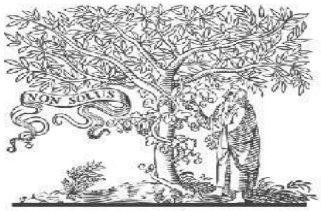
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

IJIEMR Transactions, online available on 21st jan 2021. Link :http://www.ijiemr.org/downloads.php?vol=Volume-10&issue=Issue 01

10.48047/IJIEMR/V10/ISSUE 01/70

Title **OPTIMIZED PREDICTION METHOD FOR BREAST CANCER DETECTION USING MACHINE LEARNING WITH DIFFERENT FEATURE EXTRACTING STRATEGIES**

Volume 10, ISSUE 01, Pages: 349-355

Paper Authors **A.HEMANTH KUMAR, T.SANGEETHA, R.V.SHANMUKHI, T.VINEETHA SANJANA, P.SAHITHI, S.HARSHINI**

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

OPTIMIZED PREDICTION METHOD FOR BREAST CANCER DETECTION USING MACHINE LEARNING WITH DIFFERENT FEATURE EXTRACTING STRATEGIES

[1]A.HEMANTH KUMAR, [2]T.SANGEETHA, [3]R.V.SHANMUKHI, [4]T.VINEETHA SANJANA, [5]P.SAHITHI, [6]S.HARSHINI

[1]Assistant Professor, Department of ECE, Sree Venkateswara College of Engineering, Northrajupalem(VI), Kodavaluru(M), Nellore (DT), Andhra Pradesh, India.

[2,3,4,5,6]B.Tech Scholars, Department of ECE, Sree Venkateswara College of Engineering, Northrajupalem(VI), Kodavaluru(M), Nellore (DT), Andhra Pradesh, India.

ABSTRACT: Breast Cancer is one of the most severe diseases that is faced by women leading nowhere other than increased death rates in society. The survival rate increases on detecting breast cancer early as better treatment can be provided. So, correct early detection can assist lower breast most cancers mortality prices. Therefore an Automated breast cancer diagnosis has been developed to reduce the time taken for diagnosis and decreases the spread of cancer. This paper presents, Optimized Prediction method for Breast Cancer detection using Machine Learning with different Feature Extracting strategies. In this paper, Three machine learning algorithms are proposed namely Support Vector Machine (SVM), Logistic Regression, and Decision Tree (DT) algorithm on the Wisconsin (Diagnostic) Data Set. Different features are proposed such as texture, morphological entropy based, scale invariant feature transform (SIFT), and elliptic Fourier descriptors (EFDs). These features are passed as input to ML classifiers. Among all algorithms, SVM performed better with the accuracy of about 98.4%.

KEYWORDS: ML classifiers, SV, DT, LR, Feature Extracting strategies, Breast Cancer .

I. INTRODUCTION

Breast Cancer is one of the main reasons for the demise of women. It is the second dangerous cancer after lung cancer. Like any other Cancer, Breast Cancer begins when healthy cells change and starts to grow in a disordered manner, forming a mass of cells called a tumor [1]. A tumor can be benign or cancerous. A cancerous tumor is called malignant and they can grow and spread to other parts of the body of the patient.

A benign tumor is a tumor that can grow in a particular part of the body but it does not spread to the other parts of the body [2].

Approximately 29.9% of mortality from cancer in women is due to breast cancer. The incidence of this disease is lower in developing countries than in developed countries, about 10% of women with breast cancer in Western countries. However, millions of people over the world are suffering from this life-threatening disease mortality. With the exponential growth and major advancement in the field of medical-science, breast cancer is curable if the threat is detected at early stages otherwise it can cause severe health issues and can cost life of patient.

Many challenges are faced by a woman fighting against breast cancer which includes pain during radiation therapy and chemotherapy, the huge cost in terms of money that comes with it and, much more so it becomes very essential to predict breast cancer as soon as possible. There are pieces of evidence (as suggested by WHO) that state that women who consume drinks that contain alcohol, have above average birth-weight and above-average height attained when they are an adult are more at risk to develop breast cancer. It is also suggested

that physically active women, eat whole grains, vegetables, fruits, and consume less red meat, alcoholic drinks and, sugarsweetened drinks are at lower risk of developing breast cancer.

