International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8, Issue-1, May 2019

Machine Learning Approach for Breast Cancer Prediction

S.Vasundhara , B.V. Kiranmayee , Chalumuru Suresh

Abstract: ***Breast cancer is horrendous disease after skin cancer which is most common in woman and it is a foremost cause for the upsurge in mortality rate. Screening mammography is the operative procedure for detecting masses and abnormalities allied to breast cancer. Digital mammograms are utmost operative source that helps in early detection of cancer in women with no symptoms and diagnose cancer in women with symptoms like pain in lump, nipple discharge which diminutions deaths and upsurges chances of survival. Usually clinician cannot spare more time on a patient to weigh the complaints and suggest a possible diagnosis by considering past records. During this stage, there is more chance to medical errors and wrong diagnosis. By using machine learning in diagnosing breast cancer improves accuracy by reducing misclassifications and saves time in diagnosing. The proposed work is instinctive classification of mammogram images as Benign, Malignant and Normal using various machine learning algorithms. Classification is an identification technique used to classify consolidated data into different categories. Initially pre-processing of mammograms is performed by using Gabor wavelet filtering technique, Adaptive global threshold and morphological operations like opening and closing techniques for cancer analysis to shrink false positives. Finally classification of the pre-processed images is performed and mammograms are classified into benign, malignant and normal with the use of 3 classifiers Support Vector Machines, Convolutional Neural Network and Random Forest. The performance of the trained classifiers is evaluated using metrics to attest which model is efficient in classifying mammograms and predicting breast cancer.***

Index Terms: ***Breast Cancer, Classification, Machine Learning, Mammograms.***

I. INTRODUCTION

Machine Learning is implausible and powerful tool which solves hard-hitting classification problems. In Today's modern-day Technology, ML is intended for Image Recognition, Speech Recognition, Medical Diagnosis, and Statistics. It has gained an expeditious popularity in health sector all over the world and primarily used to identify or diagnose disease. In health sector approximately $100 billion investment is done for taking enhanced and suitable decisions using ML concept. It provides security, identify sign of harm and helps in increasing proficiency of clinical trials. The physicians can expedite their decision making and quickly choose the healthier medications for the patients by spotting or identifying disease using ML. Using ML in health care can reduces mortality rate, hospital length-of-stay, re-admissions and predicts chronic diseases, propensity-to-pay, no-shows. Trained algorithms can be used to identify abnormalities in images, and point areas that need attention, thus improving the accuracy of these processes. A well trained algorithm used to speed up diagnosis process and acts as collaborative partner to Radiologists in identifying abnormalities in explicit areas, and illumes noise. Breast Cancer is direful disease after skin cancer, it is most common in women than men. The elements which leads to Breast Cancer Includes Age, Alcohol consumption, Genes, Obesity, Radiation exposure. The most common indications of Breast Cancer affected patients includes change in size, appearance of Breast and change in skin color over Breast. According to Statistics, in 2018 approximately 330,080 breast cancer cases are anticipated to be spotted in Women. In lifetime on an average 1 in 8 women is diagnosed to breast cancer. Breast cancer occurs in diverse regions of breast like and in-between, ducts and lobules. There are diverse types of breast cancers, which are often divided as non-invasive breast cancer and invasive breast cancer. Non-invasive breast cancer is a cancer which doesn't spread outside breast and progresses in breast ducts. Most common type of cancer is Invasive breast cancer which progresses in breast ducts cells and spreads outside the breast. Usually Breast screening is done through physical exam and further perform diagnostic tests which includes Mammogram and Ultrasound. Mammogram is preliminary step in screening test which is mostly used in early stage to diagnose cancer for a person who doesn't have any warnings and symptoms to find out abnormalities. Normally abnormalities are classified as Benign (non-cancerous) and malignant. Ultrasound for producing detailed image and screening tumor in breast uses high frequency sound waves. After perceiving abnormality then diagnosing is done using Biopsy or Breast MRI. In Biopsy a sample of tissue from lump or mass is removed to determine whether cancerous or non-cancerous. A Breast MRI is a typical screening tool used for women having higher risk. Breast cancer treatment depends on tumor size and cancer stage. Surgery is most prevalent treatment used for breast cancer, in addition to surgery a patient undergo other treatment like chemotherapy, radiation or hormone therapy. Using machine learning in diagnosing breast cancer helps in improving diagnosis of cancer, quality of care and unnecessary biopsies.

