# Resnet Based Feature Extraction with Decision Tree Classifier for Classificaton of Mammogram Images

**T. Sathya Priya[a], and Dr. T. Ramaprabha[b]**

[A]
 Research Scholar, Department of Computer Science, Vivekanandha College of arts
and science for women, (Autonomous), Namakkal, India.
[b]Professor, Department of Computer Science, Vivekanandha College of arts and
science for women, (Autonomous), Namakkal, India

_____

**Abstract:** Right now, breast cancer is considered as a most important health problem among women over the world. The detection of breast cancer in the beginning stage can reduce the mortality rate to a considerable extent. Mammogram is an effective and regularly used technique for the detection and screening of breast cancer. The advanced deep learning (DL) techniques are utilized by radiologists for accurate finding and classification of medical images. This paper develops a new deep segmentation with residual network (DS-RN) based breast cancer diagnosis model using mammogram images. The presented DS-RN model involves preprocessing, Faster Region based Convolution Neural Network (R-CNN) (Faster R-CNN) with Inception v2 model based segmentation, feature extraction and classification. To classify the mammogram images, decision tree (DT) classifier model is used. A detailed simulation process is performed to ensure the betterment of the presented model on the Mini-MIAS dataset. The obtained experimental values stated that the DS-RN model has reached to a maximum classification performance with the maximum sensitivity, specificity, accuracy and F-Measure of 98.15%, 100%, 98.86% and 99.07% respectively.

**Keywords:** Breast cancer, Classification, Segmentation, Feature extraction, Mammogram

_____

## 1. Introduction

Breast cancer is a familiar type of disease that exists in female. Usually, breast cancer consumes maximum time for developing and signs expressed latter. There is no medicine for complete recovery of cancer, but the lifetime can be extended only if it is detected in the earlier stage. Hence, the previous prediction of breast cancer is suggested by America Cancer Society (ACS) which has stated that, the screening test is highly essential for extending the lifetime [1]. Recently, digitalized diagnostic systems widely applied mammogram screening models for classifying the breast lesions. Typically, Computer Aided Diagnosis (CAD) model depends upon Machine Learning (ML) methods for detecting tumors present in computerized mammogram images. These methods should be determined with diverse and descriptive features for classifying images into several classes.

Several developers have proposed mammogram images for 2-class (normal and abnormal) classification and accomplished effective simulation outcome. Mazurowski et al. [2] projected a template on the basis of prediction model for breast tumor. The data set depends upon massive Digital Database for Screening Mammography (DDSM) images and attained highest accuracy. Wei et al. [3] projected a relevance feedback learning approach and perform classification with the application of SVM radial kernel using a data set of enormous photographs. Tao et al. [4] related the function of 2 classification models, termed as curvature scale space as well as local linear embedded metric with the application of a database of and accuracy of 2 classifiers. Abirami et al. [5] employed a wavelet features for 2-class classification of digital mammograms which has gained maximum accuracy for Mammographic Images Analysis Society (MIAS) data set.

Elter and Halmeyer [6] processed classification under the application of Artificial Neural Network (ANN) and Euclidean metric classification, correspondingly, and attained a performance to greater extent. The developers have applied 2-class classification; however 2-class classification is insufficient to eliminate unwanted biopsy as in abnormal cases the tumor might be benign or malignant. Suckling [7] presented Extreme Learning Machine (ELM) model for classifying mammograms of MIAS database. The newly developed has surpassed the alternate models using the similar database. Jasmine et al. [8] carried out 2-class classification with the projected model on the basis of wavelet analysis using Artificial Neural Network (ANN). This process is computed under the application of MIAS database and reached better accuracy. Xu et al. [9] related the function of 3 NNs and recommended Multilayer Perceptron (MLP) function as count of features enhanced. This model accomplished accuracy to a greater limit with the application of mammogram images.

In last decades, Deep Learning (DL) under the application of NN which is facilitated as state-of-the-art outcomes in massive computer vision models, like object prediction as well as classification. DL methods are used in diverse clinical imaging applications like tissue classification in histopathology as well as histology images. Therefore, limited studies are accessible under the application of DL for mammogram images categorization. In [10], Convolutional Neural Network (CNN) have been applied for segmenting the breast tissue

of mammographic texture. Multi-scale features as well as auto-encoders (AE) have been used for determining breast density measure. CNNs are employed for classifying micro-calcifications; however, the data set is small sized. Mert et al. [11] developed a radial basis function neural network (RBFNN) with independent component analysis (ICA) for 2-class categorization. A maximum accuracy has been accomplished on WBDC data set with enormous images. In recent times, for 2-class classification, Dheeba et al. [12] applied particle swarm optimization (PSO) related based Wavelet Neural Network (PSO-WNN) as well as deep belief network (DBN), correspondingly, and attained effective results on data set with images.

