Comparative study of Relevance Vector Machine with various machine learning techniques used for detecting breast cancer

B.M.Gayathri
Research scholar
S.D.N.B Vaishnav College for women, Chromepet
Chennai, India
gayathri_bm2003@yahoo.co.in

Dr.C.P.Sumathi
Department of Computer Science
S.D.N.B Vaishnav College for women, Chromepet
Chennai, India
drcpsumathi@gmail.com

Abstract— **Now-a-days breast cancer has become one of the leading cause of cancer death among women. This cancer is caused mostly due to the lifestyle changes, avoiding breast feeding etc. Detecting breast cancer takes long time due to manual diagnosis. Even though there are many diagnostic systems are available still, it takes more time for proper classification. For detecting breast cancer, mostly machine learning techniques are used. This work deals with the comparative study of Relevance vector machine(RVM) which provides Low computational cost while comparing with other machine learning techniques which are used for breast cancer detection .The aim of this work is to compare and explain how RVM is better than other machine learning algorithms for diagnosing breast cancer even the variables are reduced.**

Keywords—Breast Cancer,Relevance Vector Machine,Machine Learning,Naïve Bayes,Analysis of Variance(ANOVA),Linear Discriminant Analysis(LDA),Extreme Learning Machine

I. INTRODUCTION

The most dangerous disease in the world is cancer and one of the cancer that kills the women is breast cancer. Detecting the breast cancer manually takes lot of time and it is very difficult for the physician to classify it. Hence for easy classification, detecting the cancer through various automatic diagnostic techniques is necessary. There are various methods for detecting breast cancer such as biopsy, mammogram, (Magnetic Resonance Imaging) MRI and Ultrasound. Breast cancer happens due to uncontrolled growth of cells and these growths of cells must be stopped as soon as possible by detecting it earlier. There are two classes of tumor, one is benign tumor and the other is malignant, in which benign tumor is non-cancerous and the latter is cancerous. Many researchers are still performing research for developing a proper diagnostic system for detecting the tumor as early as possible and also in an easier way, so that the treatment can be started earlier and the rate of survivability can be increased. For developing the computerized diagnostic system, machine learning algorithms plays an important role. There are many machine learning algorithms which are used to classify the tumor easily and in effective way. This work deals with the comparative analysis of Relevance vector machine (RVM) with various machine learning algorithms which are used to detect breast cancer and also the number of variables used in it.

II. MACHINE LEARNING ALGORITHMS

Machine learning is one of the branch of computer science, which is useful in pattern recognition and computational learning theory of artificial intelligence. Machine learning can be used to construct algorithms which can learn and make relationship with mathematics and also with computational statistics. By using machine learning, the user can create new algorithms which can learn and predict the data without explicitly being programmed.

A. Categories of Machine Learning:

There are three different categories of Machine learning. They are supervised learning, unsupervised learning and Reinforcement learning. Each and every category is used based on the requirement.

1) *Supervised Learning:* If there is a proper structure of inputs passed to the system which gives outputs based on the pattern which is already stored is known as supervised learning. In this proper label names are given.
2) *Unsupervised Learning:* If there is no proper structure or labels and if the system has to discover the pattern of its own, then it is an unsupervised learning.
3) *`Reinforcement Learning:* If the system interacts with dynamic environment then it is reinforcement learning. For ex: if the user plays a game in a system, with the system as opponent.

Other categories are Semi supervised learning and transduction. In these Semi Supervised consists of missing targets and transduction consists of problem instances which is passed during the learning time, except some of the parts of the targets are missing.

