(IJRST) 2019, Vol. No. 9, Issue No. I, Jan-Mar **e-ISSN: 2249-0604, p-ISSN: 2454-180X**

MACHINE LEARNING AND CANCER: EMPLOYABILITY OF RANDOM FOREST, SUPPORT VECTOR MACHINE, BAYESIAN NETWORK ALGORITHMIC TOOLS IN THE EARLY DETECTION OF BREAST CANCER

ADARSH DHIMAN

Delhi Public School, Ambala

ABSTRACT

One of the significant widespread disease these days for women is breast cancer. Right treatment and early detection is a significant step to take to prevent this disease. However, it is not easy, because of a few vulnerabilities and recognition of mammograms.

Machine Learning (ML) strategies can be utilised to create apparatuses for doctors that can be utilised as a compelling system for first location and conclusion of breast cancer growth which will significantly improve the survival rate of patients.
This paper analyses about three of the most prominent ML strategies generally utilised for breast cancer disease location and finding,
in particular, Support Vector Machine (SVM), Random Forest (RF) and Bayesian Networks (BN).
The Wisconsin breast cancer malignancy informational collection was utilised as a preparation set to assess and think about the execution of the three ML classifiers as far as critical parameters, for example, exactness, review, accuracy and zone of ROC. The outcomes got in this paper give an outline of the condition of craftsmanship ML methods for bosom growth identification.

Keywords: Breast cancer, Machine Learning, Random Forest, Support Vector Machine, ROC, Bayesian Networks

I. INTRODUCTION

ML procedures have been generally utilised in the medicinal field and have filled in as a value indicative device that helps doctors in examining the accessible information and additionally structuring therapeutic master frameworks. This paper exhibited three of the most prevalent ML systems usually utilised for bosom malignancy recognition and conclusion, in particular, Support Vector Machine (SVM), Random Forest (RF) and Bayesian Networks (BN). The notable highlights and system of every one of the three ML strategies were depicted. Execution examination of the researched systems has been completedutilising the Original Wisconsin Breast Cancer Dataset. Reenactment results acquired has demonstrated that characterisation execution differs dependent on the strategy that is chosen. Results have demonstrated that SVMs have the most astounding execution as far as exactness, specificity, and accuracy. Notwithstanding, RFs have the most astounding likelihood of accurately ordering tumour.

The most dangerous disease in the world is cancer,and one ofcancer that kills women is breast cancer. Detecting thebreast cancer manually takes much time,and it is tough for the physician to classify it. Hence for easy classification,detecting cancer through various automatic diagnostictechniques is necessary. There are various methods fordetecting breast cancer such as biopsy, mammogram,(Magnetic Resonance Imaging) MRI and Ultrasound. Breastcancer happens due to uncontrolled growth of cells,and thesegrowths of cells must be stopped as soon as possible bydetecting it earlier. There are two classes of atumour, one isa benigntumour,and the other is malignant, in which a benigntumour is non-cancerous,and the latter is cancerous. Manyresearchers are still performing research for developing aproper diagnostic system for

detecting the tumour as early aspossible and also in a more natural way so that the treatment can bestarted earlier and the rate of survivability can be increased.

For developing the computerised diagnostic system, machinelearning algorithms plays an important role. There are manymachine learning algorithms which are used to classify a tumour easily and ineffective way. This work deals with thecomparative study of Relevance vector machine (RVM)with various machine learning algorithms which are used to detect breast cancer and also the number of variables used init.

1.1. Breast Cancer

The breast is milk creation organ in a female having lobules and areola associated through pipes. A breasttumour is the most regular type of disease representing around one-fourth of unfortunate passings,and late discovery put ladies at higher danger of death — around 70-80% of breast growths created in lobules while channels a tumour involved just of around 20% of breast malignancy cases.

The breast cancer can be of three types as shown in figure 1, below.

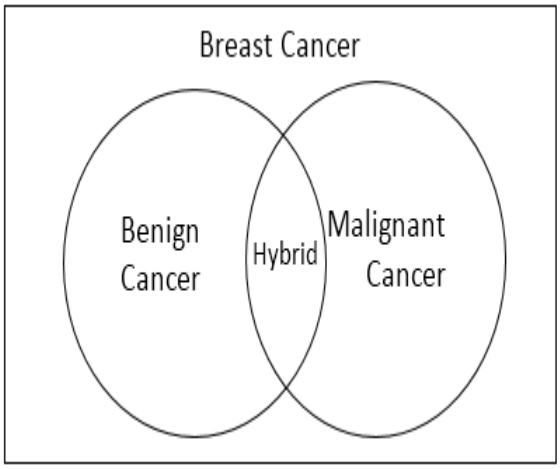
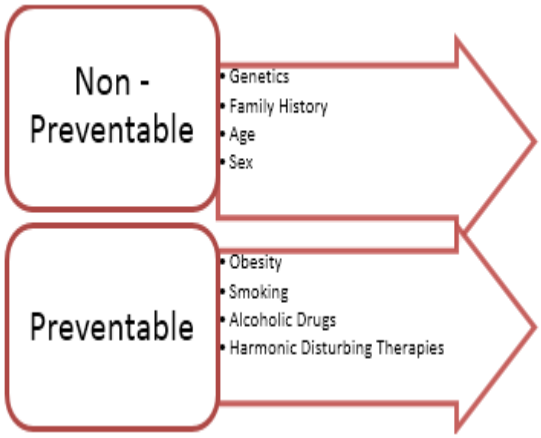


Figure 1: Different types of breast cancers

However, the breast tumour is an unpredictable infection and couldn't be ascribed to a single reason instead there are different hazard factors, which add to the plausibility of affliction. These dangers elements might be grouped into two classes as appeared in figure 2, underneath.