Revised Manuscript Received on 22 May 2019.
***** Correspondence Author
S. Vasundhara, Department of Computer Science Engineering, VNR Vignana Jyothi Institute of Engineering & Technology, India.
B.V.Kiranmayee, Department of Computer Science Engineering, VNR Vignana Jyothi Institute of Engineering & Technology, India.
Chalumuru Suresh, Department of Computer Science Engineering, VNR Vignana Jyothi Institute of Engineering & Technology, India.

II. LITERATURE SURVEY

Anuj Kumar Singh.*et.al* [1] has discussed image pre-processing techniques such as averaging and thresholding techniques for detecting breast cancer and obtained region boundaries using image gradient techniques. Poonam Sonar.*et.al* [2] proposed classification model of Mammograms using hybrid SVM-KNN, Region of Interest from mammograms is extracted using Active Counter technique and Fuzzy C Means Clustering. GLCM (Gray level co-occurrence method) is used for feature extraction in ROI. Classification of breast tissues to normal or abnormal is done by training extracted features with hybrid SVM-KNN classifier.

M.A. Al- masni.*et.al* [3] recommended procedure of using deep learning techniques for Computer-aided diagnose (CAD) .For Pre-processing Multi-threshold peripheral equalization techniques are used to remove compression effect and resized. For feature Extraction, finding mass locations ROI-based Convolutional Neural Network is used and classification of mammograms is done into benign and malignant. G. R. Jyothilakshmi.*et.al* [4] projected a methodology for detecting mass abnormalities and classification. In this approach Median filter is used for noise removal, Global threshold for Feature Extraction, Gray-Level Co-occurrence Matrices (GLCM) for Feature analysis and for classification Support Vector Machines (SVM) classifier is used on Features acquired from GLCM technique. Xiaofei Zhang.*et.al* [5] intended mammogram prototypical classification on tomosynthesis images. In this model Convolutional Neural Network is used for classification of mammograms. C.Abrami, R.Harikumar.*et.al* [6] projected model for detecting digital mammograms micro calcification. Discrete wavelet Feature is used for Feature Extraction on mammogram data and classification is done by neural network using Radial basis Aidarus .M. Ibrahim.*et.al* [7] proposed a solution for mammogram classification using shearlet Transform. Using shearlet transform Features like mean, standard deviation, entropy and measure of uniformity is extracted and subjected to Kernel Principal Component Analysis (KPCA) for dimensionality reduction and KNN classifier (K- Nearest Neighbor) is used for feature classification on extracted features from mammograms.

A.S. Pitchumani Angayarkanni.*et.al* [8] projected a model for MRI classification. In this approach MRI pre-processing is done using Gabor Wavelet feature and histogram equalization. Gray Level Co-occurrence matrix (GLCM) is used for Texture based segmentation. For classification Decision Tree is used on extracted texture features of MRI data. Naglaa S. Ali Ibrahim.*et.al* [9] proposed a model for image segmentation. In this model Noise removal is done using 2D median filter. Morphological operations is performed and enhances contrast image using Band Limited Histogram Equalization (BLHE) method. Region of interest in mammograms is detected using Otsu's N thresholding.

N.M.Sangeetha.*et.al* [10] recommended an approach for abnormality detection using mammogram. In this approach shape extraction is done by applying manual threshold, morphology operations like Dilation and Erosion. Boundary extraction is performed by applying threshold using clustered method. So by extracting features mammograms are classified to benign and malignant.

