# Survey on Different Methods on Mammogram Image Classification

**Dr. Ashoka S B[1], Nirmala Y[2]**

[1]Assistant Professor, Department of Computer Science (MCA) PG-Coordinator, Maharani Cluster University Bengaluru, Karnataka

[2]Assistant Professor, Department of Computer Science (MCA) Maharani Cluster University Bengaluru, Karnataka

**Abstract:** ***The image processing is utilized to convert the image nature to improve and enhance the image for human interpretation. Image processing also used to render images for machine perception. Image segmentation is a technique which has opened a wide field of methods to analyze data, which could be in the form of images, videos. This becomes useful in breast cancer detections as images are used to extract features from the region of interest in the provided mammogram. Mammography tests have been identified as a potential source for pre-detection of a cancerous cell and thus, detection done with the technique of image segmentation has been proven effective in early diagnosis of breast cancer. This paper describes the different types of image preprocessing on mammogram images, followed by image classification using different pattern classifiers can be utilized to achieve an efficient diagnosis of breast cancer. The analysis shows that there is still sufficient scope of improving efficiency and accuracy of the classification techniques available.***

**Keywords:** Mammogram Images, Preprocessing, Feature Extraction, Classification

## 1. Introduction

Breast cancer is one among the leading cancers within the female population. it's the leading explanation for death thanks to cancer in women. The prevention of carcinoma as not yet been found; early detection is vital. Mammography [1] may be a low x-ray procedure for the visualization of breast. Mammography is that the key screening tool for the first detection of carcinoma. In initial step the image is preprocessed by image enhancement techniques and segmentation methods. Removal of noise from mammogram images is extremely important and for improving the standard of images several techniques are available like Contrast Stretching, Histogram Equalization [2] is one of the method used in the image's histogram for contrast adjustment. Histogram equalization accomplishes by efficiently spreading out the foremost frequent intensity values. the tactic is beneficial in images with backgrounds and foregrounds that are both bright or both dark. Mammogram images, Histogram equalization is employed to form contrast adjustment in order that the image abnormalities are going to be better visible. Adaptive median filter [3] are proposed to realize a balance between straightforward averaging (in homogeneous regions) and all-pass filtering (where edges exist). They adapt to remove required speckles and selectively properties of the image from different parts of the image. The mean filter [3] replaces each pixel by the typical value of the intensities in its neighborhood. It can locally reduce the variance and is straightforward to implement. Thresholding [4] usually involves selecting one gray level value from an analysis of the grey-level histogram, to segment the histogram into background and breast tissues. All the pixels with grey level value but the edge are marked as background and therefore the rest as breast.

Once the segmentation is finished, next step is to extract the specified features of the image by using different feature extraction techniques. The discrete wavelet transform (DWT) [5] translates the image into an approximation sub-band consisting of the size coefficients and a group of detail sub-bands at different orientations and determination scales composed of the wavelet coefficients. DWT provides an appropriate basis for separating the noise from a picture. The features are extracted using different techniques such as texture feature, position feature and shape feature e.t.c. Texture processing [5] algorithms are usually divided into three major categories: structural, spectral and statistical. Structural methods consider textures as repetitions of basic primitive patterns with a particular placement rule [5]. Statistical methods are supported statistical parameters like the grey Level Co-occurrence Matrices [5], the grey Level Difference Method, Principal Components Analysis and grey Level Run Length Matrices. Spectral methods are supported the utilization of a special space by a transform like the fourier transform, wavelet transform or curvelet transform [6]. These transforms analyzing the facility spectrum, images are oriented at various directions in multiple scales, with flexible aspect ratios and that they are helpful to separate a special data [6].