III. LIST OF MACHINE LEARNING ALGORITHMS

Some of the Machine learning algorithms are Artificial Neural Network (ANN),Naïve Bayes, Support Vector Machine(SVM),Decision tree(DT),Relevance vector machine(RVM) etc. Many researchers have done research in

978-1-5090-0612-0/16/$31.00 ©2016 IEEE

breast cancer by using various dataset such as by using Mammogram images as dataset, SEER dataset, Wisconsin dataset and also dataset from various hospitals. In all these dataset, the authors reduces the features by using various feature selection methods, but still the number of variables exceed more than 5. But the author B.M.Gayathri and C.P.Sumathi has reduced the feature below 5 by using Wisconsin Original dataset as database and done research using various machine learning algorithms and the accuracy of those works does not show much difference than those researches which has variables more than 5, with Wisconsin original dataset as database. The following section compares number of variables with the accuracy of work of B.M.Gayathri and C.P.Sumathi with other works which uses same dataset for classification. The table 3 shows the list of articles which uses Wisconsin original dataset for detecting breast cancer and also the number of variables used and its accuracy. Among these different algorithms RVM shows better accuracy with less attributes.

IV. RELEVANCE VECTOR MACHINE

It is one of the machine learning technique which uses Bayesian inference to get a proper solution for probabilistic classification. The RVM is a combination of Gaussian process with covariance function. It is specified as

$$k(x,x) = \sum_{j=1}^{N} 1 \frac{1}{\alpha} \varphi(x, x_j) \varphi(x', x_j) \quad (1)$$

where φ is the kernel function. Variable x represents the input vectors and N shows the target. The most common kernel functions are Fisher's kernel, linear kernel, RBF kernel etc. and One of the advantages of the relevance vector machines is that there is probabilistic predictions, using arbitrary kernel functions and does not require regularization parameter setting[1].and also it provides inferences with low computational cost[2]. The relevance vector machine is not only used in medical classification but also it is used in other areas like remote sensing, weather forecasting etc.

A. Literature Review on Relevance Vector Machine(RVM)

B.M.Gayathri and C.P.Sumathi[1] used RVM for detecting breast cancer. The authors have created a user friendly environment for diagnosis. In this work authors compares the RVM with SVM and the accuracy of RVM shows better than SVM. The dataset used is Wisconsin original dataset. The dataset consist of 12 variables and they are reduced to four variables by using feature selection algorithms. In this work the author shows the difference between the variables and their accuracy and it does not show much difference and the accuracy was up to 96%.

A.Bharathi and Ananda kumar[2] developed a machine learning approaches for classifying cancer, so that a proper algorithm can be identified by using the accuracy of these approaches. In this work the lymphoma and Leukemia dataset were used and the features were selected by using Analysis of Variance (ANOVA) model. After selecting the features they were applied to Fast Support Vector Machine Learning (FSVML), Fast Extreme Learning Machine Learning (FELML) and Relevance Vector Machine Learning (RVMMLML). From these machine learning algorithm the author found that RVM shows better accuracy of more than 95%. Elie Tcheimegni[3] created a Kernel based Relevance vector machine for classifying cancer diseases. In this work the author has developed a hierarchical Bayesian model using sigmoid kernel, Radial basis function (RBF). By using the data driven approach the author tries to classify various diseases by using clinical data. The author compares the RVM with SVM, in which RVM shows good accuracy than SVM of more than 90%

A.Bharathi and A.M.Natarajan[4] has done cancer classification using SVM and RVM based on ANOVA. The main aim of this work was to select smallest genes from microarray data to classify cancer in an effective way. For feature selection ANOVA was used. Features were selected based on ranking scheme. The selected genes were implemented in SVM and RVM, in which RVM shows better accuracy of more than 93%.

The dataset used was Lymphoma dataset and Leukemia dataset. Dr.S.SanthoshBaboo and Mrs.S.Sasikala[5] proposed a technique for classifying cancer by using microarray gene expression dataset. Relevance vector machine (RVM) was used for diagnosing cancer and the performance of RVM was evaluated by using the Lymphoma and Leukemia dataset. From the analysis, it was found that RVM works better than Support vector machine (SVM) and Extreme Learning Machine (ELM).The accuracy of RVM was more than 95% than SVM and ELM.

From the survey, it is found that even though RVM is used for cancer classification, it is not directly applied for breast cancer detection. Authors B.M.Gayathri and Dr.C.P.Sumathi has applied RVM exclusively for detecting breast cancer by using Wisconsin original dataset which consist of 12 attributes and 699 sample records. The dataset was retrieved from UCI repository.