Chiranji Lal Chowdary.*et.al* [11] proposed texture feature based classification. In this method intuitionistic fuzzy histogram, fuzzy c-mean is used for image pre-processing. Spatial gray level dependence for feature extraction and classification is done using Support Vector machine, K-nearest neighbor and rough set data analysis. Abhishek Midya.*et.al* [12] worked on mammograms masses classification using multi resolution. In this work feature extraction is done by deriving two-angle co-occurrence matrices using Haar-wavelet and classification using artificial neural network. Joaquim C. Felipe.*et.al* [13] proposed a method for shape retrieval and classification on mammograms. For shape description Zernike moments is used, to identify relevant attributes statistical association rules are used and to reduce dimensionality fractal theory is used-nearest neighbor algorithm is used for classification on mammograms.

S. Deepa.*et.al* [14] proposed procedure for Mammogram image classification. Dual Tree Complex Wavelet Transform (DT-CWT) for image decomposition and statistical features are extracted and for classification SVM classifier is used in this procedure. Hyun I. Kim.*et.al* [15] implemented a model for Mammogram image classification. In this model Harris Corner Detection is used for image pre-processing, SVM classifier is used for classification.

Elaheh Mahraban Nejad.*et.al* [16] proposed a technique for classification of Histopathological images. In this technique histopathological images are first subjected to pre-processing then are classified two benign and Malignant using Convolutional Neural Network.

Johanna Johnsi Rani G.et.al [17] has proposed methodology for Breast cancer classification and prediction. In this method the system classifies parameters Tumor, Lymph node and Metastases using Natural Language Processing (NLP) and Information Extraction (IE) techniques .Decision tree, Naïve Bayes, LAD Tree and Random Forest are applied on extracted features for classification.

Mohammad Taheri.*et.al* [18] projected a methodology for Breast cancer classification. In this method as part of image pre-processing noise removal is done using adaptive median filter and unwanted objects are removed using Harris Corner Detection and classification of extracted features is done by SVM. Basak Oztan.*et.al* [19] proposed methodology for Breast cancer grade classification. In this approach Fourier based image registration,

3D watershed segmentation is used to identify ductal structures and for classifying extracted structures Support Vector Machines (SVM) classifier is used. Chulwoo Pack *.et.al* [20] projected screening Breast Cancer model using dynamic learning. In this model by extracting features by correlation score and tissue tends to show conductivity and permittivity value. Classification on extracted features are performed using MLP using DLR and Conventional model.

III. PROPOSED METHOD

Radiologists encounter stressful and time consuming task while diagnosing breast cancer for getting better means by catching true cancer and reducing false positives where misclassifications leads to unnecessary biopsies. So to overcome this obstacle the proposed method is to use machine learning algorithms to instinctive classifications of mammograms and predict breast cancer .In this method various Algorithms like Convolutional Neural Network, Support Vector Machine and Random Forest tree are used on digital mammograms. In the first phase, Initially image pre-processing is done on image data using Gabor Wavelet transform Filter, Global thresholding and morphological operations on image data and further processed data is used for training classification algorithms and classified into benign, malignant and normal and performance of algorithm is calculated. The below figure represents the proposed work.

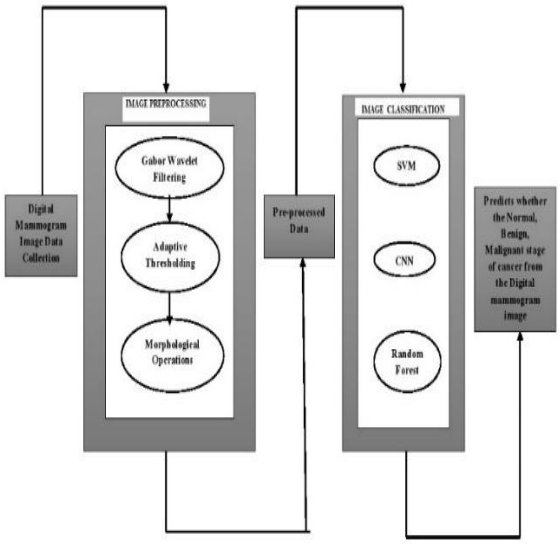

Fig 1: Breast Cancer Prediction Approach.