The breast cancer can primarily be identified with physicalsymptoms. Such identification paves the way for confirmation teststo ensure favourable prognosis. Figure 3, enlists the main symptoms ofbreast cancer.

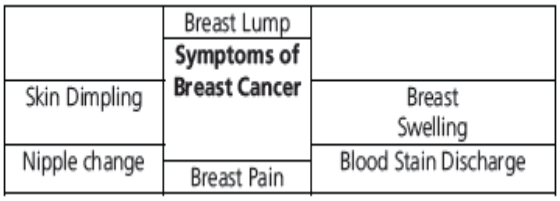


Figure 3: Symptoms of Breast Cancer

1.2 MACHINE LEARNING ALGORITHMS

Machine learning is one of the branch of computer science, which is useful in pattern recognition and computationallearning theory of artificial intelligence. Machine learning canbe used to construct algorithms which can learn and makea relationship with mathematics and also with computationalstatistics. By using machine learning, the user can create newalgorithms 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 have supervised learning, unsupervised learning andReinforcement learning. Every category is used based on the requirement.

1) Supervised Learning: If there is a proper structure ofinputs passed to the system which gives outputsbased on the pattern which is already stored is knownas supervised learning. In this proper label, names are given.

2) Unsupervised Learning:If there is no properstructure or labels and if the system has to discoverthe pattern of its own, then it is unsupervised learning.

3) `Reinforcement Learning:If the system interacts withthe dynamicenvironment, then it is reinforcement learning.For ex: if the user plays a game in a system, with thesystem as an opponent.

Other categories are Semi-supervised learning andtransduction. In these Semi-Supervised consists of missingtargets and transduction consists of problem instances which arepassed during the learning time, except some of the parts ofthe targets are missing.

II. LITERATURE REVIEW ON BREAST CANCER DETECTIONUSING OTHER DATASETS.

Mandeep Rana[13] et al. have done the comparative study ofspecific machine learning techniques such as Support vectormachine (SVM), Logistic regression, K-Nearest Neighborhood(KNN) and Naïve Bayes for predicting the recurrence of breastcancer and also diagnosing breast cancer using thesetechniques. The dataset used for this study is taken from UCIrepository (Wisconsin prognostic breast cancer dataset),and allthe 32 variables were used in this work,and the accuracy ofbreast cancer detection was 95.6% and 68% for recurrence andnon-recurrence of breast cancer. E.Venkatesan andT.Velmurugan14 have done the performance analysis ofdifferent classification algorithms such as j48, AD tree(Alternative Decision tree) and Best first tree (B+ tree). Thedataset is taken from Swami Vivekananda diagnostic centrehospital at Chennai. It consists of totally 220 patient records,and nine attributes are used for analysis. Out of the entire fouralgorithms, j48 algorithm shows the result of 99%.

Konstantina kourou1[15] et al. has discussed a variouspredictive model of recent machine learning approaches usedin finding the cancer progression. In this work, the author hasmade a reviewof various publications which are relevant to ML.Every type of paper and its classification differs basedon the dataset and its variables. Mostly those papers consist ofmammographic features of up to 14 variables,and the accuracywas up to 83% for mammographic data and 71% for otherdatasets. Ahmad LG16 et al. analysesthree different algorithmssuch as Decision tree (DT), Artificial Neural Network (ANN)and Support Vector Machine (SVM) out of which SVM showshigher accuracy than other two algorithms. The database used inthis work was taken from Iranian centre for breast cancer(ICBC). Totally eight predictor variables were used,and SVM gavethe accuracy of up to 95%.

H.S. Hota[17] has developed a model based on assemblingSVM and C5.0 for identifying breast cancer. The dataset usedin this work was taken from Wisconsin prognostic datasetwhich consists of 32 features. For performing variablereduction Rank based feature selection was used. Theperformance of Radial Basis Function shows that for fivefeatures the accuracy was 92.59%.

Cuong Nguyen[18] et al. created a computer-aided diagnosticsystem to classify a malignant and benign tumour. In this workthe Backward Elimination (BE) approach was used forfeature selection in combination with Random forest tree isused. The dataset was taken from Wisconsin prognosticdatabase. It consists of totally 33 variables,and it was reducedto 17 to 18 variables,and the accuracy of this hybridisedalgorithm shows around 99%.

Table2 shows some of the sample articles of RVM inother branches such as weather forecasting, finding heartdiseases, optical cancer.

Table I. 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[21]	Lasers in surgery and medicine	S.K.Majumder,Gosh N.Gupta PK(2005)	Optical cancer

Table 2. 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

III. FUTURE PROSPECTS

This work is the comparative study of RVM with variousML algorithms, to show that RVM classifies better than otherML algorithms even when the variables are reduced. Fromtable 3 it is found that RVM shows better accuracy than anyother algorithms and in the related works of RVM, it is seenthat RVM is not mostly used for detecting breast cancer byusing Wisconsin Original dataset. RVM is generally used fordetecting cancer by using the benchmark dataset ofLymphoma and Leukemia. Hence, authors, B.M.Gayathri andDr.C.P.Sumathi1 have used Wisconsin original dataset fordetecting breast cancer which shows good result than anyother Machine learning (ML) algorithms. Table 5 shows theuses of RVM in other branches also. As a future work, RVMcan be combined with other ML algorithms so that it can befine-tuned to improve the accuracy.