This paper develops a novel deep segmentation with residual network (DS-RN) related breast cancer analyzing method with the help of mammogram images. The projected DS-RN method is composed of preprocessing, Faster Region based Convolution Neural Network (R-CNN) (Faster R-CNN) with Inception v2 model based segmentation, feature extraction as well as classification. For classifying the mammogram images, decision tree (DT) and random forest (RF) classification methods have been employed. A brief simulation process is carried out to approve the advancement of the proposed method on the Mini-MIAS dataset.

2. Proposed method

Fig. 1 reveals the processes involved in the presented model. As depicted in figure, the input mammogram image is preprocessed and then segmented using Faster RCNN model. Besides, ResNet model is applied as a feature extraction model to determine the useful set of feature vectors from the segmented image. At last, the classification process is done by DT model.

2.1. Preprocessing

In this method, the preprocessing stage is applied to improve the result of classification process. First, the input image is fed into mean shift filtering approach to avoid the noise from the image. Next, thresholding is performed to transfer the extracted image to binary format. Then, a contour drawn phase is done to retrieve the contours over the objects in images. Afterwards, the higher contour mask is applied to maintain high sized object by mask generation. Besides, the noise in mask is removed and computes the contrast enhancement using Contrast Limited Adaptive Histogram Equalization (CLAHE) approach. At last, contrast is enhanced in which preprocessed image is fed to image segmentation.

2.2. DL based Segmentation Process

Here, a DL based Faster RCNN with Inception v2 method is utilized for classifying as well as preprocessed images. Initially, this model undergoes training with human modeled images for training the ROI. Based on the training phase, segmentation model has been utilized to find the affected area on new test image. RCNN is considered as an object prediction method in which it is developed using 2 objectives. Deep fully convolutional network is assumed as first stage which means the regions whereas Fast R-CNN predictor utilizes the previous regions. In case of object prediction, the complete method is grouped in single network. Fast R-CNN approach transforms the data regarding RPN using NN units.
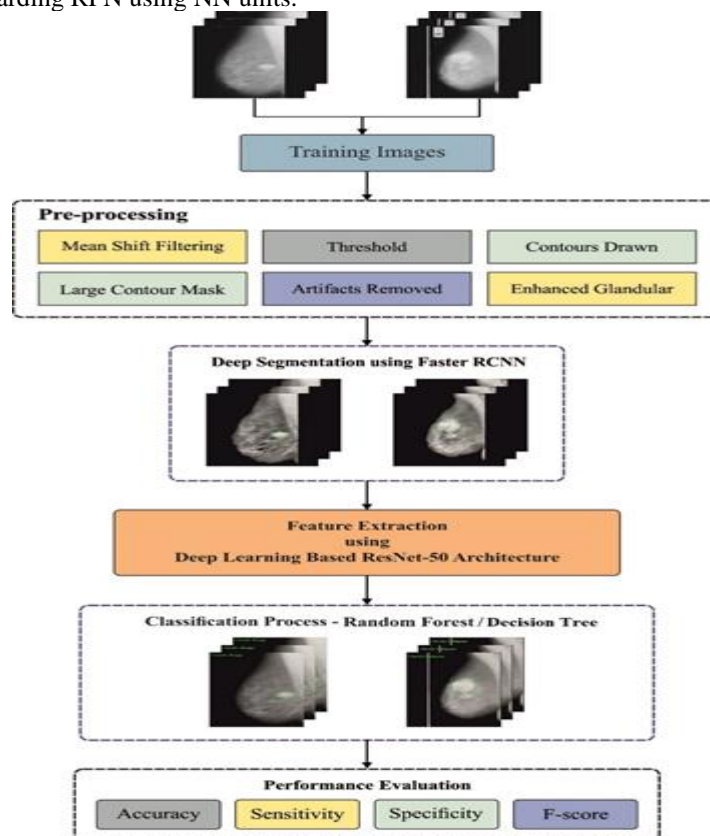