A. Image Pre-processing

Pre-processing is considered as vital step and the decisive goal of this technique is image quality improvisation by suppressing uninvited noise. Mammograms interpretation is difficult when compared with other medical images hence pre-processing is essential. Redundancy and optical inspection reliability is increased by Image pre-processing.

a) Gabor-wavelet Filter:

Gabor wavelets are developed by Dennis which are closely related to Gabor filters. In wavelet Transformation, image information is mentioned by only using shape of wavelet. Gabor wavelets are produced by from dilating one atom and provide complete image representation. Gabor filters are typically used in feature extraction, disparity extraction, texture analysis by distinct classes of band pass filters i.e. they allow a definite frequency band and discard others. The equation of Gabor filter is as follows.

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right) \quad (1)$$

Gabor filter when applied to mammogram image outcomes utmost response at edges and at points where texture changes. The following images represents before and after filter transformation when applied on mammogram image.

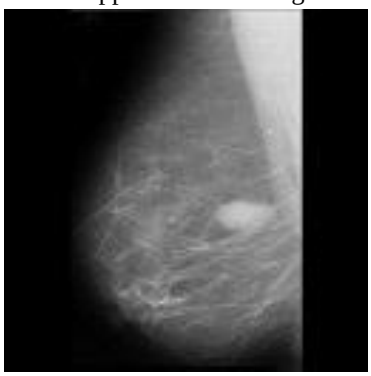

Fig: 2.1 Mammogram before Gabor filter

Fig: 2.2 Mammogram after Gabor filter

b) Adaptive Global threshold:

In image processing to acquire binary image from grayscale images Threshold is used. Adaptive threshold is seemed to be efficient when equaled to conventional thresholding technique. Global threshold sets all pixel to black in image. The important task is to decide adaptive threshold for all images. In this approach, for all images adaptive threshold value applied is 120. A grayscale image consists of pixel range from 0 to 255.

In Global thresholding each pixel is compared with threshold value. If value is greater it is assigned one value i.e. it refers to foreground. If value is lesser than threshold it is assigned zero value i.e. it refers to background. The following images represents mammogram image after applying adaptive threshold on filtered image.

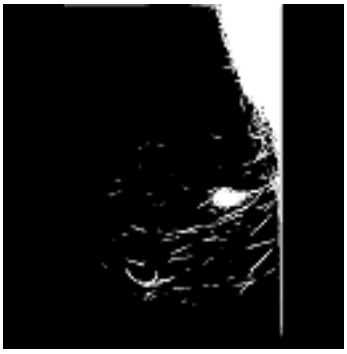

Fig 3: Mammogram after Adaptive Global threshold.

c) Morphological Operations:

Morphological operation is acquired by performing logical operation on the pixels enclosed by structuring element over the binary image by middling over image pixel at some point. Erosion and Dilation are basic morphological operations. The extensively used operations are opening and closing by integrating Erosion and Dilation filtering tasks.

- **Opening** is used to confiscate noise in images and streamlines image by rounding corners where kernel uses fits from inside. The mathematical representation of opening process is represented as follows

$$X \circ B = (X \ominus B) \oplus B \quad \textbf{(2)}$$

Where X- Input image,
B- Structuring image.

- **Closing is** contradictory of opening operation which has same effect of opening operation. It smoothens contours and maintain shapes and sizes of objects by padding small gaps and holes in a single – pixel object. The mathematical representation of closing process is represented as follows

$$X \circ B = (X \oplus B) \ominus B \quad \textbf{(3)}$$

Where X- Input image,
B- Structuring image.

The following figure represents results of mammogram image after performing opening and closing operations on image.

