IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# A Framework on Classification of Mammogram Images for Breast Cancer Detection using Image Processing with Data Mining Techniques

**V.Punithavathi[1], Dr.D.Devakumari[2]**

[1]Research Scholar,
PG and Research Department of Computer Science
Government Arts College,
Coimbatore, Tamil Nadu, India.

[2]Assistant Professor,
PG and Research Department of Computer Science
Government Arts College,
Coimbatore, Tamil Nadu, India.

***Abstract –*** Breast cancer is genetically and clinically heterogeneous disease. An area of interest for data mining applications is the study of biomedical data which is the combination of image processing with the data mining techniques or algorithm to analyze the hidden data by creating the patterns. This paper deals with the classification of breast cancer within digital mammography images. Digital Image processing facilitates in medical area widely in case of detecting and classifying the mammogram images. Classification of breast cancer in mammogram image is an important step to classify whether the patient is affected by cancerous (malignant) or non-cancerous (benign) tumors. Researchers have been used various algorithms and investigation methods to investigate mammogram images depending on the ultimatum of the disease, status of the disease and the quality of image. The main aim of classification of mammogram images is to select the best treatment. Doctors suggest best treatment for the patient if the result is positive and can avoid from death. Researchers uses so many data mining algorithms used for classification of mammogram images are k-Nearest Neighbor, Nave Bayes, Artificial Neural Network and Support Vector Machine algorithm.

***Keywords***—**Benign, Classification, Data Mining Algorithms, Malignant , Mammogram Images**.

## 1. INTRODUCTION

Mammogram, breast x-ray imaging is an effective, low cost, reliable method in early breast cancer detection. Mammogram images are classified as normal, benign and malignant [3]. Breast cancer is a malignant (cancer) tumor starting with breast cells. A screening mammogram aims to find breast cancer when it's too small to be felt by a woman or her doctor. Cost effectiveness is one of the major requirements for a mass screening program to be successful. The ultimate diagnosis of all types of breast disease depends on a biopsy. In most cases the decision for a biopsy is based on mammography findings. Biopsy results indicate that 65-90% of suspected cancer detected by mammography turned out to be benign [8]. Hence, it would be valuable to expand a computer aided technique for mass classification based on extracted features from the region of interests (ROI) in mammograms. This can reduce the number of unnecessary biopsies in patients with benign disease and thus avoid patients physical and mental suffering, with an added bonus of reducing healthcare costs. Finding breast cancers before they grow and spread greatly improves a woman's chance for successful treatment [3].

A diagnostic mammogram is a breast X-ray examination when a patient shows signs and symptoms of breast disease, or who with earlier mammography findings needs an imaging follow-up [5]. Feature selection and feature extraction [6] are dimension reduction techniques. Feature selection is generally used in breast cancer classification. Feature selection filters redundant and irrelevant features from original data. Feature selection, a data mining pre-pre-processing step selects and extracts valuable information in massive related materials. "It is investigated that

logistic regression model differentiate between benign and malignant in decision making for early breast cancer detection and identifies most important features associated with breast cancer [7]". "Feature selection's advantage including improvement of prediction reduces training times and ensures faster classifier performance [8]". Benign tumors are non-cancerous whereas malignant tumors are cancerous in which, the cells grow abnormally [4].

## II LITERATURE REVIEW

In breast cancer classification various classification techniques are employed on the segmented image in order to classify the lesions, malignant or Benign. Some of the existing classification method for mammogram images is described in this section.

The literature contains various approaches for supervised feature learning, risk factors, classification to classify the lesions in mammogram images. Hua Li, Shasha Zhuang, Deng-ao Li , Jumin Zhao,and Yanyun Ma, proposed an enhanced DenseNet neural network model, the DenseNet-II neural network model. This model has produced effective and accurate classification of benign and malignant mammography images [11]. The datasets in this paper was collected from mammography images from Shanxi Medical University. This study has produced good accuracy in classification.

M. Kanchana and P. Varalaxmi discussed on the methods for mass detection and classification. This proposed work has also discussed on their merits and demerits of the methods. It also provides various classifiers and combinations for mass classification [9]. The datasets were collected from DDSM (Digital Database for Screening Mammography) and MIAS (Mammographic Image Analysis Society) dataset.