The features were reduced by using feature selection algorithms such as Linear Discriminant algorithm (LDA), so that it saves time without entering more data in the system and the system is user friendly. The main advantage of using RVM is it reduces the computational cost. Hence the accuracy of RVM is better than any other machine learning algorithm with reduced variables.

B. Implementation 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. The following table 1 and 2 shows the classification results in percentage and performance

evaluation of RVM and figure 1 shows the sample interface of the proposed model.

TABLE I. CONFUSION MATRIX FOR RVM CLASSIFIER

Dataset	300	
Condition positive	98% (TP)	0.04% (FN)
Condition Negative	0.02% (FP)	96% (TN)

TABLE II. PERFORMANCE EVALUATION FOR RVM CLASSIFIER

Dataset	**Algorithm**	**Sensitivity**	**Specificity**	**Accuracy**
300	RVM	98%	98%	97%

C. Other Related works of Breast cancer usingWisconsin original dataset for various machine learning algorithms

The table 3 shows the list of articles of different machine learning algorithms used for detecting breast cancer using Wisconsin original dataset for classification, which compares variables, dataset and their accuracy. From this comparison, we can find that Breast cancer detection using RVM shows 97% of accuracy even when the variables were reduced.

TABLE III. MACHINE LEARNING METHODS AND FEATURES USED FOR BREAST CANCER DETECTION USING WISCONSIN ORIGINAL DATASET

Article Name	**Authors/Year**	**Method**	**No. of Features**	**Accuracy**
Weighted naïve bayes classifier:- Predictive model for breast cancer detection[6]	S.Kharya Sunita soni (2016)	Naïve Bayes	9	92%
An Application of Classification Techniques on Breast Cancer Prognosis[7]	Sandeep Chaurasia, Neha Chourasia, Prasun Chakrabarthi. (2016)	Neural networks	9	96.14%
An approach for breast cancer diagnosis classification using neural network.(Wisconsin original dataset[8]	Htet Thazin Tike Thein (2015)	Neural Networks	9	99.97%
Mining Big data: Breast cancer prediction using DT-SVM Hybrid Model[9]	K.Sivakami (2015)	Decision tree, Support Vector Machine.	9	95%
Breast cancer risk detection using RVM.[10]	B.M.Gayathri ,C.P.Sumathi (2015)	Relevance Vector Machine	4	97%
Mamdani fuzzy inference system for breast cancer risk detection.[11]	B.M.Gayathri ,C.P.Sumathi (2015)	Fuzzy inference system	4	93%
Naïve bayes classifier: A probabilistic detection model for breast cancer[12] .	S.Kharya,Sunita Soni.(2014)	Naïve Bayes	9	93%