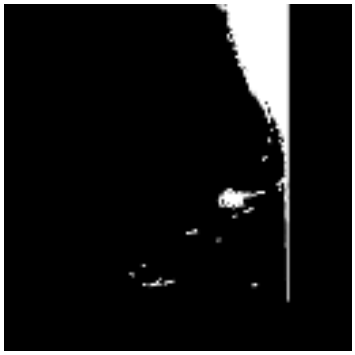

Fig 4.1: Mammogram after Opening operation.

Fig 4.2: Mammogram after closing operation

B. Image Classification

a) Convolutional Neural Network(CNN):

The Convolutional Neural Network moreover known as ConvNet is a class of deep neural networks, widely used in image processing, Natural Language Processing, pattern recognition, image recognition and image classification. CNN entail input layer, output layer and multiple hidden layers. CNN hidden layers classically contain convolutional layers, ReLU layer, pooling layers, activation function, fully connected layers and normalization layers. The CNN used in this approach involves 4 convolutional layer and 4 pooling layers. The input images are of size 64 × 64 resolution. In layer 1 convolutional2D is of 32- 3 ×3 filters, in layer 2 convolutional2D is of size 64- 3 × 3 filters. In layer 3 convolutional2D is of size 96- 3×3 filters. In layer 4 convolutional2D is of size 96- 3×3 filters and additionally there are 4 max pooling layers each of size 2×2. After flattening images are sent to dense layers. The following structure represents Convolutional Neural Network architecture for implementation python package keras is used with backend tensor flow. The pre-processed data has driven into 80 % training data and 20% testing data. CNN classifier is trained using training data. Trained classifier uses test set to classify mammograms. The output of classifier is categorical label with 0 for Benign, 1 for Malignant and 2 for Normal. Performance is calculated using accuracy metrics.

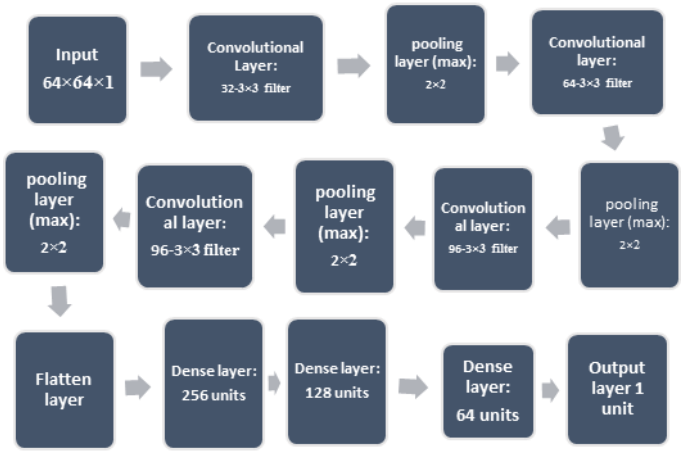

Fig 5: Convolutional Neural Network Architecture

b) Random Forest:

Random Forest is supervised classification algorithm and a considerable bagging Technique. Random forest trees Performance is modest to train and tune and much related to boosting. The key concept of bagging is reduce the variance by middling noisy unbiased models. In Random forest, each tree in the ensemble is built by replacing training set sample .Open source python library Scikit – learn is used in this process, the pre-processed data is alienated into 80% training set and 20% test set. A random forest (RF) classifier of 120 estimators is trained using training data. Trained classifier is used to classify test data mammograms. The output of Random Forest classifier is a categorical label, with 0 for Benign, 1 for Malignant and 2 for Normal. Performance is calculated using metrics Precision, Recall and F1-score.

c) Support Vector Machines (SVM):

Support vector machines (SVM) is extension of non-linear models and one of the supervised classification method. It is used in text categorization, outlier detection and image classification. The objective of SVM is to find distinctly classifying data points using N-dimensional hyper plane. In classification of data points hyper planes act as decision Boundaries. For classification SVM generates hyper planes set in a high or infinite dimensional space. In this approach SVM of linear kernel is implemented using python library (scikit-learn). The pre-processed data is split in ratio 80:20 for Training and testing. SVM classifier is trained using training set. The validation of trained classifier is done by classifying test set. The output of SVM classifier is multiclass with 0 for Benign, 1 for Malignant and 2 for Normal images. Accuracy metrics Precision, Recall and support used for performance calculation.

