TEST Engineering & Management

July-August 2020
ISSN: 0193-4120 Page No. 3014 - 3025

Diagnosis of Breast Cancer for Mammographic Images Using Various Data Mining Techniques

Poornima Deshpande[1],*, Prof. K. S. Ghuge[2],*, Prof. S. M. Jaybhaye[3]*

[1]Dept. of IT, SCOE, Vadgoan, Pune, India

[2] Assistant Professor, Dept. of IT, SCOE, Pune, India

[3] Assistant Professor, Dept. of IT, SCOE, Pune, India

Article Info
Volume 83
Page Number: 3014 - 3025
Publication Issue:
July-August 2020

Abstract

Numerous surgical oncologists believe that Breast Cancer is a noxious disease among others as quite a number of women suffer from this condition. It impacts almost 2.1 million women and it accounts for the death of 15% of women worldwide making more and more women vulnerable to this disease every year. Though there is no permanent solution to this disease, diagnosing it at an early stage can change the outcome in a very significant way. It can greatly reduce the risk of spreading the tumour to other parts of the Breast. In the past decade, CAD systems which assists in diagnosing the disorder with the help of AI tools has proved to be an eminent way for diagnosing Breast Cancer. Supervised techniques have been used for classification. This survey paper analyses various papers based on the concept of Decision Tree algorithms to diagnose breast cancer. It also lays out the detailed steps for the technique proposed in these papers which mainly consists of 4 stages. Pre-processing is the initial stage followed by segmentation and feature extraction. The system then enters into the classification phase which categorizes mammograms. These techniques focus on removing unwanted noise from the images by pre-processing them and applying filters. The targeted Region is extracted from these mammograms. The mammograms are further classified under 2 labels as benign(non-lethal) or malignant(lethal). Further, distinct algorithms of Decision tree (CART and C4.5) are surveyed to classify and detect cancer. Datasets such as mini-MIAS is generally used. The studies show that the results can also be improvised further with the Random Forest algorithm which is based on a supervised learning model. Random Forest works on an ensemble learning algorithm called Bootstrap Aggregation or Bagging for Classification and Regression.

Article History
Article Received: *19 August 2019*
Revised: *27 November 2019*
Accepted: *29 January 2020*
Publication: *6 May 2020*

Keywords: *Pre-processing, Region of Interest, Segmentation, GLCM, Decision Tree Classifier.*

1. INTRODUCTION

Breast Cancer leads to innumerable deaths among human beings. The conventional form of detecting Breast Cancer in hospitals is CT scans using X-rays. Screening is necessary because Breast Cancer shows no prodromes when the size of the tumor is small and when cancer can be easily treated. Also, MRI that makes use of a powerful magnetic field is an acceptable technique but all these methods now lead to inaccuracy due to the swarming population and a smaller number of doctors. These techniques are also inefficient as they fail to predict cancer correctly. The prodromes of this disease cannot be detected in the early stages which causes the death of women in substantial numbers.

World Health Organization (WHO) states that 2.1 million females are impacted by this disorder on a yearly basis [1]. In 2018, the number was approximately the same. According to the latest report “Breast Cancer Facts and Figures 2019-2020”

July-August 2020
ISSN: 0193-4120 Page No. 3014 - 3025

data shows infected number of patients with this condition in 2019 which is approximately 3.8 million females [1]. There are 41,760 deaths of women alone in the US by Breast Cancer. The most common symptom includes a painless lump. Less common symptoms are redness of the skin, breast pain or heaviness and persistent changes, such as swelling and thickening. Breasts are made up of glands for the production of milk. These glands are called lobules and ducts form a link between them. Lymphatic tissues fill up the existing part. The occurrence of this condition is caused by the uncontrollable division of the cell tissue. The cell tissue becomes abnormal and starts to divide uncontrollably. These abnormal cells form a lump or mass and grow into a tumor. Cell tissues become abnormal due to the deposition of calcium These calcium deposits are obtained on Breast tissue in 2 forms namely Micro-Calcification and Macro-Calcification. Thus, the tumor is classified into classes as Benign and Malignant also.

Benign tumors are non-lethal but malignant tumor has the capability to proliferate to different areas and infect them. This is where Automated Classifiers come into the picture. In the age of technology and AI, Machine Learning acts as a stupendous tool. By applying CAD (Computer-Aided Detection) tools cancer can be diagnosed at an early stage. Thus, it can greatly reduce the risk of death of women worldwide. Data can be congregated from MIAS repository. The goal here is to determine Breast Cancer in Mammographic images in a pre-mature phase.

Firstly, the mammograms of the breast are initially processed by removing undesirable noises mentioned below in detail. Undesirable noise from mammograms can be eliminated using various filters such as Median or Gaussian filters. Further, the exact targeted region is extricated from mammograms viz segmentation technique. Input obtained via this phase is converted into levels of gray. The matrix obtained from Gray level co-occurrence matrix will act as an input matrix to decision tree. The aforementioned technique will classify images as benign or malignant.