M. Karabatak and M. Cavedet Ince has been provided automatic diagnostic systems. In this proposed work Association Rule (AR) and Neural Network (NN) was compared with Neural Network model to measure the system performance [10]. In this work Wisconsin breast cancer database were used to find the performance of the automated diagnostic system.

Saima Anwar Lashari et al., has been suggested to use various data mining techniques. In this proposed work various data mining techniques were used such as KNN, SVM, Fuzzy KNN, and Decision Trees in order to classify the diseases like Liver disease, Diabetes and Cancer [13]. This study produces the best result for 2D mammogram images.

Kharya and Shweta has been discussed on various data mining approaches to distinguish Benign from Malignant [14]. This paper provides an overview of present research on breast cancer using data mining techniques. It also enhances the breast cancer diagnosis and prognosis. This paper also discusses on Association Rule, Neural Network and Decision Trees to classify the tumor.

Sertan Kaymaka, Abdulkader Helwana and Dilber Uzuna has been proposed Back Propagation Neural Network (BPPN) for the classification of image. This proposed work has further improved for the automatic classification of breast cancer images by using Radial Basis Neural Network (RBFN)." Images were represented using Local Binary Pattern and Curvelet transformed images"[15].

Mohammed M Abdelsamea, Marghny H Mohamed and Mohamed Bamatraf has been provided a novel neural network approach for the classification of abnormal mammographic images into benign or malignant based on texture representation [16]. This proposed work showed 95.2% accuracy to classify the tumor.

Akhilesh Kumar Shrivas and Ankur Singh has used various data mining based classification techniques [17]. This proposed work uses Bayes Net and Support Vector Machine(SVM) for the classification of mammogram images. The performances were measured in terms of accuracy, sensitivity and specificity.

Aswini Kumar Mohanty et al., has used Association Rule with MARI (Mining Association Rule in Image Database) for classification of tumors. The normal and abnormal images are correctly classified as True Positive and True Negative respectively [18].

Monika Hedawoo, Abhinandan Jaisawal and Nishita Mehta has been described a breast cancer classification using two-computational intelligence system [19]. The two- computational intelligence paradigms were K- Nearest Neighbor and Naïve Bayes Algorithm. These were used for classification of data whether it is an cancerous or non-cancerous data.

Dr. M Raja Sekar has provided Convolutional Generative Adversarial Networks (CGANs). These CGANs were used to divide the regions of interest which consists of cancer cells [20]. The database used in this proposed work was DDSM. In this work the comparison of SVMs were carried out. The accuracy produced in this work for mammogram images with respect to precision was 86%.

Xiaofei Zhang et al., suggested Convolutional Neural Network (CNN) for image classification. It also provides the different models of CNNs which were built to classify both the 2-D mammograms and 3-D Tomosynthesis [21].

S. Devisuganya and R.C. Suganthe has been proposed a system mainly for automatic segmentation and classification of

mammogram images as benign, malignant or normal [3]. It also provides a hybrid method of data mining technique.

B. Surendiran and A. Vadivel has been proposed a feature selection method using Analysis Of Variance (ANOVA) [6]. DiscriminantAnalysis (DA) which facilitates classifying mammogram masses. In this proposed work the results were compared with dimension reduction techniques like Principal Component Analysis(PCA). The accuracy obtained is 82% in classification of images.

## III PROPOSED FRAMEWORK OF CLASSIFICATION METHODS

Mammogram images are taken from MIAS database.