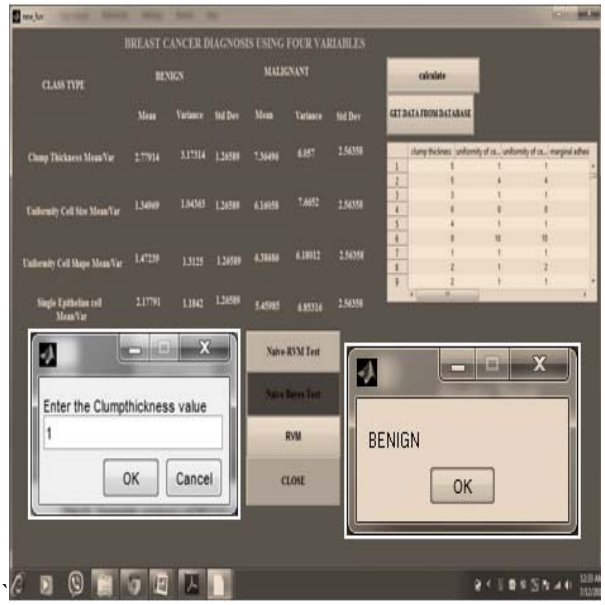

Fig 1 Sample output of RVM classifier

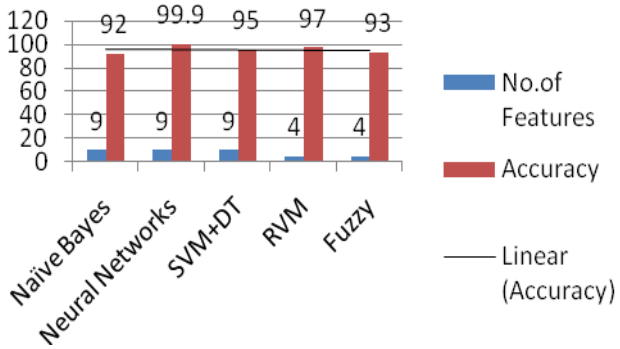

Fig 2 Accuracy difference between Machine Learning Algorithms

In the figure 2 the trend line shows that there is no much accuracy difference, even the features are reduced. The table 4 shows the use of machine learning algorithms in other medical diagnosis.

V. LITERATURE REVIEW ON BREAST CANCER DETECTION USING OTHER DATASETS.

Mandeep Rana [13] et al. have done the comparative study of certain machine learning techniques such as Support vector machine (SVM), Logistic regression, K-Nearest Neighborhood (KNN) and Naïve Bayes for predicting the recurrence of breast cancer and also diagnosing breast cancer using these techniques. The dataset used for this study is taken from UCI repository (Wisconsin prognostic breast cancer dataset) and all the 32 variables were used in this work and the accuracy of breast cancer detection was 95.6% and 68% for recurrence and non recurrence of breast cancer. E.Venkatesan and T.Velmurugan[14] have done the performance analysis of different classification algorithms such as j48, AD tree (Alternative Decision tree) and Best first tree (B+ tree). The dataset is taken from Swami Vivekananda diagnostic centre hospital at Chennai. It consists of totally 220 patient records and nine attributes are used for analysis. Out of the entire 4 algorithms j48 algorithm shows the result of 99%.

Konstantina kourou[15] et al. has discussed about various predictive model of recent machine learning approaches used in finding the cancer progression. In this work the author has made review on various publications which are relevant to ML. Each and every type of paper and its classification differs based on the dataset and its variables. Mostly those papers consist of mammographic features of up to 14 variables and the accuracy was up to 83% for mammographic data and 71% for other datasets. Ahmad LG[16] et al. analyzes 3 different algorithms such as Decision tree (DT), Artificial Neural Network (ANN) and Support Vector Machine (SVM) out of which SVM shows higher accuracy than other 2 algorithms. The database used in this work was taken from Iranian centre for breast cancer (ICBC). Totally 8 predictor variables were used and SVM gave the accuracy of up to 95%.

H.S. Hota[17] has developed a model based on assembling SVM and C5.0 for identifying breast cancer. The dataset used in this work was taken from Wisconsin prognostic dataset which consist of 32 features. For performing variable reduction Rank based feature selection was used. The performance of Radial Basis Function shows that for 5 features the accuracy was 92.59%.

Cuong Nguyen[18] et al. created a computer aided diagnostic system to classify malignant and benign tumor. In this work

the Backward Elimination (BE) approach was used for feature selection in combination with Random forest tree is used. The dataset was taken from Wisconsin prognostic database. It consists of totally 33 variables and it was reduced to 17 to 18 variables and the accuracy of this hybridized algorithm shows around 99%.

The table 5 shows some of sample articles of RVM in other branches such as weather forecasting, finding heart diseases, optical cancer.

TABLE IV. MACHINE LEARNING ALGORITHMS IN OTHER MEDICAL DIAGNOSIS

Title	Journal	Author	Application
ECG Arrhythmia Detection and Classification Using Relevance Vector Machine.[19]	International conference on modeling optimization and computing.	Gayathri.S , M. Suchetha , V.Latha (2012)	Heart disease
Detecting lung nodules in chest CT images with Ensemble Relevance vector machine.[20]	Applied Mechanics and Materials,	Chao Dong, Lian-fang Tian, Jing Zhang and Bin Li (2012)	Heart disease
Classification of Electrocardiogram signals with Extreme Learning Machine and Relevance Vector machine[21]	International Journal of computer science Issues	S.Karpagachelvi, M.Sivakumar, Dr.M.Arthanari. (2011)	Heart disease
Classification of Electrocardiogram signals with Extreme Learning Machine and Relevance Vector machine[20]	International Journal of computer science Issues	S.Karpagachelvi, M.Sivakumar, Dr.M.Arthanari. (2011)	Heart disease
Relevance vector machine for optical cancer diagnosi[s21]	Lasers in surgery and medicine	S.K.Majumder,Gosh N.Gupta PK(2005)	Optical cancer