2. LITERATURE SURVEY

Burau Bektas, Ilkim Ecem Emre, Elif Karta [2] worked with the mini-MIAS database in this study wherein the prediction is done to find out if a lump can be found out in a breast. The lump can then be categorized as benign or malignant. Undesirable noise can be eliminated and images can be filtered with the help Gauss, Average, Weiner, Median filters. To make regions of the breast more visible CLAHE (Contrast limited Analysis Histogram Equalization) technique is used. This study shows that with HOG and LBP new datasets can be generated. Further, for comparing performance, classification algorithms such as C5.0(normal and boosted), Naïve Bayes, CART, and RFC are implemented.

Authors Dr. S. N. Singh, Shivani Thakral in [3], have used WBCD taken from original dataset for analyzing as well as diagnosing this condition. Algorithms used are J4.8, Simple CART and Bayes classifier (Naive Bayes, Bayesian LR). Accuracy is used as the main factor in the classification of data of both the algorithms by using the WEKA tool. It is concluded that amid all of them, this particular classification method is efficient which is providing an accuracy of 98.13%. It provides better time complexity

S. Murungan, B. Muthu Kumar, S. Amudha are the authors of “Classification and Prediction of Breast Cancer using Linear Regression, Decision Tree and Random Forest” [4]. The work is done upon the WBCD procured from the UCI Machine Learning Repository. Center point focuses on the class label of tumor as lethal or non-lethal. Classification of this paper has been accurately determined upto 84.14% whereas the rate of prediction is upto 88.14% [4]. Methods used are LR, DT and RFC. In LR, the association between the distinct properties is computed. Only those attributes are considered for anatomy which are strongly

July-August 2020
ISSN: 0193-4120 Page No. 3014 - 3025

interconnected to one another.

Liu Yi, Wu Yi [5], the authors have conducted experiments on variegated training set of WBCD database. DT model is merged with characteristics feature which is a commonly used method. It reduced the complexity of DT by deleting some highly relevant features and prefer tumor thickness, mitosis, single epithelial cell size, and cell shape consistency. Model accuracy is determined along with the sensitivity. Specificity and ROC curves are also determined for anticipating the performance of the decision tree classifier. The results have obtained an accuracy of 94.3% considering a variety of metrics.

The authors Hind Elouedi, Walid Meliani, Zied Elouedi, Nahla Ben Amor [9] proposed a mixed process for classifying breast cancer as dangerous or non-dangerous by combining two different techniques i.e. decision trees and clustering. It takes a unique approach and makes a refined treatment based on the WBCD database. Rather than classifying cancer as just benign or malignant, this study takes an idiosyncratic perspective. C4.5 is used to determine the quality of these clusters. The clustering algorithm uses the K-means algorithm. Here different types of malignant cases are described to enhance the results.

This paper is a computable process to determine breast cancer from mammogram images. In [11], the authors Shruthi G.K, Dr. Sunitha M.R. have proposed the method of classification based on three groups, twelve divisions, and six structures. The three groups are normal, benign and malignant tumor. It also uses a random forest classifier to classify images for high accuracy. For noise reduction, breast cancer is classified into stages. The database used is the Mammographic Image Analysis Society database. For pre-processing, the techniques used are image enhancement and noise removal where high intensity and additive noises are removed. Here, the purpose of disk is to extract the boundary features which acts as a structuring element. While segmenting the image is partitioned and the adaptive diffusion color model is used for extracting ROI. With the help of RFDC over 250 images are tested and when compared with SVM, RF classifier is known to achieve high accuracy.

The authors in [16] B. Padmapriya, T. Velmurugan have basically done a survey work for analyzing different algorithms of data mining such as ID3 and C4.5 for classification purposes. This work reviews technical articles on breast cancer diagnosis. Articles are reviewed for the purpose of exploring the current scenario in data mining field. It basically acts as the compilation for research purposes.

Ronak Sumbaly, N. Vishnusri, S. Jeyalatha [18] worked on various DM techniques for detecting tumor. It discusses the layout of the breast and its anatomy. The authors have also scrutinized the risk factors which can be prevented along with some non- preventable risk factors. Input data in this model is trained. Afterwards tested using k-fold cross-validation where k=10. Further, the WEKA tool is used for implementing the J48 algorithm. After implementing the J48 algorithm the data is split into benign and malignant. The dataset consists of 699 tuples and 10 different attributes from which the correctly classified instances are 661 which shows 94.5637% precision. Among them, the incorrectly classified instances account for 38 with a precision rate of 5.4363%. Alternate approaches are discussed such as Neural Networks, Naïve Bayes Classifier, and Digital Mammography. In this paper [20] Dr. J. Arunadevi, K. Ganeshamoorthi, has worked on predicting Breast Cancer by using 3 classifiers on mammographic images. The task is to use a minimal number of attributes in order to handle data overhead. Time as well as cost reduction can be done by reducing the variables involved. By using 3 classifiers (KNN, SVM, and Neural Networks) the outcome parameters are impacted greatly. The feature selection facilitators used are GLM and RF. For the task of classifying, small database with N characteristics are chosen. Some of the parameters used for evaluation are Accuracy, Error, Kappa, Precision, and Recall.