The early diagnosis and detection of breast cancer can decrease the death rate and provide means for prompt treatment [3]. The breast cancer is diagnosed and detected using combination of approaches including imaging, physical examination and biopsy. The imaging techniques used to detect the breast cancer are mammography and ultrasound. In this method X-ray is used to create images of the breast. These images are known as mammograms. Radiologists who are trained in to read mammograms to detect sign of breast cancer. The effectively of screening process can rely on radiologists' explanation. The patients affected by palpable breast cancer may have sonogram and mammogram examination with both normal and benign or nonspecific appearance. Biopsy is used to confirm the symptoms of breast cancer, but it is invasive surgical operation causing psychological and physical impact on patients [4].

We can diagnose breast cancer by using all these features found using Imaging Tests with the assistance of different machine learning algorithms. Machine learning helps with early diagnosis of breast cancer and determines the nature of the cancer by analyzing the tumour size. Machine learning methods are the leading approaches to obtain favourable outcomes among classification and prediction problems. Breast cancer research could benefit from. ML techniques used to identify cancer and predict the presence or absence of tumours. Machine learning techniques can also be used to predict tumour malignancy [5]. Three machine learning algorithms are used in this study namely Support Vector Machine (SVM), Logistic Regression, and Decision Tree (DT).

II. LITERATURE SURVEY

Jaber Alwidian, Bassam H. Hammo, Nadim Obeid, et. al. [5] have proposed an efficient weighted classification system based on algorithms of association rules called WCBA in breast cancer diagnosis. They have proposed a new pruning and prediction method by producing more accurate association rules to increase classification accuracy. M. Nilashi, O. Ibrahim, H. Ahmadi, L. Shahmoradi, et. al. [6] a new knowledge-based system was used to classify breast cancer using clustering techniques called Expectation Maximization (EM) and classification and regression trees (CART) to generate fuzzy rules on the Wisconsin dataset. The accuracy computed in this system based on the fuzzy reasoning method was estimated at 93.20%.

Moh'd Rasoul Al-Hadidi, Abdulsalam Alarabeyyat, Mohannad Alhanahnah, et. al. [7] propose a new method to detect the breast cancer with high accuracy. This method consists of two main parts, in the first part the image processing techniques are used to prepare the mammography images for feature and pattern extraction process. The second part is presented by utilizing the extracted features as an input for a two types of supervised learning models, which are Back Propagation Neural Network (BPNN) model and the Logistic Regression (LR) model. In this paper we examined the accuracy of these models. The results showed that the LR model utilized more features than the BPNN.

Sheetal Shahari, Asmita Wakankar, et. al. [8] present analysis of breast thermogram

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

based on segmentation of region of interest which is extracted a hot region followed by color analysis. The shape, size and borders of the hottest regions of the images can help to determine features which are used to detect abnormalities. The abnormality of breast thermograms is indicated by these features and it is confirmed by comparing with doctors diagnosis. Hence suitability of infrared thermography as a diagnostic tool in breast cancer detection is establihed through this study.

S. k. Mandal, A. Gupta and Animesh Hazra, et.al. [9] presents Ensemble Algorithm in detection of breast cancer with high accuracy. In this Study and Analysis of Breast Cancer Cell Detection using Naïve Bayes, SVM and Ensemble Algorithms, a comparative study was conducted between different machine learning techniques also measuring the time complexity of all the techniques used in this paper, their successful work helped the achieve an accuracy of 95.16%, 95.53%, 95.91% for Naive Bayes, Support Vector Machine and Ensemble technique respectively.