Fig. 1. Block diagram of presented model

2.3. ResNet 50 based feature extraction

CNN has completely dominated the vision space recently. A CNN is composed of input layer, output layer, and various hidden layers. The hidden layers of CNN usually contain convolutional layers, pooling layers, Fully Connected (FC) layers as well as normalization layers (ReLU). Excess layers are employed for tedious methodologies. The CNN structure has expressed tremendous function under various Computer Vision as well as ML issues. CNN computes the prediction as well as training operation in abstract level, with the remaining details given in the following subparagraphs. CNN method is applied widely in smart ML fields because of the processing record breaking efficiency. Linear algebra depends upon the performance of CNN. Matrix vector multiplication is central premises of how data and weights are displayed. The layers are composed of various characteristics for image set. For sample, when a face image is provided as input for CNN, the system would learn fundamental properties like edges, bright spots, dark spots, shapes and so forth. The consecutive set of layers is composed of shapes and objects which are related to image that can be recognized like Eyes, Nose and Mouth. The next layer is compose of factors which resembles as original faces, besides, shapes and objects of the network applies for defining human face. CNN maps the portions rather than entire image, and breaks the image classification process into tiny portions.

A grid is described to depict the features extraction by CNN for estimation. The following task is named as filtering, which lines the feature with image patch. The element is enhanced by using corresponding feature pixel, and the process is completed and classified by overall count of pixels in feature space. The consequent value for a feature is fixed within the feature patch. This operation is followed by residual feature patches and attempts in all feasible match- repeated field of filter that is named as a convolution. The subsequent layer of a CNN is assumed as "max pooling", that contributes in diminishing the image stack. For pooling an image, window size should determine, the stride has to be described. The window is filtered over the image in strides, along with max value which is saved for every window. Max pooling limits the dimensionality of a feature map while maintaining the significant data. The normalization layer of a CNN is computed as Rectified Linear Unit (ReLU), which contributes the negative values inside the extracted image to 0.

This is followed for all filtered images; the ReLU layer enhances non-linear features of a method. The next step by CNN is stacking the layers, thus the resultant of a layer is considered as input of subsequent layer. Layers which are repeated for "deep stacking". The last layer inside the CNN structure is named as FC layer is named as the classification method. FC layers are stacked jointly, with middle layer voting on phantom "hidden" classes. Obviously, the excess layer enables the network for learning better integrations of features for making effective solution. The measures are employed for convolution layer and weights for FC layers were attained by using backpropagation (BP) that is processed by Deep Neural Network (DNN). BP in NN applies the error in last solution for computing the modifications which exist in the system.

| Conv_Layer 1 | Conv_Layer 2 | Conv_Layer 3 |
|---|---|---|
| 7 × 7, 64, stride 2 | 3 × 3, maxpool, stride 2 | 3 × 3, maxpool, stride 2 |
| Output Size: 112 × 112 | $\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 3, 256 \end{bmatrix} \times 3$ | $\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 3, 512 \end{bmatrix} \times 4$ |
| | Output Size: 56 × 56 | Output Size: 28 × 28 |

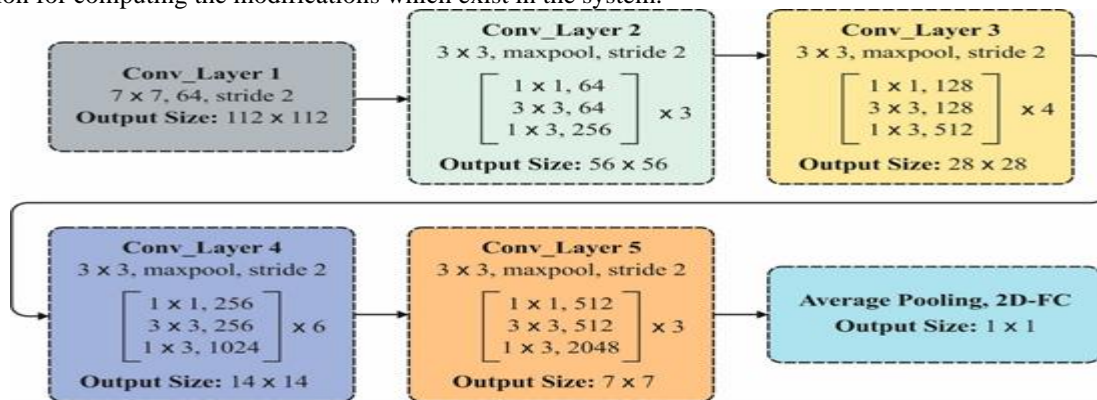| Conv_Layer 4 | Conv_Layer 5 | Average Pooling, 2D-FC |
|---|---|---|
| 3 × 3, maxpool, stride 2 | 3 × 3, maxpool, stride 2 | Output Size: 1 × 1 |
| $\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 3, 1024 \end{bmatrix} \times 6$ | $\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 3, 2048 \end{bmatrix} \times 3$ | |
| Output Size: 14 × 14 | Output Size: 7 × 7 | |