The authors R.D. Ghongade and D.G. Wakde

July-August 2020
ISSN: 0193-4120 Page No. 3014 - 3025

of "Computer-aided Diagnosis System for Breast Cancer Using RF Classifier" [21] have worked with the mammographic image which is taken from MIAS database and gaussian filter further is applied to it. CLAHE augments the levels of gray in an image. It is thus used for the feature extraction with characteristics which include Contrast and Correlation. Energy, Homogeneity, Mean, Standard Deviation, Entropy, Variance, Smoothness, Kurtosis, Skewness and Inverse Different Moment (IDM). Area, solidity, eccentricity, perimeter and major axis length are some of the extricated features. Lastly, the final result to categorize the mammograms into lethal, non-lethal or normal comes from RF algorithm.

Shubham Sharma, Archit Aggarwal, Tanupriya Choudhury [22] have used the WBCD database. It aims to present the breast cancer prediction such as the Random Forest (RF), KNN, and Naïve Bayes (NB). The dataset is used to compare the performance of these techniques like precision and accuracy. In the paper, out of 569 observations recorded training set uses 398 while testing set is using 171. This paper shows that RF has an outstanding confusion matrix with an accuracy of 97.74%. Though RF has an excellent recall measure the concluding outcome shows that KNN has the better accuracy as well as precision and F1 Score when compared to NB and RF.

3. ARCHITECTURE AND MODELLING

This survey guides and focuses on various approaches for diagnosing breast cancer using Decision Trees. The entire flow of this paper is divided into the following phases:

A. Initially, in the data collection phase, the data is collected via datasets namely MIAS.

B. The second level consists of images from datasets. These Images are primarily processed to annihilate redundant noise. Various noise removal filters are applied in pre-processing stage.

C. In the third phase, the image is segmented to extricate the exact Area of Concern.

D. The fourth level includes extraction of features. Here, necessary features are extracted out from mammogram. It helps in recognizing crucial features.

E. In fifth level, classification task is executed. The mammograms are categorized under 2 labels as benign and malignant.

July-August 2020
ISSN: 0193-4120 Page No. 3014 - 3025

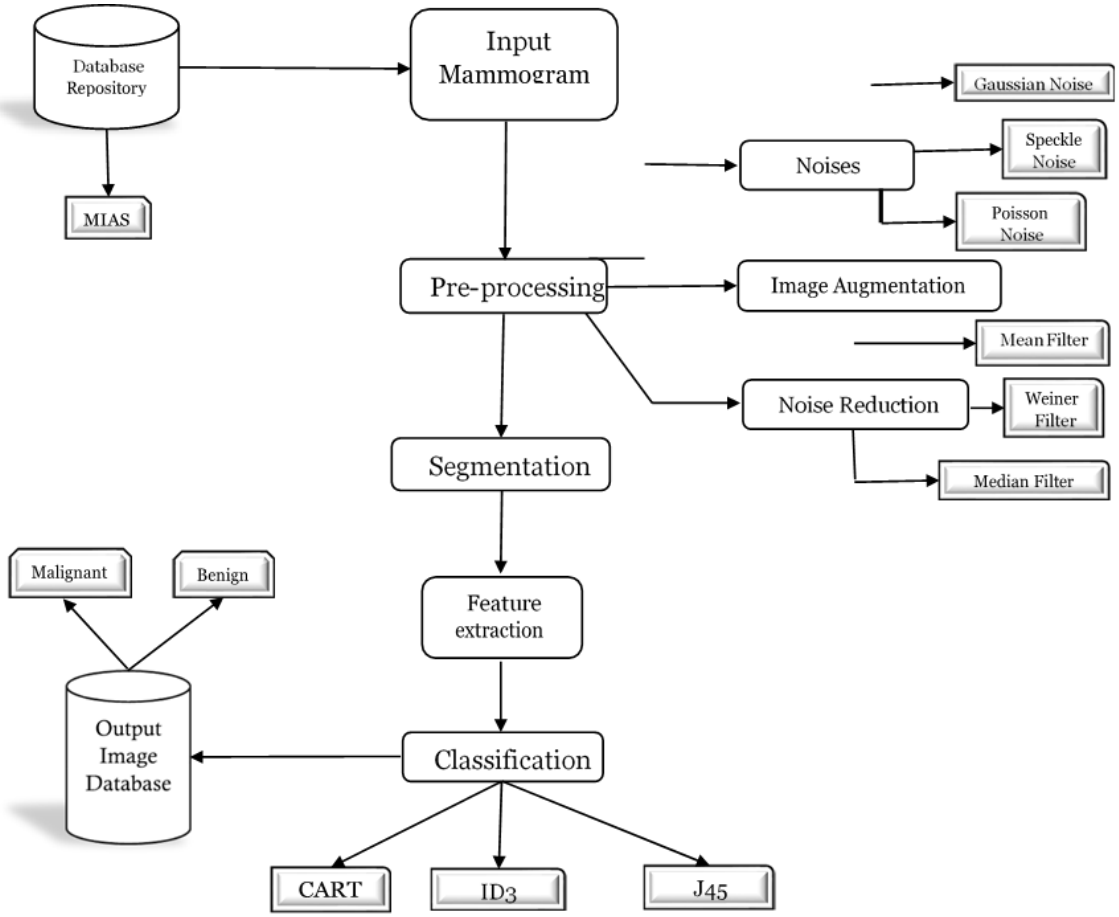


Fig. 1. *Block Diagram of Breast Cancer Diagnosis*

4. METHODOLOGY

4.1 PRE-PROCESSING STAGE

The data is accumulated from the above-mentioned dataset. The images from these datasets are used for pre-processing of images and classifying them for Computer-Aided Detection of Mammograms.