Nithya and Santhi et. al. [10] were able to propose a method for diagnosis in digital mammograms with the help of Grey Level Co-occurrence Matrix (GLCM) features. Computerized tool were considered to play an essential role in diagnosis of breast cancer. Mammography is an important tool which is often used during early stages for detection of breast cancer. Effectiveness of this model is determined by classifying mammogram images into cancer and non-cancer images. Accuracy of this model was calculated to be 96%.

J. Roca-Pardiasa,C. Cadarso-Surezb, P. Tahocesc, M. Ladod, et. al. [11] presents A logistic Generalized Additive Model (GAM). They used linear kernel smoothers with this logistic GAM, they speed up their system using many techniques. In this simulation model they used odds-ratio curves. This model help in early detection process for the breast cancer.

III. OPTIMIZED PREDICTION METHOD FOR BREAST CANCER DETECTION

The block diagram of Optimized Prediction method for Breast Cancer detection using Machine Learning with different Feature Extracting strategies is represented in below Fig. 1.

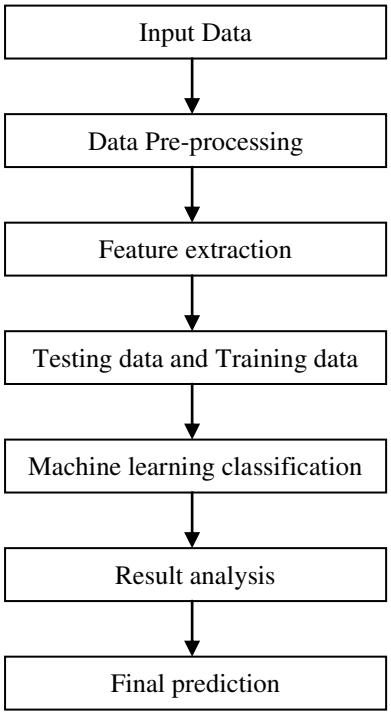


Fig. 1: BLOCK DIAGRAM OF BREAST CANCER DETECTION

The proposed model uses Wisconsin Diagnostic Breast Cancer (WDBC) Dataset from the UCI repository. This dataset was created With 212 malignant instances and 357 benign instances, this dataset contains a total of 569 instances i.e 37.26% malignant

tumour and 62.74% benign tumour. There are exactly 32 attributes WDBC dataset.

The dataset from the UCI Machine Learning Repository will be used and do some cleaning of data (i.e. making the data suitable for our use by removing the empty columns and assigning numbers to text data like malignant is assigned to 1 and benign to 0) for prediction purpose.
Foremost important step toward classification problems is to extract and select the relevant features based on the type and characteristics of problem. Features such as texture, morphological, scale-invariant feature transforms (SIFT), elliptic Fourier descriptors (EFDs) and Entropy base features are extracted. The EFDs features are useful for discriminating the images having epileptic shape. SIFT features that have been used to analyze the problems of panoramas reconstruction, face identification and visual object tracking. Due to their robustness, characteristics such as illumination changes, rotation, noise, scaling and blurry effects, SIFT features have been used in wide area of research. The texture features are most widely used in solving classification issues particularly to classify the colon biopsies. The morphology of tissue is important to determine that tissues are normal or not. Morphological features are extracted from images by converting the morphology of images into set of quantitative values used in classification.

Then split the data into training data (which in our case will be 75%) and testing data (25%). Then at the end scale the data. We have employed three machine learning algorithms, namely: Logistic Regression, Support Vector Machine and Decision Tree (DT).

Support Vector Machine (SVM) is one of the supervised ML classification techniques that is widely applied in the field of cancer diagnosis and prognosis. SVM functions by selecting critical samples from all classes known as support vectors and separating the classes by generating a linear function that divides them as broadly as possible using these support vectors. Therefore, it can be said that a mapping between an input vector to a high dimensionality space is made using SVM that aims to find the most suitable hyperplane that divides the data set into classes.

Decision Tree (DT) is a supervised Machine Learning (ML) method used for classification and regression analysis. Decision tree is based on the divide and conquers methodology. It divides the partition by two methods: Numerical partitions: Typically, partitions are formed on the basis of discrete values with some conditions and Nominal partition: The partitions are formed on the basis of nominal attributes. It leads splitting of the tree depending on attributes values.

