and information sciences", Vol. 4, No. 1, 2011, ISSN 1913-8989

[3] Elie Tcheimegni, Dr. Manohar Mareboyana, and Dr.Claude Turner, "Application of the relevance vector machine and support vector machine to clinical data", International Conference on ICT for Africa 2013, Harare, Zimbabwe, February 20 -23,

[4] Dr. Santhosh Baboo, S., and Mrs.Sasikala, S, "Multicategory Classification Using Relevance Vector Machine for Microarray Gene Expression Cancer Diagnosis", International Journal of Advanced Research in computer, Vol.1, No.4, 2010, ISSN 0976-5697.

[5] Liyang Wei, Yongyi Yang, Robert M., and Nishikawa, " A Relevance Vector Machine technique for the automatic detection of clustered microcalcifications", IEEE transactions on Medical Imaging, 2005, Vol. 24, No.10, 2005, Pages 1278-1285, ISSN 0278-0062.

[6] Shovan K. Majumder, Ghosh, N., and Gupta PK."Relevance Vector Machine for Optical Diagnosis of Cancer, Lasers in Surgery and Medicine, 2005, Vol.36, No.4, Pages 323-333.

[7] Michael E Tipping, " Sparse Bayesian learning and Relevance Vector Machine", Journal of Machine learning Research, 2001, Vol. 1, Pages 211-244

[8] Behnood Gholami, " Relevance Vector Machine for Neonate pain intensity assessment using digital imaging", IEEE transactions on Biomedical Engineering,, Vol.57, No.6, Pages 1457-1465, 2010

[9] Alexe G., Dalgin G.S., (et al), "Analysis of breast cancer using Principal Component Analysis and clustering" Journal of Biosciences, Vol. 32, No.5, Pages 1027-1039, 2007

[10] http://www.cs.ccsu.edu/~markov/ccsu_courses/DataMining-3.html, "Data preprocessing"

[11] Punitha, A., Santhanam, T., "Feature Space Optimization in Breast Cancer diagnosis using Linear Vector Quantization", Information technology journal, Vol. 6, No.8, Pages 1258-1263, 2007

[12] Hota, H.S., "Identification of Breast Cancer Using Ensemble of Support Vector Machine and Decision Tree with Reduced Feature Subset", International Journal of Innovative technology and exploring Engineering, Vol.3, No.9, ISSN: 2278-3075, 2014

[13] Guoqiang Sun, Yue Chen(et al), " Day-Ahead Wind Speed Forecasting Using Relevance Vector Machine ", Journal of applied mathematics, Vol. 2014, Article ID 437592, Pages 6, 2014

[14] Kanhu Charan Nayak, Rajesh Kumar Tripathy, Sudha Rani Panda, "Relevance Vector Machine

Based Analyses of MRR and SR of Electro discharge Machining Designed by Response Surface Methodology", International Journal of manufacturing engineering**,** Vol.2013, Article ID 170746, Pages 9, 2013

[15] Pijush Samui, Venkata Ravibabu Mandla (et.al), " Prediction of Rainfall Using Support Vector Machine and Relevance Vector Machine", "Open access e-journal, Earth science India", 2011, Vol. 4, No.IV, Pages 188-200, 2011