Fig. 2. Layered Structure of ResNet-50

ResNet employs the residual block to resolve the decomposition and gradient disappearance issues which generally exists in general CNNs. The residual block is not based on the network depth, however enhances the function of a system. In particular, ResNet networks have accomplished best performance in ImageNet in classifying process. The function of residual function is given in the following:

$$y = F(x, W) + x \qquad (1)$$

where refers the input of residual block; denotes the weight; implies the result. The fundamental architecture of ResNet50 is depicted in Fig. 2.

2.4. Classification

Once the features are extracted from the ResNet-50 model, classification process takes place using DT model.

DT is an inductive learning method which provides classification tree under the application of training data and samples. It depends upon "divide and conquer" principle. It is defined as a non-parametric which is independent of properties of data distribution, hence applicable for embedding non-spectral data with classification principle for enhancing the class separability is attained. The final DT offers a representation of model which appeals to human as it renders classification process evidently. It is suitable for resolving classification issues using maximum number of classes and adjusted to manage the regression problems. DT

applies a hierarchical architecture in which all levels are used for attribute scores which have 2 results. For classifying an object, the root of tree is initialized, estimate the test, and apply the branch for best results. The process is repeated finding a leaf, where an object is organized. The final leaf is assumed as final outcome of following set of unique decision rules over the tree. The tree is extended until all training instances are classified exactly; and over fitting of data is eliminated by pruning the training dataset.

The structure of DT needs supervised training; so it is essential to apply training dataset with a response and explanatory variables. A classification architecture is determined by DT is evaluated from training data utilizing a statistical process. The nodes are where trees branch or divide the dataset; terminal nodes are known as leaves that include most homogeneous classes. When a training set , there is  count of classes  and a total of  cases, the expected data from such methods are,

$$info\ (T) = -\sum_{j-1}^{K} \frac{freq(Cj,T)}{|T|} \log_2 \left( \frac{freq(Cj,T)}{|T|} \right) \qquad (3)$$

where,  implies the probability of class  in training set . When separation the training set  in accordance with response variable  (e.g., NDVI), there can be 'n' count of cases. A usual data in separation is weighted sum over the subset as:

$$infox\ (T) = -\sum_{i=1}^{n} \frac{|Ti|}{|T|} info\ (Ti). \qquad (4)$$

The information gain (IG) by portioning training set following variable  (NDVI) is

$$gain\ (T) = info\ (T) - infox\ (T). \qquad (5)$$

The gain condition chooses a test to maximize the IG. However, it has a strong bias, which facilitates the tests with multiple results. It is rectified by establishing gain ratio , which is determined as:

$$gain\ ratio\ (X) = gain \frac{X}{split} info\ (X) \qquad (6)$$

where, split info  is the potential data created by splitting  into  subsets that is not relevant to classification as provided by:

$$split\ info\ (X) = -\sum_{i=1}^{n} \frac{|Ti|}{|T|} \log_2 \left( \frac{|Ti|}{|T|} \right). \qquad (7)$$

The gain ratio signifies the proportion of helpful data for classification. An experiment that maximizes the ratio, subjects to control huge IG is chosen. A training sample (root node) is separation following the revealed situation for creating branches. A few nodes, instance sharing in uniform distribution, which means a single LULC class in all groups and so, data content is higher. It is determined as 'leaves'.

3. Experimental Validation

For examining the final outcomes of the DS-RN method, overall count of 322 mammogram images has been employed from Mini-MIAS dataset [13]. It is composed of images acquired from 3 class labels such as benign, malignant and normal. The details about a dataset are offered in Table 1 and sample set of images are depicted in Fig. 3.