Logistic Regression (LR) is an appropriate type of regression model when sure that our dependent variable is of binary type. Logistic regression is used to find the relationship between one dependent variable (which is binary in nature) and one or more independent variables.

After classification module, the performance parameter values are compared for three classification techniques. The results are extracted to decide regarding be benign or malignant of cancer mass. Performance parameters used in this study are Accuracy, Precision and Recall.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

IV. RESULT ANALYSIS

This section describes the parameters that are used for measuring comparative performance of machine learning techniques. Our investigation is based on the Original Wisconsin Breast Cancer Data set that is obtained from the UCI Machine Learning Repository, an online open source repository. This dataset contains 212 malignant instances and 357 benign instances. The performance of described model of Breast cancer detection is evaluated in terms of three parameters as Accuracy, Precision and Recall.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \dots (1)$$

$$Precision = \frac{TP}{TP + FP} \dots (2)$$

$$Recall = \frac{TP}{TP + FN} \dots (3)$$

Where,

1) True Negative (TN) – A percentage of benign that are correctly classified.
2) True Positive (TP) – A percentage of malignant that are correctly classified.
3) False Positive (FP) - The percentage of benign that are incorrectly classified such as malignant.
4) False Negative (FN) - The percentage of malignant that are incorrectly classified as benign.

Table 1 represents the comparative performance of three classifiers.

Table 1: COMPARATIVE PERFORMANCE ANALYSIS

Classifiers	Accuracy	Precision	Recall
SVM	98.4	97	96.8
DT	90.2	91	90.7
LR	91	90	91

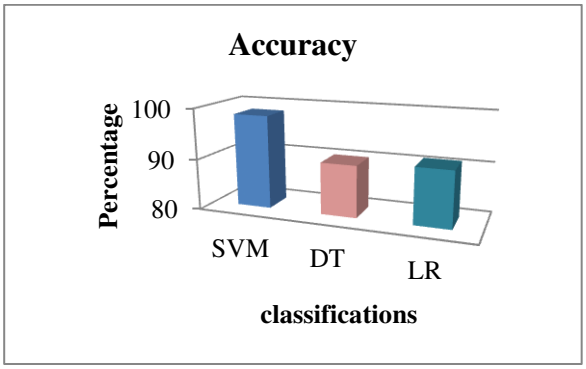


Fig. 2: ACCURACY ANALYSIS

Fig. 2, Fig. 3 and Fig. 4 shows the graphical representation of three classifiers Accuracy, Precision and Recall parameters respectively.

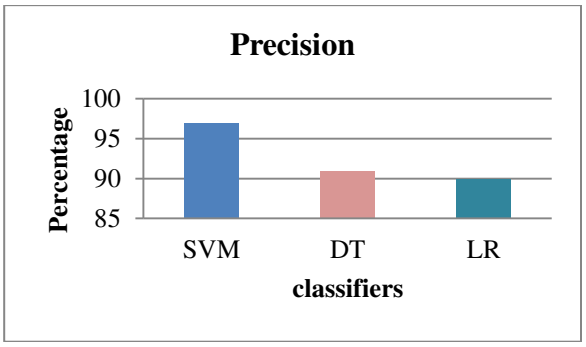


Fig. 3: PRECISION ANALYSIS

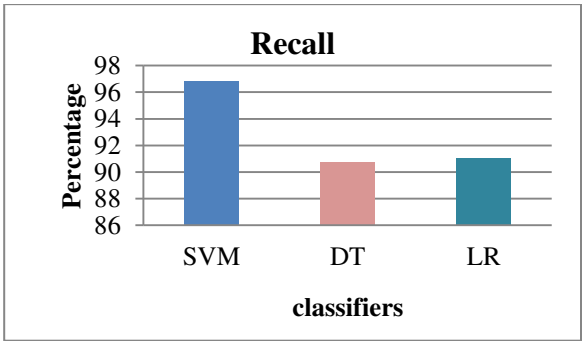


Fig. 4: RECALL ANALYSIS

From results it is clear that, SVM based Breast Cancer detection model with different Feature Extracting strategies is efficient than

other classifications as DT and LR. Obtained parameter values for SVM classifiers are Accuracy as 98.4%, Precision as 97% and Recall as 96.8%.