For the training procedure, a random number can be selected. Usually, 80% of the database goes to the training set and the remaining 20% data goes to the test dataset. Common noises encountered in an image are Poisson along with Gaussian and Speckle Noise. In some cases, Salt and Pepper noise are also being found. Undesirable noises when not removed congruously leads to inaccurate results. The patient can also be severely impacted [11].

This aforementioned process is carried out to reduce undesirable noises. The areas other than the target region is also eliminated to enhance the region of interest. Distinct techniques are present in order to pre-process the mammogram images.

The pre-processing phase includes two steps:

1. Image Augmentation: Image Augmentation deals with the inclusive quality of the images. Over-dark or Over-bright pixels in the image can be reduced considering the intensity of the image.
2. Noise Reduction: Varied filters can be applied such as Median along with Mean and Gaussian Filter. Weiner filter and Adaptive Median have also been proved efficient. These filters can be applied for the purpose of noise reduction. They are used to preserve the edges within an image and are also used for the smoothening of the images.

A. Mean filter

July-August 2020
ISSN: 0193-4120 Page No. 3014 - 3025

The above-mentioned filter is known to refine the attributes of image for naked eye [23]. It is a spatial filter. Usually it works by replacing the values present in the center of the window with the average of all the elements present in the window. Basically, this filter replaces the component with the average number intrinsically as well as of the neighbors.

B. Median filter

Random noise can be reduced to a great extent using this filter. It is considered distinct from the above filter as it replaces the pixels in the center of the frame with the median of all the pixels present in the frame.

C. Gaussian filter

This filter eliminates the uproar by blurring the images and remove speckles from an image. The image is blurred by a Gaussian function which results in a Gaussian blur. It is usually used in images with high noise.

D. Adaptive Median filter

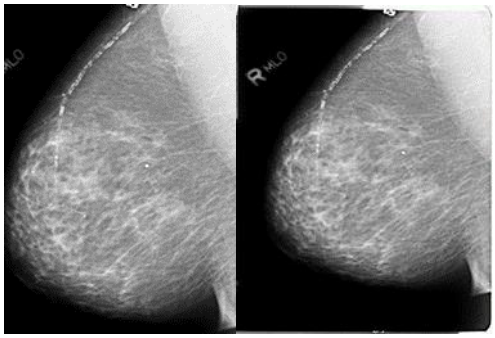
The aforementioned filter intensifies the pixels leaving out the composition of image to avoid blurring. It does so by preserving the edges. It filters out the noise from corrupted areas.

E. Weiner filter

This is an ideal filter. It strives to diminish the mean square error [23].

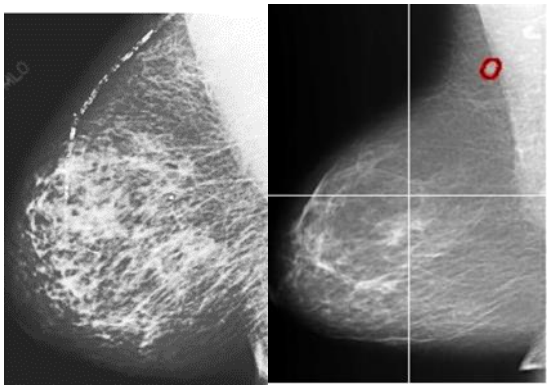
TABLE OF ABBREVIATION

Sr No.	*Abbreviations*	*Description*
1	*AI*	*Artificial Intelligence*
2	*CAD*	*Computer-Aided Detection*
3	*CART*	*Classification and Regression Trees*
4	*WBCD*	*Wisconsin Breast Cancer (Diagnostic)Dataset*
5	*MIAS*	*Mammographic Image Analysis Society*
6	*DM*	*Data Mining*
7	*DT*	*Decision Trees*
8	*LR*	*Linear Regression*
9	*RFC*	*Random Forest Classifier*
10	*HOG*	*Histogram of Oriented Gradients*
11	*LBP*	*Local Binary Pattern*
12	*AC*	*Actual Class*
13	*PC*	*Predicted Class*
14	*TP*	*True Positive*
15	*TN*	*True Negative*
16	*FP*	*False Positive*
15	*FN*	*False Negative*