I. RESULTS

Evaluating performance of models is one of the most essential and final step in building a model. In this approach Precision, Recall, F1-score, support, Accuracy score are considered as Performance metrics for evaluating model. For performance evaluation the dataset is split into training and testing initially, after that values are predicted using test set on trained model. The accuracy score of model is calculated by considering actual result and predicted value. Mathematical form for calculating accuracy score is as mentioned below.

$$Accuracy = \frac{TP+TN}{Total} \quad (4)$$

Precision is percentage of relevant results mathematical form is as mentioned below. It is fraction of true positives to Actual Results.

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

Recall is percentage of total relevant results correctly classified by trained model. It is fraction of True positives to Predicted Results. Mathematical form for calculating Recall is as shown below.

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

Harmonic mean of Recall and Precision is **F1-score**.The Mathematical form for calculating F1-score is as shown below.

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (7)$$

Support parameter represents number of occurrences of each class. From collected Digital Mammogram Images train and test data split is 80% and 20%. Accuracy score when evaluated on trained models CNN models results with 98% accuracy score. Random Forest results with 82.5% and SVM results with 80.5%.The following figure represents accuracy analysis of 3 classifiers.

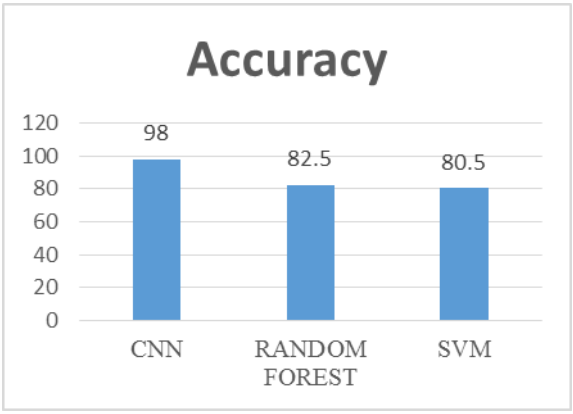

Fig 6: Performance Analysis of CNN, Random Forest and SVM classifier

To evaluate SVM classifier performance metrics Precision, Recall, F1-score and support are used for each class. The following figure represents Performance evaluation for SVM model.

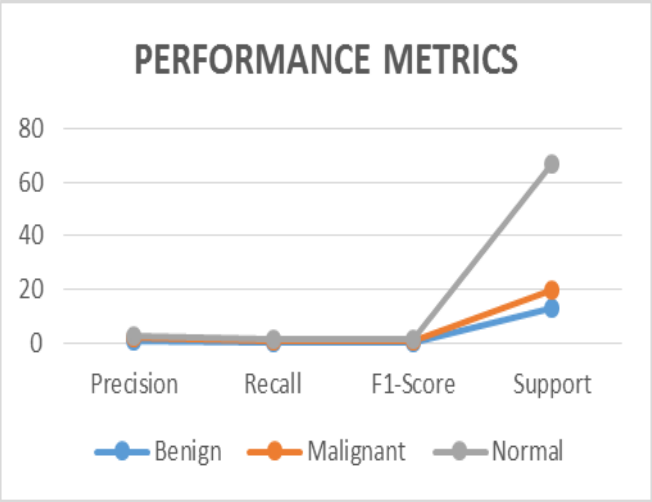

Fig 7: Performance metrics for SVM

The following table represents Performance metrics Precision, Recall, F1-score and Support values on trained SVM classifier

Precision	Recall	F1-Score	Support
1	0.38	0.56	13
1	0.29	0.44	7
0.78	1	0.88	47

Table I: Performance metrics of SVM classifier

To evaluate Random Forest classifier performance metrics Precision, Recall, F1-score and support are used for each class. The following figure represents Performance evaluation for Random Forest model.