V. CONCLUSION

In this paper, Optimized Prediction method for Breast Cancer detection using Machine Learning with different Feature Extracting strategies is described. The proposed model uses Wisconsin Diagnostic Breast Cancer (WDBC) Dataset from the UCI repository. 75% of data used for Training and 25% of data used for Testing. Used machine learning algorithms are Logistic Regression, Support Vector Machine and Decision Tree (DT). The performance of described model of is evaluated in terms of three parameters as Accuracy, Precision and Recall. From results it is clear that, SVM based Breast Cancer detection model with different Feature Extracting strategies is efficient than other classifications as DT and LR.

VI. REFERENCES

[1] Valerii E. Orel, Oleksdandr Rykhalskyi, Liubov Syvak, Ivan Smolanka, Ivan Smolanka, Anton Loboda, Andriy Lyashenko, Iryna Dosenko, Alexander Mokhonko, Tetiana Golovko, Alexander Ganich, Varerii. B. Orel, "Computer-assisted Inductive Moderate Hyperthermia Planning For Breast Cancer Patients", 2020 IEEE 40th International Conference on Electronics and Nanotechnology (ELNANO), Year: 2020

[2] Mengwan Wei, Yongzhao Du, Xiuming Wu, Jianqing Zhu, "Automatic Classification of Benign and Malignant Breast Tumors in Ultrasound Image with Texture and Morphological Features", 2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID), Year: 2019

[3] Nirmine Hammouch, Hassan Ammor, "Microwave Imaging for Early Breast Cancer Detection using CMI", 2019 7th Mediterranean Congress of Telecommunications (CMT), Year: 2019

[4] S.P. Sachin Raj, N. Sri Madhava Raja, M.R. Madhumitha, V. Rajinikanth, "Examination of Digital Mammogram Using Otsu's Function and Watershed Segmentation", 2018 Fourth International Conference on Biosignals, Images and Instrumentation (ICBSII), Year: 2018

[5] Jaber Alwidian, Bassam H. Hammo, Nadim Obeid, "WCBA: Weighted classification based on association rules algorithm for breast cancer disease", Applied Soft Computing, Volume 62, 536-549, 2018.

[6] M. Nilashi, O. Ibrahim, H. Ahmadi, L. Shahmoradi, A knowledge-based system for breast cancer classification using fuzzy logic method, Telematics and Informatics, 34(4) (2017) 133-144

[7] Moh'd Rasoul Al-Hadidi, Abdulsalam Alarabeyyat, Mohannad Alhanahnah, "Breast Cancer Detection using K-Nearest Neighbor Machine Learning Algorithm", 2016 9th International Conference on Developments in eSystems Engineering (DeSE), Year: 2016

[8] Sheetal Shahari, Asmita Wakankar, "Color analysis of thermograms for breast cancer detection", 2015 International Conference on Industrial Instrumentation and Control (ICIC), Year: 2015

[9] S. k. Mandal, A. Gupta and Animesh Hazra, "Study and Analysis of Breast Cancer Cell Detection using Naïve Bayes, SVM and Ensemble Algorithms", IJCA (0975 – 8887) Volume 145 – No.2, July 2016

[10] B. Santhi, R. Nithya, "Classification of Normal and Abnormal Patterns in Digital

Mammograms for Diagnosis of Breast Cancer", Aug. 2011

[11] J. Roca-Pardiasa,C. Cadarso-Surezb, P. Tahocesc, M. Ladod, ”ssessing continuous bivariate effects among different groups through nonparametric regression models: An application to breast cancer detection”, Elsevier, Computational Statistics and Data Analysis, Vol. 52,(2008)