Last step is classifying the mammogram images as normal and malignant images using classifiers. Support vector machines (SVM) are supported the Structural Risk Minimization principle [7] from statistical learning theory. the thought of structural risk minimization is to seek out a hypothesis h with rock bottom true error. The Convolutional Neural Network CNN [8] moreover referred to as ConvNet may be a class of deep neural networks, widely utilized in image processing, 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. Random Forest [9] is supervised classification algorithm and a substantial bagging Technique. Random forest trees Performance is modest to coach and tune and far associated with boosting. The key concept of bagging is reducing the variance by middling noisy unbiased models. the target of this paper is to research different approaches that are developed to deal with the simplest classification methods of mammogram images so as to extend the accuracy.

## 2. Overview of Different Images Enhancement and Segmentation

**Techniques**

Image enhancement techniques are necessary, in order to enhance the quality of the image and to remove the noise of the image. The medical images like mammograms that are very difficult to be interpreted, those images need preparation phase in order to improve the quality of image and make the segmentation results more accurate. There are some different techniques available for image enhancement such as, Histogram of the mammographic image is considered as a probability distribution. According to the information theory, the uniform distribution of the histogram contains most information. Adaptive histogram equalization enhances contrast in radiological images and provides a large global dynamic range, but small local feature gray-level varies. Contrast Limited Adaptive Histogram Equalization (CLAHE) is a special case of adaptive histogram equalization where the histogram is calculated for the contextual region of a pixel. In CLAHE the maximum contrast adjustment is limited by imposing a user-specified maximum. This helps the resulting image not to become too noisy. Enhancement using Dyadic Wavelet Processing is adaptable to the various nature of diagnostic features within the image under analysis and it permits an equivalent algorithm for both tumor and mass detection. Adaptive Median filter (AMF) [3] is one of the most advanced approaches in filtering which is used to assess the effects of impulse noise on an image. Here, the image pixels are compared with the surrounding pixels for classification. The issues of conventional median filters can be addressed by using adaptive median filters where the window size can be varied adaptively. Genetic programming (GP) based quantum noise removal filter is meant to make a numerical optimal evolved expression for mammogram image restoration that optimally combines and exploits dependencies among the features of the degraded/blurred mammogram image. The performance of the filter function is estimated using various degraded mammogram images. The proposed filter effectively removes the noise and enhances the mammograms for further processing. The segmentation method aims to represent the image in straightforward method such it's simple to investigate and perceive. The resultant image offers a lot of important info. Segmentation covers the complete image by dividing it into totally different sets. Image segmentation plays an important role in cancer tumor detection. Numerous sorts of technologies square measure getting used in segmentation of the Digital X-ray photograph pictures. Segmentation method is a picture process technique that separates the Region of Interest from the background. Segmentation is performed mistreatment numerous approaches like native thresholding, K–Means clump or Otsu Segmentation Technique. Region Growing Method (RGM) [4] is one among the foremost ordinarily used image segmentation technique encompassing the subsequent approaches particularly detection of the seed purpose and region growing utilized by detected seed purpose. Just in case of mammograms wherever the background is black, the breast tissues square measure seen as bright. Here, the medians of the blocks of mammograms square measure obtained and therefore the true grey values square measure determined. For a given input image, the seed price calculated and therefore the threshold price is ready. Absolute distinction between the input image and seed price is calculated. Extraction of foreground is finished mistreatment iteration taking every component values. The process of finding region of interest or suspected space is completed by mistreatment morphology algorithms. Morphology rule is used as a result of it's able to separate abnormal space from traditional space. The process of finding region of interest or suspected areas can be completed by using Morphology. The Morphology algorithm is employed reason beyond is ready to separate abnormal area from normal area. Otsu's N is nonparametric and unsupervised method of automatic threshold selection for segmentation of images. It is optimum in the sense that it maximizes the between class variance used to measure utilized in statistical discriminant analysis.