(a)Initial image (b) Image after getting cropped

Fig. 2. *Pre-processed Images [13]*



(a)Histogram Equalization Technique (b) Augmented Image

Fig. 3. *Image Augmentation [13]*

4.2 TEXTURE EXTRACTION USING GRAY LEVEL CO-OCCURRENCE MATRIX(GLCM)

This matrix is also known as the Gray Level

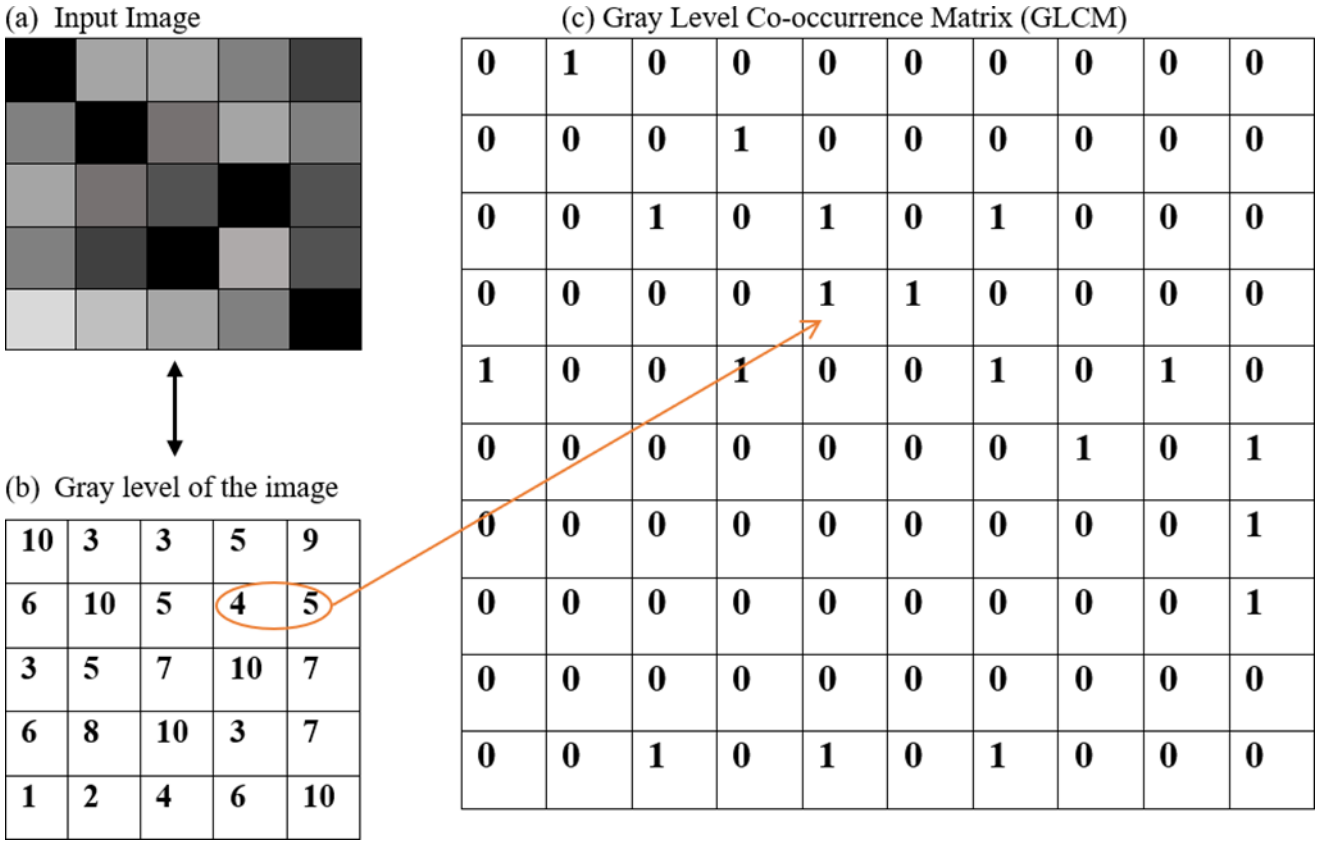
July-August 2020
ISSN: 0193-4120 Page No. 3014 - 3025

Spatial Dependence Matrix which is based on texture segmentation system and performs supervised texture segmentation. The characteristic of an image can be extricated via this method. The obtained matrix represents the variegated attributes of an image. F1, F2, F3, Fn is the form for expressing texture within the image. GLCM is proposed by Haralick in the 1970s and helps in extracting second-order statistical features [10].

The area of interest which is obtained from the segmentation phase can be taken as an input i.e. the target area can be converted into the group of attributes by using particular formulas. The matrix formed contains the gray level value of each pixel from the ROI.

The function identifies the image traits by computing how often the recurring pixels appear in pairs which are distinct values and in definite spatial relationship. Thus, relationship is determined and the features are extracted.

The features extricated must be cautiously chosen. Classification is done on this area obtained via GLCM. Instead of the entire area, reduced representation provides for a more accurate result. Assuming that the direction and distance is specified, mostly it extricates the basic features such as Energy, Entropy, Homogeneity, Contrast, and Correlation[25].