Table 1 Dataset Description

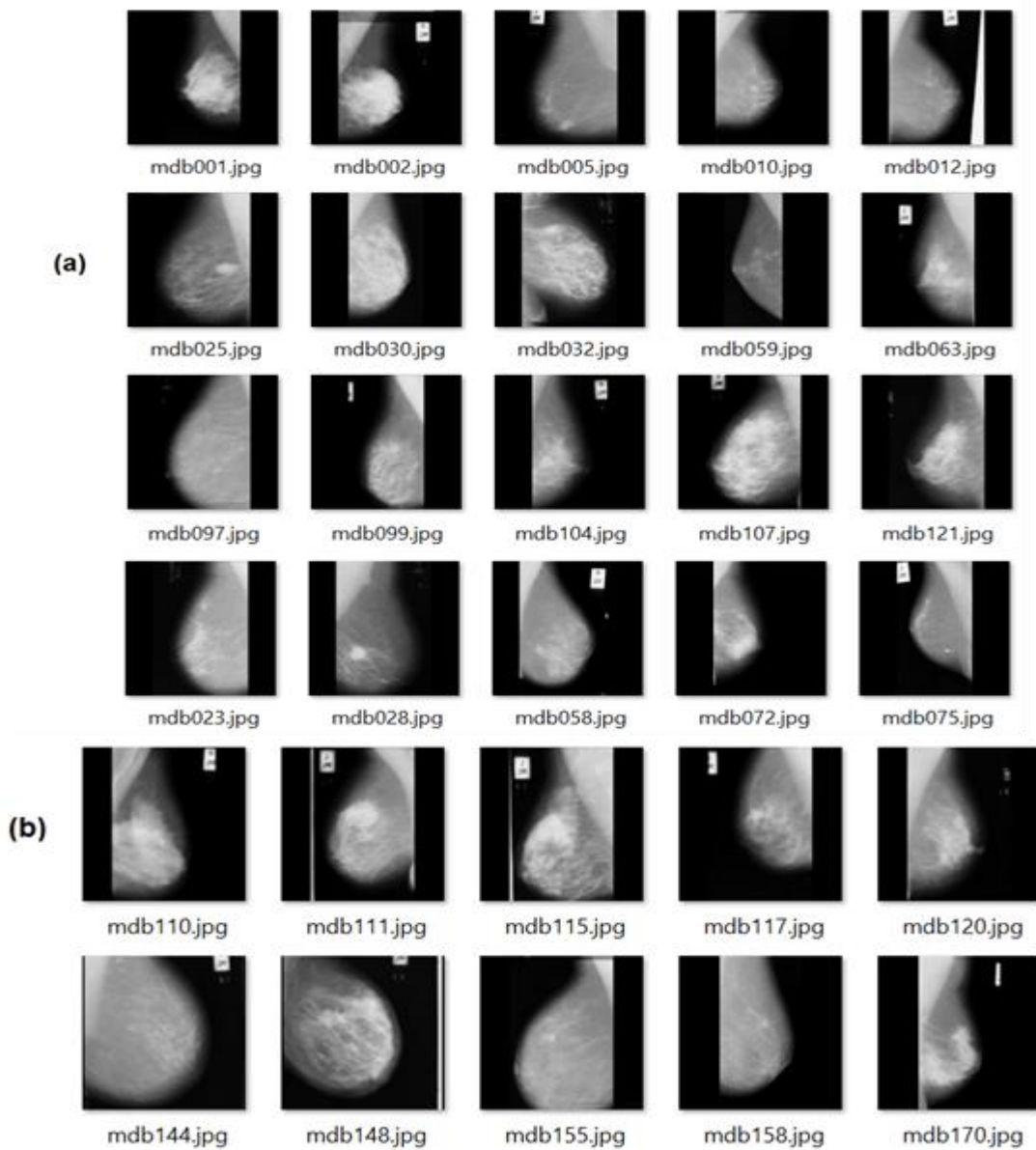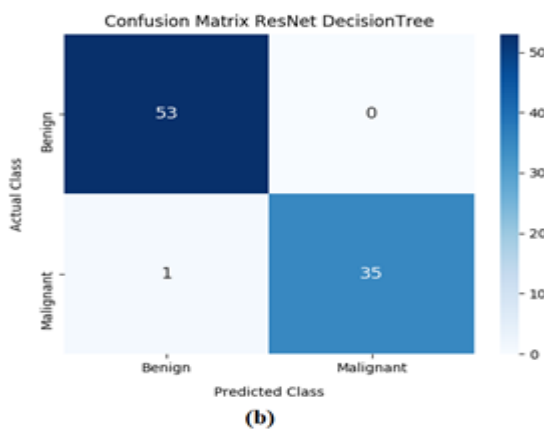| Description | Values |
|---|---|
| Total Number of Images | 322 |
| Number of Classes | 3 |
| Number of Images in Normal Class | 206 |
| Number of Images in Benign Class | 64 |
| Number of Images in Malignant Class | 52 |