TABLE V. RELEVANCE VECTOR MACHINE IN OTHER BRANCHES

Article Name	Journal Name	Author /Year	Area
Wavelet-multivariate relevance vector machine hybrid model for forecasting daily evapotranspiration[22]	Stochastic Environmental research and risk assessment. ,Springer	Roula Bachour, Inga Maslova, Andres Ticlavilca, Wynn R. Walker , Mac McKee (2015)	Weather Forecasting
The Wavelet Transform with best decomposition Level and Relevant Vector Machine Based Approach for Chaotic Time Series Forecasting[23]	3rd International Conference on Mechatronics, Robotics and Automation (ICMRA)	WANG Xiao-LU1,LIU Jian1,LU Jian-Jun (2015)	Weather Forecasting
Relevance vector machines as a tool for forecasting geomagnetic storms during years 1996–2007.[24]	Journal of Atmospheric and Solar-Terrestrial Physics	T.Andriyas,S.Andriyas (2015)	Weather Forecasting
Prediction of Rainfall Using Support Vector Machine and Relevance Vector Machine.[25]	Earth science India	Pijush Samui , Venkata Ravibabu Mandla , Arun Krishna2 and Tarun Teja(2011)	Weather forecasting

VI. Discussion & Conclusion

This work is the comparative study of RVM with various ML algorithms, to show that RVM classifies better than other ML algorithms even when the variables are reduced. From table 3 it is found that RVM shows better accuracy than any other algorithms and in the related works of RVM, it is seen that RVM is not mostly used for detecting breast cancer by using Wisconsin Original dataset. RVM is generally used for detecting cancer by using the benchmark dataset of Lymphoma and Leukemia. Hence authors B.M.Gayathri and Dr.C.P.Sumathi[1] have used Wisconsin original dataset for detecting breast cancer which shows good result than any other Machine learning (ML) algorithms. Table 5 shows the uses of RVM in other branches also. As a future work RVM can be combined with other ML algorithms so that it can be fine tuned to improve the accuracy.

Acknowledgment

I take this opportunity to thank my guide Dr.C.P.Sumathi, Professor, Head of the department of Computer Science for giving me valuable support and guidance. I also thank my husband Mr.S.Muralidharan for encouragement given to me for completion of this paper.

References

[1] B.M.Gayathri and C.P.Sumathi, "Breast cancer risk detection using RVM", International journal of applied engineering research,Vol.10 No.5, pp.3979 -3982, September 2015.

[2] A.Bharathi and K.Anandakumar."Cancer classification using relevance vector machine learning",Journal of medical imaging and health informatics Volume 5, No. 3, pp. 630-634, June 2015 .

[3] Elie Tcheimegni,"Kernel based relevance vector machine for cancer classification of diseases",Bowie stateuniversity,2013,202,publication Number:3558597.

[4] A.Bharathi and A.M.Natarajan,"Cancer classification using support vector machines and relevance vector machine based on Analysis of Variance features", Journal of computer science, Vol.7 No.9, pp: 1393-1399, September 2011.

[5] S.Santhosh Baboo and S.Sasikala, "Multicategory classification using Relevance vector machine for microarray gene expression cancer diagnosis", International Journal of advanced research in computer science, Vol.1, No.4 pp., 462-473, December 2010.

[6] S Kharya and S Soni,"Weighted Naïve Bayes classifier –Predictive model for breast cancer detection", International Journal of computer applications,Vol.133,No.9,pp.32-37,January 2016.