10	3	3	5	9
6	10	5	4	5
3	5	7	10	7
6	8	10	3	7
1	2	4	6	10

0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0
0	0	1	0	1	0	1	0	0	0
0	0	0	0	1	1	0	0	0	0
1	0	0	1	0	0	1	0	1	0
0	0	0	0	0	0	0	1	0	1
0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0
0	0	1	0	1	0	1	0	0	0

Fig. 4. *Extracting features from the images converted to gray levels*

4.3 CLASSIFICATION

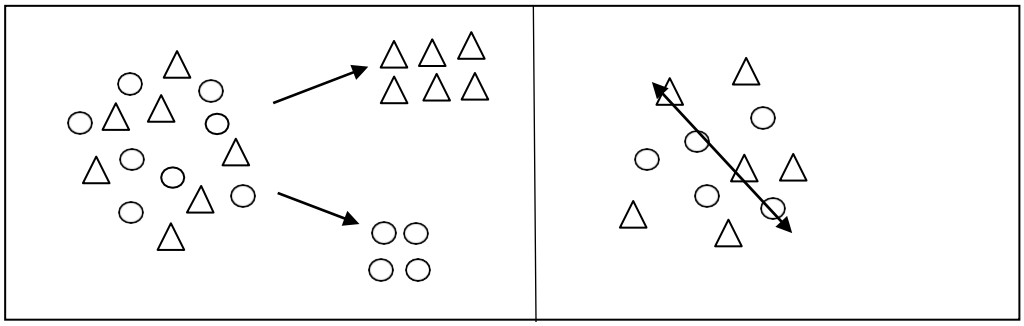
The training data can be accumulated from the given dataset. Learning and prediction are the two phases of Decision Tree Classifier. DT is the most robust and sought-after classification technique. It is categorized into supervised learning method. Obtained data is used for training the model in learning stage wherein the models learn the trained data. In succeeding stage, model predicts a response using the trained data. The model infers specific

July-August 2020
ISSN: 0193-4120 Page No. 3014 - 3025

rules from the previously learned data. A training model is thus created. It anticipates the label of the target variable. Root of the tree is considered as an initial point for anticipating the class of that particular variable. It follows the structure of a tree of decisions and their predictable results. Each node leads to the classification results and Every decision taken leads to the sub-nodes. Decision tree models are highly interpretable.

The model has to be trained and tested and training set determines the accuracy of the classifier [7]. Decision trees are often modeled on the idea of two steps: Induction and Pruning. In induction, the tree is built and in pruning, all the unnecessary branches are cut-off thus reducing ambiguity. CART and C4.5 are some of techniques used widely:

A. Classification And Regression Tree : CART is also reffered to as an estimation model. It estimates the outcome of the model based on variables that are labelled. CART makes a decision by answering a set of If-Else questions. It is reffered as the umbrella term in which the questions are sequentially structured. In the former algorithm the targeted variable can be classified under a particular label and hence called "catergorical". However the latter envisages the value of tree as the variable is called "continuous". To a great extent, CART can encapsulate the non-linearity in the data.



(a) Variable classified categorical b) Continuous variable

Fig. 5. *CART algorithm*

B. C4.5 : Using the training data, ID3 creates a classifier. To test the variables at each stage this model employs the approach from top to bottom for a greedy search.

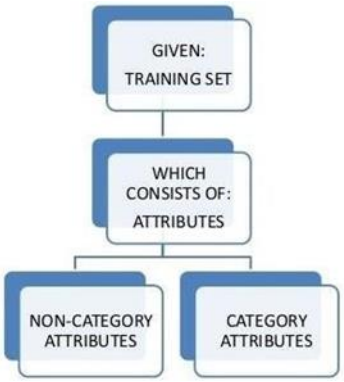


Fig. 6. *ID3 Algorithm*

July-August 2020
ISSN: 0193-4120 Page No. 3014 - 3025

In this technique, the result is predicted in following forms:

Precision = ta÷ta+faRecall=(ta÷ta+fn)

1. Accuracy = ((ta+tb)÷ (ta+fa+ fb+tb)) ∗ 100%
2. Sensitivity = ta÷ta+fb ∗ 100%
3. Specificity = (tn÷ tn+fp) ∗ 100%

Where,
ta = TP tb = TN fa = FP fb = FN

With aid of DT we classify the results of patients under 2 labels: Benign also represented as 0 and Malignant which can be represented as 1.

Suppose, the patient's mammogram shows Malignant tumor prevailing in the final result, then the above parameters best describe execution of our classifying approach.

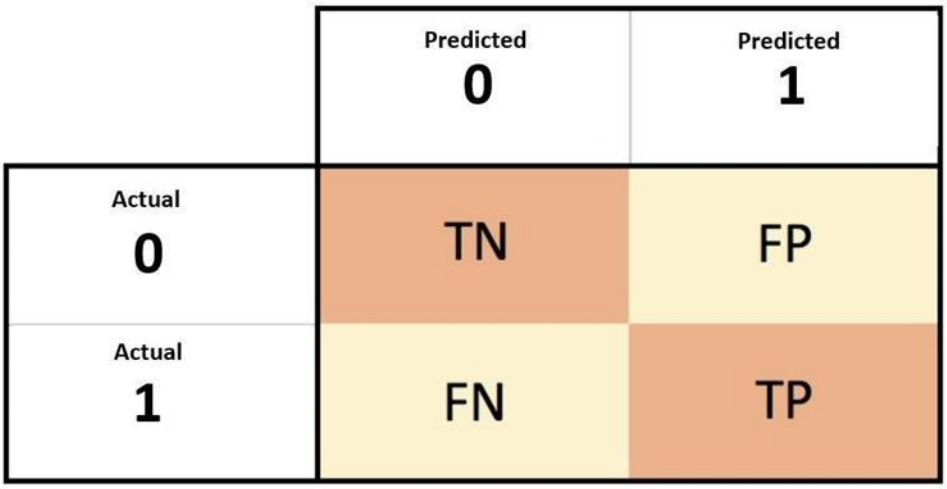


Fig. 7. *Confusion Matrix estimating Positive and Negative values*

TP and TN are the outcomes predicted precisely whereas FP and FN are the outcomes that anticipate incorrectly and hence we need to curtail it.