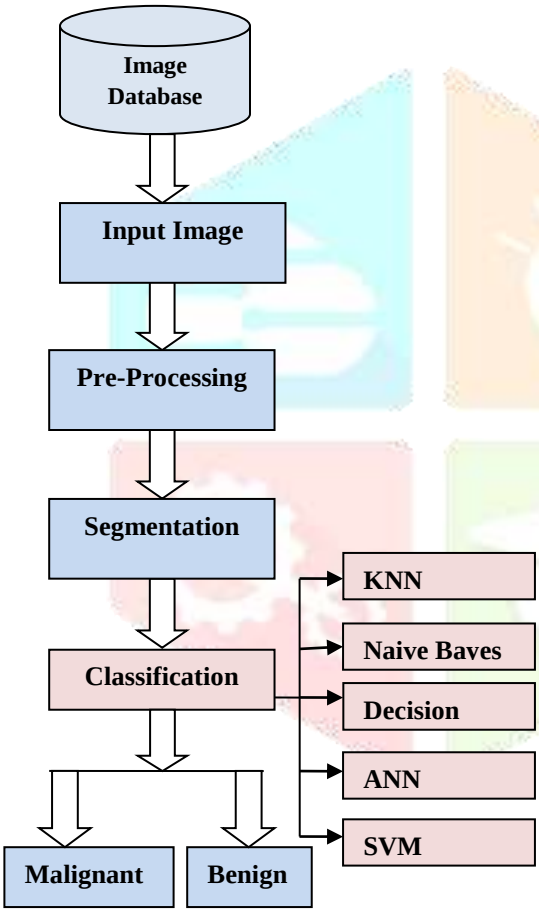


**Fig. 1. Proposed Framework of Classification Methods for Mammogram Images**

The pre-processing steps are carried out and the noises are removed using filtering techniques. The segmentation steps are done to find the tumour affected area. There are so many methods to classify the lesions in breast.

In this proposed work five different classifier methods were studied in order to know the performance of each classifier. One of the supervised method is classification algorithm. In this algorithm first trained on a set of sample images which is training set. The extracted features are input to the classifier. By applying various classifiers the dataset are analyzed. The accuracy measures such as Precision and Recall are used to evaluate the performance of the classifiers [1].

Fig.1. shows various methods of Image Classification using data mining algorithms and each method are classified into so many sub methods. These methods facilitates for the Breast Cancer detection in mammogram images. Image classification is carried out with the extraction of identified area in mammogram images for classifying cancerous and non-cancerous tissues in breast.

### A. K-Nearest Neighbor

Layer Based segmentation is used for object detection and image segmentation which is the combined form of the output of a bank of object detectors. In order to define shape masks and explain the appearance, and depth ordering composite is needed. [21]. It is easy to implement. Training is done in faster manner.

### B. Naive Bayes Method

Bayesian classifiers [12] are statistical classifiers. This method can envisage class membership probabilities. Naive Bayes Classifier is a probabilistic model based on Baye's theorem. It shows the high accuracy and speed when tested to large databases. It is defined as a statistical classifier. It is one of the frequently used method for supervised learning. It gives an proficient approach of handling any number of attributes or classes which is entirely based on probabilistic theory. Bayesian classification provides practical learning algorithms and prior knowledge on observed data [2]. It makes computations process easier. It provides better speed and accuracy for huge datasets.

### C. Decision Trees

A decision tree is a tree where each non-terminal node serves as a test or decision on the treated data item. Alternative of a confident branch depends upon the outcome of the test. To classify a certain data item, start at the root node and follow the assertions down until it reaches a terminal node (or leaf). A resolution is made when a terminal node is approached. Decision trees can also be depicted as a special form of a rule set, characterized by their hierarchical organization of rules [9]. It does not require domain prior knowledge in the construction of decision tree. It minimizes the ambiguity of complicated decisions, Decision tree It is easy to interpret and it also handles both numerical and categorical data.

### D. Artificial Neural Network

Artificial Neural Networks (ANNs) are arithmetical learning algorithms that are motivated by properties of the biological neural networks. ANNs are loosely based on biological neural networks in a sense that they are implemented as a system of interconnected processing elements, sometimes called nodes, which are functionally analogous to biological neurons. Fig.2 shows the back propagation neural network. Blue circle on left side represents the Input layer, blue circle on right side represents output layer and yellow circles in middle represents Hidden layer. The connections between different nodes have numerical values, called weights, and by altering these values in a systematic way, the network is eventually able to approximate the desired function. Each node in the network takes many inputs from other nodes and calculates a single output based on the inputs and the connection weights. The output in general fed into another neuron by repeating the process. The input layer receives the inputs and the output layer produces an output. The layers that lie in between these two layers are called hidden layers. It easily identifies complex relationships between dependent and independent variables. It also able to handle noisy data.