## 3. Different Feature Extraction

### 3.1 Techniques

Once segmentation is done next important stage is to extract the feature values from the image. Multiple techniques are available for extracting the dataset, Gray Level Co-Occurrence Matrix (GLCM) is suitable to obtain the statistical texture features of mammograms. GLDM uses probability Density Functions (PDFs) to aid in the statistical analysis of the given image. Discrete Wavelet Transform (DWT) separates data elements into different frequency components. It also allows the given function to be analyzed at various levels of resolution. DWT transforms data vector into a numerically different vector of same length by a linear transformation. Statistical techniques implement the spatial distribution of gray level values, by using local features at each point in the image and deriving a set of statistics from the distributions of the local feature. Relying on the number of pixels meaning the local feature, statistical techniques can be further being divided into first order, second-order and higher-order statistics. Common features such as, Mean is a measure of the average intensity of the neighboring pixels of an image. Variance is the average of squared deviation of all pixels from mean. Correlation measures the probability occurrence of the specified pixel pairs. 2D Gabor wavelets have been widely used in computer vision applications and modelling biological vision. Gabor elementary functions are suitable for modelling simple cells in visual cortex.

### 3.2 Overview of Different Pattern Classifiers

The feature values are extracted from the particular region from the mammogram images, next level is classifying the mammogram images as benign or malignant. There are plenty of classifiers available for classification such as, Support Vector Machine (SVM) is a constructive learning method in the field of statistical learning theory. It reduces the bound on learning algorithms. Due to this reason, SVM performs well when applied to the test data. Relevance Vector Machines (RVM) is based on Bayesian estimation for regression and classification problems. generalization error which occurs due to the data unseen by the learning machine during training phase instead of mean square error over the dataset like other machine RVM method can yield a better solution only when function depends on only a very small number of training samples called relevance vectors. The textural, statistical and spectral features are extracted by Bayesian classifier from each mammogram image and gives the models of real MCs which are used as training samples through a simplified learning phase of the Bayesian classifier. The classification outcomes are in binary form like 0 or 1 and used to create a binary image. Artificial Neural Network(ANN) comprises of various mathematical models used to replicate the pattern of functioning of biological neurons and the human nervous systems. ANN based approaches suffer from multiple local minima and major disadvantage of ANN is that they take a long time to converge. K- Nearest Neighbour Classifier (K- NN) uses a lazy learning classification strategy;instead of making any model with the test data it simply stores it. When a test input is given, the classifier uses the training data directly to classify it. K- NN classifies an unknown category vector into K similar vectors present in the training set. Since K- NN is based on instances, the feature sets to be rescaled before classification. Convolutional Neural Network (CNN) is built with many neurons which are having biases and learnable weights. Each neuron receives some inputs. Then performs a dot product and optionally follows it with a non-linearity. The entire network expresses a single differentiable score function which are taken from the pixels of raw image on one edge to class scores at the other.

## 4. Result and Discussion

The paper describes the different types of image preprocessing on mammogram images, followed by image classification using different pattern classifiers which can be utilized to achieve an efficient diagnosis of breast cancer. The analysis shows that there is still sufficient scope of improving efficiency and accuracy of the classification techniques available. The results of the considerable number of steps are clearly explained in the methodology. The main aim is to identify the best and average algorithm which helps to classify the mammogram images according to density. The various methodologies are listed in table 1.

**Table 1:** Different methods for Preprocessing, segmentation, feature extraction and classification of mammogram images