[7] S Chaurasia,N Chourasia and P Chakrabarthi,"An application of classification techniques on breast cancer prognosis",Vol.59,No.3,pp.6-10,December 2012.

[8] Htet Thazin Tike Thein,"An approach for breast cancer diagnosis classification using neural networks,"Advanced Computing.An International Journal,Vol.6,No.1,pp.1-11,January 2015.

[9] K.Sivakami, Mining Big data: Breast cancer prediction using DT-SVM Hybrid Model, International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-1, Issue-5 ,pp.418-429,August 2015.

[10] B.M.Gayathri and C.P.Sumathi,"Breast cancer risk detection using RVM", International Journal of Applied Engineering Research(IJAER),Vol.10,No.17,pp.37717-37723, August 2015.

[11] B.M.Gayathri and C.P.Sumathi,"Mamdani fuzzy inference system for breast cancer risk detection", 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC),pp.1-6, December 2015.

[12] S Kharya and S soni,"Naïve Bayes classifier:A probabilistic detection model for breast cancer", International Journal of Computer Applications ,Vol.92.No.10,pp.26-32,April 2014.

[13] Mandeep Rana,"Breast cancer diagnosis and recurrence prediction using machine learning techniques", International journal of research in Engineering and Technology, Vol.4, No.4, pp.372-376, April 2015.

[14] E.Venkatesan and T.Velmurugan, Performance analysis of decision tree algorithms for breast cancer classification, Indian journal of science and technology,Vol.8,No.29,pp.1-8,November 2015.

[15] Konstantina Kourou and Themis P. Exarchos, "Machine learning application in cancer prognosis and prediction", Computational and structural biotechnology journal, Vol.13, pp.8-17, November-2015.

[16] L.G.Ahmad, A.T.Eshlagh,A.Poorebrahimi,M.Ebrahimi and A.R.Razavi "Using three machine learning techniques for predicting breast cancer recurrence", Journal of Health and medical informatics,Vol.4,No.2,pp.1-3,April 2013.

[17] H.S.Hota," Identification of breast cancer using Ensemble of support vector machine and decision tree and reduced feature subset", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-3, Issue-9,pp-99-102, February 2014.

[18] Cuong Nguyen, Yong Wang and Nam Nguyen, "Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic", Journal of Biomedical science and Engineering, Vol.6, pp.551-560, May 2013.

[19] S.Gayathri, Sucetha.V and Latha," ECG arrhythmia detection and classification using relevance vector machine", International conference on modeling optimization and computing,Vol.38,PP.1333-1339,June 2012.

[20] S.Karpagachelvi, M.Sivakumar and M.Arthanari, "Classification of electrocardiogram signals with extreme learning machine and relevance vector machine", International Journal of computer science Issues,Vol.8,No.1,pp.338-345,January 2011.

[21] S.K.Majumder,N.Gosh and P.K.Gupta, Relevance vector machine for optical cancer diagnosis., Lasers in surgery and medicine,Vol.36,No.4,pp.323-333,April 2005.

[22] Roula Bachour,Inga Maslova, M.Andres,R. Ticlavilca, Wynn Walker and Mac McKee, "Wavelet-multivariate relevance vector machine hybrid model for forecasting daily evapotranspiration", Stochastic Environmental research and risk assessment.,Springer,Vol.30,no.1,pp.103-117,January 2016.

[23] Wang Xiao,Lu,Liu Jian and Lu Jian-Jun, The Wavelet transform with best decomposition Level and Relevant Vector Machine Based Approach for Chaotic Time Series Forecasting, 3rd International Conference on Mechatronics, Robotics and Automation (ICMRA),pp-947-953,August 2015.

[24] T.Andriyas and S.Andriyas,"Relevance vector machines as a tool for forecasting geomagnetic storms during years 1996–2007", Journal of Atmospheric and Solar-Terrestrial Physics, Volumes 125–126, Pages 10–20,April 2015.

[25] Pijush Samui , Venkata Ravibabu Mandla , Arun Krishna and Tarun Teja, "Prediction of rainfall using support vector machine and relevance vector machine", Earth science India,Vol.4,No.IV,pp.188-200,October 2011.