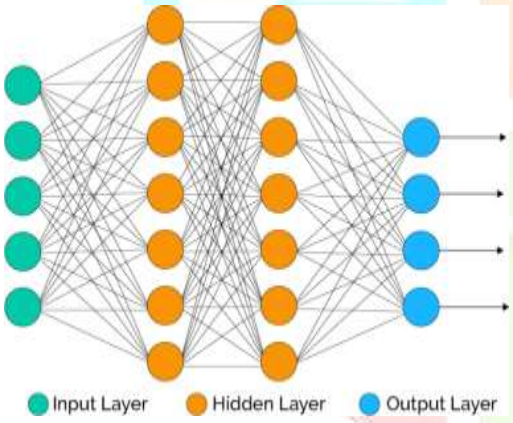


**Fig. 2. Back Propagation Neural Network**

### E. Support Vector Machine (SVM)

Support vector machines (SVMs) [11] are supervised learning methods that generate input-output mapping functions from a set of labeled training data. The mapping function can be either a classification function which are used to categorize the input data or a regression function which are used for the estimation of the desired output. Support Vector machines (SVM's) are a relatively new learning method used for binary classification. SVMs have established highly viable performance in many real-world applications.[22] It produces better accuracy among the other classifiers. Easily handle complex nonlinear data points. Over fitting problem is not as much as other methods.

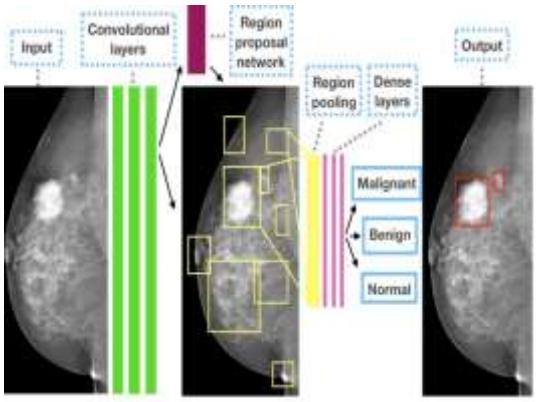


**Fig. 3. Classification of Breast Lesion**

In Fig. 3 the overall flow of breast lesion detection, segmentation and finally it shows the classification of tumors such as Malignant, Benign and Normal.

## IV CONCLUSION

This paper concentrates on various Image Classification techniques for mammography using data mining algorithms. Data mining algorithms with Image processing plays a vital role in medical field. This paper provides a study of various classification methods on breast cancer classification in mammogram images. The classification of tumor is very important step after exact feature extraction and feature selection process in digital mammogram images. From this study the prognostic problem is mainly analyzed under ANNs and SVM among the classification methods such as Decision Trees and KNN. This paper also facilitates the clinicians in decision making to find out the cancerous, non-cancerous tissues or normal in a patient.

## REFERENCES

[1] A.Subasini, Nirase Fathima abubacker and Dr.Rekha "Analysis of classifier to improve Medical diagnosis for Breast Cancer Detection using Data Mining Techniques" International Journal of Advanced Networking and Applications, Vol. 5, No. 6, pp. 2117-2122 , 2014.

[2] A.S.Pitchumani Angayarkanni and Dr.Nadira Banu Kamal, "Automatic Classification Of Mammogram Mri Using Dendograms" Asian Journal Of Computer Science And Information Technology , 2: 4, pp. 78- 81, (2012).

[3] S. Devisuganya and R.C. Suganthe, "Breast Cancer Detection: A Framework to Classify Mammograms", Middle-East Journal of Scientific Research 24 (Techniques and Algorithms in Emerging Technologies)", pp. 573-581, ISSN 1990-9233, 2016.

[4] CJ. Vyborny and ML .Giger, "Computer vision and artificial intelligence in mammography," 162:699,708,1993.

[5] C.H.Lee,D.D. Dershaw, D. Kopans, P. Evans, B.Monsees, D. Monticciolo and L.W. Burhenne, 2010"Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound and other technologies for the detection of clinically occult breast cancer" Journal of the American College of Radiology, Vol. 7, No. 1 pp.18-27, Jan 2010.