| Pre Processing | Segmentation | Feature Extraction | Classification | Accuracy | Number of images |
|---|---|---|---|---|---|
| ---- | Region Growing Method (RGM) | Cross Section Local Binary Pattern (CSLBP) and Difference of Gaussian (DoG) | Support Vector Machine (SVM) | 89.11% | 75 images from MIAS |
| Adaptive median filtering | Region Growing Method (RGM) based on seed value | Gray Level Dependency Matrix and Cross Section Local Binary Pattern (CSLBP) | Support Vector Machine (SVM) | 95.11% | 166 images from MIAS dataset and the 83 were real time images |
| ---- | Region Growing method | Gray level co- occurrence matrix, Angular secondary moment, Entropy Correlation and Variance | Radial Basis Function Neural Network (RBFNN) | 94.44% | 330 images from MIAS |
| Contrast stretching, Stochastic Modelling | Region growing | Trace Extraction, Singular Value Decomposition | SVM | 94.9% | 322 images from MIAS |
| --- | x and y coordinates with radius r by the experts(MIAS dataset) | Discrete Wavelet Transform (DWT) | RBFNN SVM K-NN | RBFNN 94.6% SVM 90.54% and K-NN 87.8% | 148 mammogram images |
| Morphology algorithm | ---- | Grey level co-occurrence matrix (GLCM) | SVM | 85% | MIAS dataset |
| Contrast Limited Adaptive Histogram Equalization (CLAHE) | ---- | Discrete Wavelet and Curvelet Transform | Convolution Neural Network (CNN) SVM | CNN 81.33% and SVM 83.74% | 2,576 images from DDSM, 150 images from MIAS |
| Median filter and Gaussian filter | Thresholding technique | Grey level co-occurrence matrix (GLCM) and Grey level Distribution matrix (GLDM) | Convolution Neural Network (CNN) | 73% | 40 images from DDSM dataset |
| Wavelet de-nosing filter | hard and soft threshold functions | Statistical feature extraction methods | FussCyier | 75.64% | 112 images from MIAS |
| Gamma Correction | Region growing | Grey level co-occurrence matrix (GLCM) and Intensity histogram method | SVM | GLCM+ SVM 87% and histogram +SVM 85.26% | 190 images from DDSM |
| -- | Wavelet and Gabor filter | Correlation feature selection method | Decision tree | 80% | 322 images from MIAS |
| Median filter and (CLAHE) and Discrete Fourier Transform (DFT) | Region Growing method | GLCM | SVM | 86.84% | 322 images from MIAS |

## 5. Conclusion

To improve the Mammogram image quality and contrast, the original image is converted into equalized image and histogram of the equalized image by using different techniques. According to the survey for pre-processing the image Adaptive median filter and Region Growing are very effective methods. From the extracted region of interest (ROI), pixel intensity values are extracted and feature values are generated to extract texture and statistical features by using various techniques. DWT and GLCM gives the best result and helps to classify the mammogram images. After extraction of feature values next step is to classify the mammogram images, for that many algorithms are proposed which are mentioned in the table 1. Under that algorithm SVM achieves better accuracy.

## References

[1] L.d.O. Martins *et al.*, Detection of Masses in Digital Mammograms using Kmeans and Support Vector Machine, Electronic Letters on Computer Vision and Image Analysis, pp. 39-50, 2009.
[2] Thangavel. K *et al.*, Mammogram Mining with Genetic Optimization of Ant-Miner Parameters, R, International Journal of Recent Trends in Engineering, Vol 2, No. 3, November 2009.
[3] Juan Shan *et al.*, Automated breast cancer detection and classification using ultrasound images: A survey, Pattern Recognition 43 (2010) 299 – 317.
[4] Sumit Chopra *et al.*, Segmentation of the Mammogram Images to find Breast Boundaries, IJCST Vol. 2, Issue 2, June 2011.
[5] Monika Sharma *et al.*, Feature Extraction of Mammograms, International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-2 Number-3 Issue-5 September-2012.
[6] Mohamed Meselhy Eltoukhy *et al.*, Breast cancer diagnosis in digital mammogram using multiscale curvelet transform Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia.
[7] Snehal A. Mane *et al.*, Mammogram Image Features Extraction and Classification For Breast Cancer Detection. International Research Journal of Engineering and Technology (IRJET): 02 Issue: 07 | Oct-2015.
[8] Ketan Sharma *et al.*, Classification of mammogram images by using CNN classifier. International Conference on Advances in Computing, Communications and Informatics (ICACCI), September 2016.
[9] S.Vasundhara *et al.*, Machine Learning Approach for Breast Cancer Prediction. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1, May-2019.