(IJIASE) 2018, Vol. No. 4, Jan-Dec **e-ISSN: 2454-9258, p-ISSN: 2454-809X**

EXPLORING THE APPLICATION OF MACHINE LEARNING STRATEGIES AND TOOLS IN THE DIAGNOSIS AND PROGNOSIS OF BREAST CANCER

Raj Verma

ABSTRACT

One of the major widespread disease in these days for women is breast cancer. right treatment and early detection is a major step to take to prevent this disease. however, it's not easy, because of a few vulnerabilities and recognition of mammograms.

Machine Learning (ML) strategies can be utilized to create apparatuses for doctors that can be utilized as a compelling system for early 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 utilized 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 utilized as a preparation set to assess and think about the execution of the three ML classifiers as far as key 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 utilized 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 normally utilized for bosom malignancy recognition and conclusion, in particular, Support Vector Machine (SVM), Random Forest (RF) and Bayesian Networks (BN). The fundamental highlights and system of every one of the three ML strategies were depicted. Execution examination of the researched systems has been completed utilizing the Original Wisconsin Breast Cancer Dataset. Re-enactment results acquired has demonstrated that characterization 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 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.

1.1. Breast Cancer

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

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

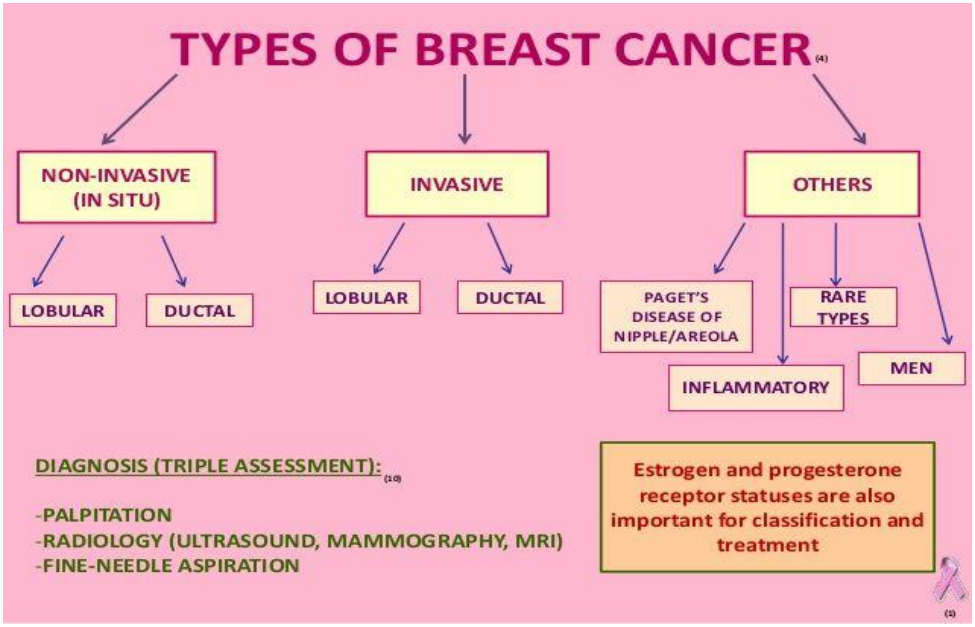
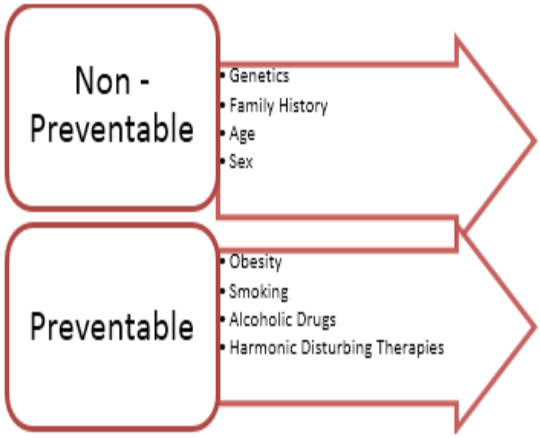


Figure 1: Different types of breast cancers

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



The breast cancer can primarily be identified with physical symptoms. Such identification paves a way for confirmation tests to ensure timely prognosis. Figure 3, enlists the main symptoms of breast cancer.

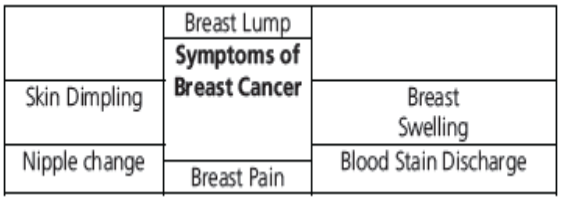


Figure 3: Symptoms of Breast Cancer

1.2 MACHINE LEARNING ALGORITHMS

Machine learning is one of the branch of computer science, which is useful in pattern recognition and 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

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 I. Machine Learning Algorithms In Other Medical Diagnosis

II. 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.Sumathi1 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.

REFERENCES

[1] M. R. Al-Hadidi, A. Alarabeyyat and M. Alhanahnah, " Breast Cancer Detection Using K-Nearest Neighbor Machine Learning Algorithm," 2016 9th International Conference on Developments in eSystems Engineering (DeSE), Liverpool, 2016, pp. 35-39.

[2] C. Deng and M. Perkowski, "A Novel Weighted Hierarchical Adaptive Voting Ensemble Machine Learning Method for Breast Cancer Detection," 2015 IEEE International Symposium on Multiple-Valued Logic, Waterloo, ON, 2015, pp. 115-120.

[3] A. Qasem et al., "Breast cancer mass localization based on machine learning," 2014 IEEE 10th International Colloquium on Signal Processing and its Applications, Kuala Lumpur, 2014, pp. 31-36.

[4] A. Osareh and B. Shadgar, "Machine learning techniques to diagnose breast cancer," 2010 5th International Symposium on Health Informatics and Bioinformatics, Antalya, 2010, pp. 114-120.

[5] J. A. Bhat, V. George and B. Malik, "Cloud Computing with Machine Learning Could Help Us in the Early Diagnosis of Breast Cancer," 2015 Second International Conference on Advances in Computing and Communication Engineering, Dehradun, 2015, pp. 644-648.

[6] B. M. Gayathri and C. P. Sumathi, "Comparative study of relevance vector machine with various machine learning techniques used for detecting breast cancer," 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, 2016, pp. 1-5.

[7] H. R. Mhaske and D. A. Phalke, "Melanoma skin cancer detection and classification based on supervised and unsupervised learning," 2013 International conference on Circuits, Controls and Communications (CCUBE), Bengaluru, 2013, pp. 1-5.

[8] S. Aruna, S. P. Rajagopalan and L. V. Nandakishore, "An algorithm proposed for Semi-Supervised learning in cancer detection," International Conference on Sustainable Energy and Intelligent Systems (SEISCON 2011), Chennai, 2011, pp. 860-864.

[9] Y. Tsehay et al., "Biopsy-guided learning with deep convolutional neural networks for Prostate Cancer detection on multiparametric MRI," 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, 2017, pp. 642-645.

[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] MandeepRana,"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] RoulaBachour,IngaMaslova, 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] PijushSamui ,VenkataRavibabuMandla , Arun Krishna and TarunTeja, "Prediction of rainfall using support vector machine and relevance vector machine", Earth science India,Vol.4,No.IV,pp.188-200,October 2011.