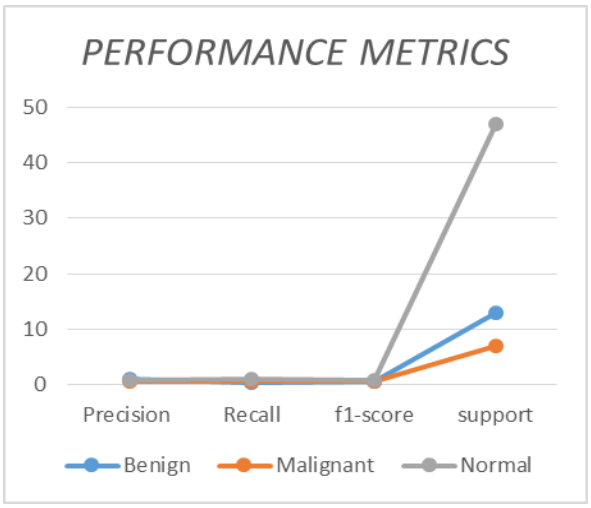

Fig 8: Performance metrics for Random Forest

The following table represents performance metrics Precision, Recall, F1-score and Support values of trained Random Forest classifier.

Precision	Recall	F1-score	Support
1	0.38	0.56	13
0.57	0.57	0.57	7
0.84	0.98	0.90	7

Table II: Performance metrics of Random Forest Classifier

II. CONCLUSION

In this Paper, Digital mammogram classification has successfully implemented using machine learning algorithms. The accuracy of classifiers of CNN, SVM and Random Forest is 98%, 80.5%, 82.5% respectively. The better classification results on CNN upon pre-processed mammogram. By this approach it has proved that CNN results in instinctive classification of digital mammograms using filtering and morphological operations. Therefore this reliable technique can be used for diagnosing breast cancer and decrease misclassifications using mammograms.

REFERENCES

1. Anuj kumar Singh , Bhupendra Gupta , 2015 , "A Novel Approach for Breast Cancer Detection and Segmentation in a mammogram", In Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015)
2. Poonam Sonar, Udhav Bhosle, Chandrajit Choudhury, 2017, "Mammography Classification Using Modified Hybrid SVM-KNN", In International Conference on Signal Processing and Communication -2017.
3. M.A. Al-masni, M.A. Al- antari, J. M. Park, G.Gi, T.Y.Kim, P.Rivera, E.Valarezo, S-M.Han, and T.-S. Kim, 2017, "Detection and Classification of the Breast Abnormalities in Digital Mammograms via Regional Convolutional Neural Network", In IEEE 2017
4. G.R. Jothilakshmi, Arun Raaza, 2017, "Effective Detection of Mass Abnormalities and its Classification Using Multi- SVM Classifier with Digital Mammogram Images" In IEEE International Conference on Computer, Communication and Signal Processing (ICCCSP-2017)
5. Xiaofei Zhang , Yi Zhang, Erik Y. Han, Nathan Jacobs, Qiong Han,Xiaoqin Wang , Jinze Liu, 2018, "Classification of whole mammogram and tomosynthesis images using deep convolutional neural networks", IEEE 2017
6. C. Abrami , R. Harikumar , S.R. Sannasi Chakravarthy , 2016 , "Performance Analysis and Detection of Micro calcification in Digital Mammograms using Wavelet Features". In IEEE 2016.
7. Aidarus M.Ibrahim and Baharum Baharudin, 2016, "Classification of Mammogram Images Using Shearlet Transform and Kernel Principal Component Analysis", In ICCOINS 2016.
8. A.S. Pitchumani Angayarkanni, B.Dr. Nadira Banu Kamal, 2012, "MRI Mammogram Image Classification Using ID3 Algorithm", In IEEE 2012.
9. Naglaa S. Ali Ibrahim, Naglaa F. Soliman, Mahmoud Abdallah, Fathi E.Abd EI-Samie , IEEE, "An Algorithm for Pre-processing and Segmentation of Mammogram Images". In IEEE 2016.
10. N.M.Sangeetha, Dr.D.Pugazhenthi, 2016, "A Novel Shape and Boundary Extraction Image Processing Technique for Detecting the Breast Abnormalities Using Digital Mammogram", In 2016 (IC-GET)
11. Chiranji Lal Chowdhary, D.P. Acharjya, 2016, "Breast Cancer Detection using Intuitionistic Fuzzy Histogram Hyperbolization and Possibilitic Fuzzy c-mean Clustering algorithms with texture feature based Classification on Mammography Images". In ACM 2016.