Fig. 3. Sample Images a) Benign b) Malignant



Fig. 4. Confusion Matrix of Proposed DS-RNDT

Fig. 4 shows the confusion matrix generated by the proposed DS-RNDT model on the classification of mammogram images. The figure showed that the DT model has outperformed other models with the classification of 53 images as benign and 35 images as malignant.
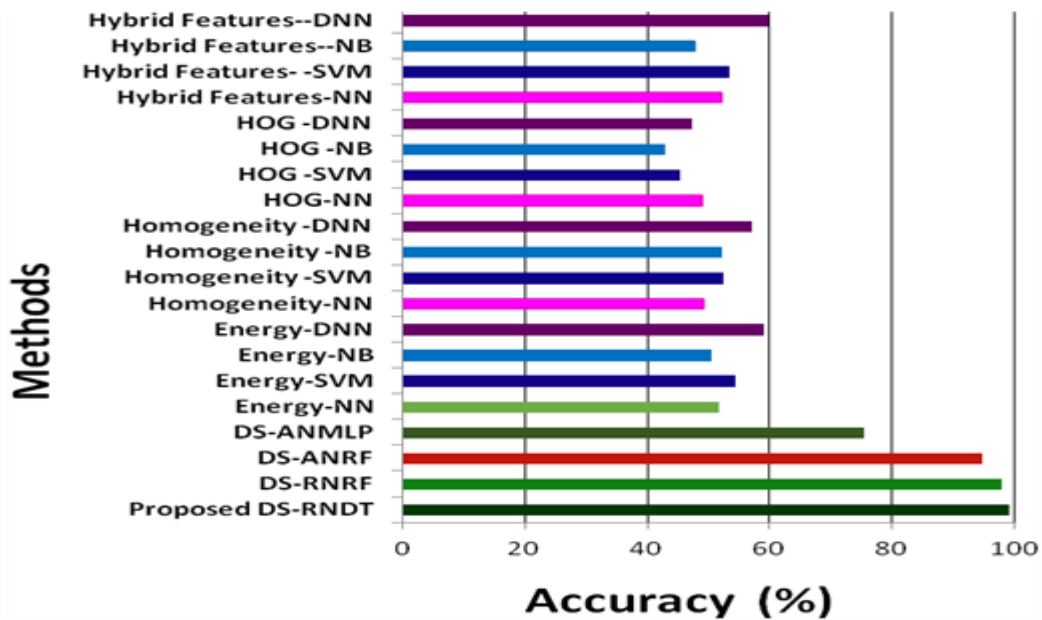
A thorough comparative study of the classification results of the DS-RNDT and DS-RNRF with former methods with respect to accuracy is showcased in Fig. 5. From the figure, it is apparent that the HOG-NB approach has defined poor classification result by accomplishing least accuracy value of 42.6%. Similarly, the HOG-SVM approach is considered to have moderate results across the HOG-NB scheme with the accuracy value of 45.2%. Additionally, the HOG-DNN scheme has displayed acceptable accuracy value of 47%. Followed by, the Hybrid Features-NB, HOG-NN, Homogeneity-NN and Energy-NB approaches have outperformed previous methods and attained nearby accuracy values of 47.6%, 49%, 49.2% and 50.2% correspondingly. Besides, the Energy-NN framework has attempted to demonstrate considerable outcomes by accomplishing higher accuracy of 51.4%. Next, the Homogeneity-NB and Hybrid Features-NN schemes have illustrated similar and acceptable accuracy value of 52%. Moreover, the Homogeneity -SVM and Hybrid Features-SVM technologies have ended up with reasonable and identical accuracy value 52.2% and 53.2%. Meantime, the Energy-SVM scheme has resulted in maximum accuracy to a considerable extent of 54.2%. Next, the Homogeneity-DNN, Energy-DNN as well as hybrid features-DNN approaches have showcased better result with the accuracy values of 56.8%, 58.8%, 59.6% correspondingly. Simultaneously, the DS-ANMLP and DS-ANRF approaches have illustrated qualified results by achieving the accuracy of 75.28% and 94.38%. Also, the newly projected DS-RNDT approaches have showcased supreme outcomes with the best accuracy of 97.75% and 98.86%.