A. TP indicates AC value to be positive and PC value to be positive as well.

B. TN indicates AC value to be negative and PC value to be negative as well.

C. FP indicates AC value to be negative and PC value to be positive.

D. FN indicates AC value to be positive and PC value to be negative.

Suppose, according to results the patient has been detected positive for Breast Cancer whereas in reality the patient doesn't have a malignant tumor, then this is termed as a false positive rate. Our goal should be to curtail these type of error rates.

July-August 2020
ISSN: 0193-4120 Page No. 3014 - 3025

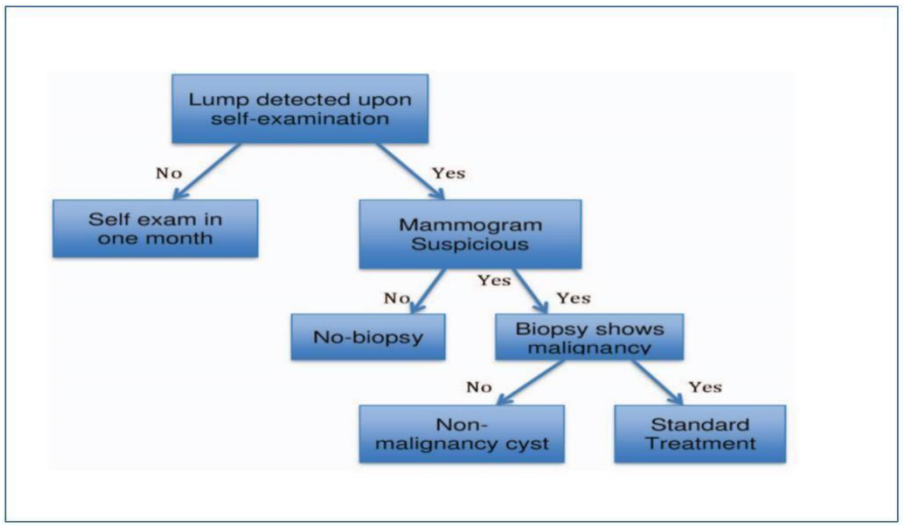


Fig. 7. *DT Classifier divided into sub-nodes to detect category of treatment [12]*

5. CONCLUSION

Concluding the survey, it can be said that Breast Cancer can be classified under 2 labels as benign and malignant. Diverse authors and researchers have concluded in their research that this classifier can be can be adroitly used to treat patients enduring from this condition.

The DT classifier has been used for the classification of the lump in the breast. It aims to achieve high accuracy. Such advanced techniques can considerably increase the chances of patient survival. The decision tree classifier gives a low error rate and high precision rate with accuracy being up to 97.7%. Further, the Random Forest classifier can be used to improvise the technique. RF has proved to be more efficient in categorising mammograms. RF takes multiple distinct DTs and predict the result. This technique is based on recursion. Machine learning applications does not aim to take over the surgical oncologist and experts. Rather it aims to improve the techniques and help them achieve a low mortality rate and high accuracy rate.

REFERENCES

[1] American Cancer Society. Breast Cancer Facts & Figures - Global Headquarters: American Cancer Society Inc., 2019-2020.

[2] Burau Bektas, Ilkim Ecem Emre, Elif Kartal "Classification of Mammography Images by Machine Learning Techniques" 3rd International Conference on Computer Science and Engineering IEEE 2018.

[3] Dr. S. N. Singh, Shivani Thakral "Using Data Mining Tools for Breast Cancer Prediction and Analysis" 4th International Conference on Computing Communicaton and Automation(ICCCA) IEEE, 2018.

[4] S. Murungan, B. Muthu Kumar, S. Amudha "Classification and Prediction of Breast Cancer using Linear Regression, Decision Tree and Random Forest" International Conference on Current Trends in Computer, Electrical, Electronics and Communication (ICCTCEEC) IEEE, 2017.

[5] Liu Yi, Wu Yi "Decision Tree Model in the Diagnosis of Breast Cancer" International Conference on Computer Technology, Electronics and Communication (ICCTEC) IEEE, 2017.

[6] Bhumika Gupta, Aditya Rawat, Akshay Jain "Analysis of Various Decision Tree

July-August 2020
ISSN: 0193-4120 Page No. 3014 - 3025

Algorithms for Classification in Data Mining" International Journal of Computer Applications (Volume 163 – No 8), April 2017.

[7] Hanif Yaghoobi, Alireza Ghahramani Barandagh, Zhila Mohammadi "Breast Cancer Diagnosis Using, Grey-level Co-occurrence Matrices, Decision Tree Classification and Evolutionary Feature Selection" 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI) IEEE, 2015.

[8] Dishant Mittal, Dev Gaurav and Sanjiban Sekhar Roy "An Effective Hybridized Classifier for Breast Cancer Diagnosis" International Conference on Advanced Intelligent Mechatronics (AIM) IEEE, 2015.