12. Abhishek Midya, Jayasree Chakraborty 2015, "Classification of Benign and Malignant masses In Mammograms Using Multi-Resolution Analysis of Oriented Patterns".
13. Joaqium C. Felipe, Marcela X. Ribeiro, Elaine P.M. Sousa, Agma J.M.Traina, Caetano Traina Jr, 2006, "Effective Shape- based Retrieval and Classification of mammograms". In ACM 2006.
14. S. Deepa, Dr.V. Subbaiah Bharathi, 2012, "An Efficient Digital Mammogram Image Classification using DTCWT and SVM", In ACM 2012.
15. Hyun I.Kim, Sung Shin, Wei Wang, Soon I.Jeon , 2013, "SVM-based Harris Corner Detection for Breast Mammogram Image Normal/Abnormal Classification", In ACM 2013.
16. Elaheh Mahraban Nejad, Lilly Suriani Affendey, Rohaya Binti Latip, Iskandar Bin Ishak,2017, "Classification of Histopathology Images of Breast into Benign and Malignant using a Single-layer Convolutional Neural Network", In 2017 ACM.
17. Johanna Johnsi Rani G, Dennis Gladis, Joy Mammen, 2015, "Classification and Prediction of Breast Cancer Data derived Uisng Natural Language Processing", In 2015 ACM.
18. Mohammad Taheri, George Hamer, Seong Ho Son, Sung Y.Shin, 2016, "Enhanced Breast Cancer Classification with Automatic Thresholding Using SVM and Harris Corner Detection", In 2016 ACM.
19. Basak Oztan, Lindsey M.Polizzotti , Cemal Cagatay Bilgn, Kira M.Hederson, George E.Plopper,Bulent Yener, 2011, "Classification of Breast Cancer Grades through Quantitative Characterization of Ductal Structure Morphology in Three-Dimensional Cultures".
20. Chulwoo Pack, Sung Shin, Seong Ho Son, Soon IK Jeon, 2015, "Computer Aided Breast Cancer Aided Breast Cancer Diagnosis System with Fuzzy Multiple-parameter Support Vector Machine", In 2015 ACM.

AUTHORS PROFILE

S. Vasundhara, currently pursuing Master of Technology in department of Computer Science and Engineering in **VNR** Vignana Jyothi Institute of Engineering and Technology affiliated to JNTU Hyderabad. Completed Bachelor of Technology in department of Computer Science and Engineering from JNTU Hyderabad. Her work is related to Machine learning.

B.V.Kiranmayee, Associate Professor in Department of Computer Science and Engineering, VNRVJIET, Hyderabad .she is currently pursuing Ph.D. degree in Computer Science and Engineering in JNTUH University. She has 17 years of teaching experience. Her specialized area of research is Data Mining.

Chalumuru Suresh, currently working as Assistant Professor in the department of CSE at VNR Vignana Jyothi Institute of Engineering and Technology affiliated to JNTU Hyderabad. He is having good experience in teaching and research. His research areas include Data Mining, BigData Analytics and Machine Learning