## 4. Conclusion

This paper has developed a DL based segmentation with classification model named DS-RN model for breast cancer diagnosis model using mammogram images. Initially, the input mammogram image is preprocessed and then segmented using Faster RCNN model. Besides, ResNet model is applied as a feature extraction model to determine the useful set of feature vectors from the segmented image. At last, the classification process is done by DT model. A comprehensive simulation process is completed to confirm the betterment of the presented model on the Mini-MIAS dataset. The obtained experimental values stated that the DS-RN model has reached to a maximum classification performance with the maximum sensitivity, specificity, accuracy and F-Measure of 98.15%, 100%, 98.86% and 99.07% respectively. In future, the proposed model can be deployed in IoT and cloud based environment to assist telemedicine.

## References

1. *American Cancer Society, http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-detection.*

2. *M. A. Mazurowski, J. Y. Lo, B. P. Harrawood, and G. D. Tourassi, "Mutual information-based template matching scheme for detection of breast masses: from mammography to digital breast tomosynthesis," Journal of Biomedical Informatics, vol. 44, no. 5, pp. 815–823, 2011.*

3. *C.-H. Wei, Y. Li, and P. J. Huang, "Mammogram retrieval through machine learning within BI-RADS standards," Journal of Biomedical Informatics, vol. 44, no. 4, pp. 607–614, 2011.*

4. *Y. Tao, S. C. B. Lo, L. Hadjiski, H. P. Chan, and T. M. Freedman, "Birads guided mammographic mass retrieval," in Proceedings of the Medical imaging, Proceedings of SPIE, 2011.*

5. *C. Abirami, R. Harikumar, and S. Chakravarthy, "Performance analysis and detection of micro calcification in digital mammograms using wavelet features," in Proceedings of the International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET '16), pp. 2327–2331, Chennai, India, March 2016.*

6. *M. Elter and E. Halmeyer, "A knowledge-based approach to the CADx of mammographic masses," in Proceedings of the Medical Imaging 2008: Computer-Aided Diagnosis, vol. 6915 of Proceedings of SPIE, San Diego, Calif, USA, February 2008.*

7. *J. Suckling, "The mammographic image analysis society digital mammogram database," Exerpta Medica. International Congress Series, vol. 1069, pp. 375–378, 1994.*

8. *J. A. Jasmine, A. Govardhan, and S. Baskaran, "Microcalcification detection in digital mammograms based on wavelet analysis and neural networks," in Proceedings of the International Conference on Control, Automation, Communication and Energy Conservation (INCACEC '09), pp. 1–6, Perundurai, India, June 2009.*

9. *W. Xu, W. Liu, L. Li, G. Shao, and J. Zhang, "Identification of masses and microcalcifications in the mammograms based on three neural networks: comparison and discussion," in Proceedings of the 2nd International Conference on Bioinformatics and Biomedical Engineering (iCBBE '08), pp. 2299–2302, May 2008.*

10. *K. Petersen, M. Nielsen, P. Diao, N. Karssemeijer, and M. Lillholm, "Breast tissue segmentation and mammographic risk scoring using deep learning," in Breast Imaging: 12th International Workshop, IWDM 2014, Gifu City, Japan, June 29–July 2, 2014. Proceedings, vol. 8539 of Lecture Notes in Computer Science, pp. 88–94, Springer, Berlin, Germany, 2014.*

11. *Mert, N. Kılıç¸, E. Bilgili, and A. Akan, "Breast cancer detection with reduced feature set," Computational and Mathematical Methods in Medicine, vol. 2015, Article ID 265138, 11 pages, 2015.*

12. *J. Dheeba, N. Albert Singh, and S. Tamil Selvi, "Computer-aided detection of breast cancer on mammograms: a swarm intelligence optimized wavelet neural network approach," Journal of Biomedical Informatics, vol. 49, pp. 45–52, 2014.*

13. *http://peipa.essex.ac.uk/pix/mias/*