[9] Hind Elouedi, Walid Meliani, Zied Elouedi, Nahla Ben Amor "A Hybrid Approach Based on Decision Trees and Clustering for Breast Cancer Classification" International Conference of Soft Computing and Pattern Recognition IEEE, 2014

[10] A. Markkongkeaw, A. Phinyomark, P. Boonyapiphat, P. Phukpattaranont "Preliminary Results of Breast Cancer Cell Classifying Based on Gray Level Co-occurrence Matrix" Biomedical Engineering International Conference (BMEiCON) IEEE, 2013.

[11] Shruthi G.K, Dr. Sunitha M.R. "A Quantitative Approach for Determining Breast Cancer Using Mammography Images" Proceedings of the 2nd International Conference on Electronics, Communication and Aerospace Technology (ICECA 2018) IEEE Conference Record, 2018.

[12] Smita Jhajharia, Seema Verma, Rajesh Kumar "A Cross-Platform Evaluation of Various Decision Tree Algorithms for Prognostic Analysis of Breast Cancer Data" International Conference on Inventive Computation Technologies (ICICT) IEEE, 2016.

[13] Vibha L, Harshavardhan G M, Pranaw K, P Deepa Shenoy, Venugopal K R, L M Patnaik "Classification of Mammograms Using Decision Trees" 10th International Database Engineering and Applications Symposium (IDEAS'06) IEEE,2006.

[14] Nusaibah Kh. AL-SALIHY , Turgay IBRIKCI "Classifying Breast Cancer by Using Decision Tree Algorithms" ICSCA, 2017.

[15] Aswini Kumar Mohanty, Manas Ranjan Senapati, Swapnasikta Bererta, Saroj Kumar Lenka "Texture-based features for classification of mammograms using decision tree" 2012.

[16] B. Padmapriya, T. Velmurugan, "A Survey on Breast Cancer Analysis Using Data Mining Techniques" International Conference on Computational Intelligence and Computing Research IEEE, 2014.

[17] H.D. Cheng, Juan Shan, Wen Ju, Yanhui Guo, Ling Zhang "Automated breast cancer detection and classification using ultrasound images: A survey" Pattern Recognition, Science Direct, Elsevier, 2009.

[18] Ronak Sumbaly, N. Vishnusri, S. Jeyalatha "Diagnosis of Breast Cancer using Decision Tree Data Mining Technique" International Journal of Computer Applications, 2014.

[19] Vibhav Prakash Singh, Ayush Srivastava, Devang Kulshreshtha, Arpit Chaudhary, Rajeev Srivastava "Mammogram Classification Using Selected GLCM Features and Random Forest Classifier" International Journal of Computer Science and Information Security (IJCSIS), 2016.

[20] Dr.J. Arunadevi, K. Ganeshamoorthi "Feature Selection Facilitated Classification ForvBreast Cancer Prediction" Proceedings of the Third International Conference on Computing

July-August 2020
ISSN: 0193-4120 Page No. 3014 - 3025

Methodologies and Communication (ICCMC), IEEE 2019.

[21] R.D. Ghongade and D.G. Wakde "Computer-aided Diagnosis System for Breast Cancer Using RF Classifier" IEEE WiSPNET Conference, 2017.

[22] Shubham Sharma, Archit Aggarwal, Tanupriya Choudhury, "Breast Cancer Detection Using Machine Learning Algorithms", IEEE 2018.

[23] R. Ramani, Dr. N. Suthanthira Vanitha, S. Valarmathy "The Pre-processing techniques for Breast Cancer Detection in Mammography Images" I.J. Image, Graphics and Signal Processing, MECS 2013.

[24] M.P. Sukassini and T. Velmurugan "A Survey on the Analysis of Segmentation Techniques in Mammogram Images" Indian Journal of Science and Technology, Vol 8(22), September 2015.

[25] K.S. Ghuge, K.S. Korabu, Aayushi Somani "Breast Cancer Detection Using Clustering And SVM" Test Engineering and Management Volume 83 Page Number: 2196 - 2205 Publication Issue: March - April 2020.

[26] Karthikeyan Ganesan, U. Rajendra Acharya, Chua Kuang Chua, Lim Choo Min, K. Thomas Abraham, and Kwan-Hoong Ng "Computer-Aided Breast Cancer Detection Using Mammograms: A Review" REVIEWS IN BIOMEDICAL ENGINEERING, VOL.6, IEEE 2013.

Poornima Deshpande was born on the 11th of February 1998 in Parbhani District of Maharashtra State. Her domain is Machine Learning and Deep Learning. She is pursuing her B.E degree from Sinhgad College of Engineering, Pune and with her focus on Information Technology.

Prof. Kalyani Ghuge is Asst. Prof. in the Dept. of IT at SCOE, Pune. Completed her M.E degree in Computer Engineering field. Her area of research is Cloud Computing, Deep Learning and Machine Learning.

Prof.S.M. Jaybhaye is Asst. Prof. in the Dept. of IT at SCOE, Pune. Her educational qualification is ME (Computer Engg). Her area of research is cloud computing, optimization, resource provisioning.