[6] B. Surendiran and A. Vadivel," Feature selection,using mass classification" International Journal of Recent Trends in Engineering and Technology, Vol.3, No.2, pp. 55-57, 2010.

[7] S.T.Luo and B.W. Cheng," Diagnosing breast masses in digital mammography using feature selection and ensemble methods", Journal of Medical Systems, Vol. 36, No.2, pp. 569-577, 2012.

[8] R.Nithya and B. Santhi, "Mammogram classification using maximum difference feature selection method" Journal of Theoretical and Applied Information Technology, Vol.33, No.2, 2011.

[9] H.D. Cheng, X.J. Shi, R. Min, L.M. Hu, X.P. Cai and H.N. Du "Approaches for automated detection and classification of masses in mammograms", Pattern Recognition Vol. 39, pp. 646-668, 2006.

[10] M. Karabatak and M. Cavedet Ince, "An Expert System for detection of Breast cancer based on association rules and Neural network Expert System with applications", Vol. 36, pp. 3465-3469, 2009.

[11] Hua Li, Shasha Zhuang, Deng-ao Li , Jumin Zhao,and Yanyun Ma "Benign and malignant classification of mammogram images based on deep learning", Biomedical Signal Processing and Control Vol **.** l 51, pp. 347–354, 2019.

[12] P .Valarmathie, V.Sivakrithika and K . Dinakaran, "Classification of mammogram masses using selected texture, shape and margin features with multilayer perceptron classifier.", Biomedical Research, pp. S310-S313, 2016.

[13] Saima Anwar Lashari, Rosziati Ibrahim, Norhalina Senan and N. S. A. M. Taujuddin " Application of Data Mining Techniques for Medical Data Classification: A Review" MATEC Web of Conferences 150, 06003 (2018).

[14] Kharya and Shweta "Using data mining techniques for diagnosis and prognosis of cancer disease." , 2012.

[15] Sertan Kaymaka, Abdulkader Helwana and Dilber Uzuna" Breast cancer image classification using artificial neural networks "Elsevier, ScienceDirect, Procedia Computer Science 120 , pp. 126–131, 2017.

[16] Mohammed M Abdelsamea , Marghny H Mohamed and Mohamed Bamatraf, "Automated Classification of Malignant and Benign Breast Cancer Lesions Using Neural Networks on Digitized Mammograms", Cancer Informatics, Vol.18, pp. 1–3, 2019.

[17] Akhilesh Kumar Shrivas and Ankur Singh,"Classification of Breast Cancer Diseases using Data Mining Techniques", International Journal of Engineering Science Invention, Vol. 5, No. 12, December 2016.

[18] Aswini Kumar Mohanty, Mrs. Arati Pradhan and Mrs. Saswati Sahoo, ," Mass Classification Method In Mammograms Using Association Rule Mining To Diagnose Breast Cancer", International Journal of Computer Science and Mobile Computing (IJCSMC) **,** Vol. 5, No. 7, pp.253 – 263, July 2016.

[19] Monika Hedawoo, Abhinandan Jaisawal and Nishita Mehta"Comparison Of Data Mining Algorithms For Mammogram Classification", International Journal Of Electrical, Electronics And Data Communication, Vol. 4, No.7, July 2016.

[20] Dr. M Raja Sekar " Breast Cancer Detection using Fuzzy SVMs", International Journal for Research in Applied Science & Engineering Technology (IJRASET**),** Vol. 5 Issue VIII, August 2017.

[21] Xiaofei Zhang , Yi Zhang, Erik Y. Han, Nathan Jacobs, Qiong Han , Xiaoqin Wang, and Jinze Liu, "Classification of Whole Mammogram and Tomosynthesis Images Using Deep Convolutional Neural Networks," IEEE Transactions on Nano Bioscience , Vol. 17, No. 3, July 2018

[22] Dezső Ribli, Anna Horváth, Zsuzsa Unger and Péter Pollner "Detecting and Classifying Lesions in Mammograms with Deep Learining" ,Springer Open Mechanical Engineering and Mechanics, 15 March 2018.