SIFT based Feature Extraction in Breast Image using Soft Computing

Dhanapriya .R[1] Saranyalakshmi .T[2] Kokila .D[3]
[1] P. G. Student [2, 3] Assistant Professor
[1, 2] Muthayammal Engineering College, Rasipuram
[3] CMS College Of Engineering And Technology, Coimbatore.

Abstract—The American Cancer Society (ACS) recommends women aged 40 and above to have a mammogram every year and calls it a gold standard for breast cancer detection. In this work a method for categorization of breast tissue thickness from mammographic images is proposed. The purpose of the method is to determine which class the breast tissue belongs to namely, fatty, fatty-glandular and dense-glandular. The SIFT algorithm is introduced for this purpose and it is used as the local feature extraction method, and k-NN algorithm is used for supervised classification. The SIFT features of each class can effectively model the breast tissue and the classification accuracy over 90% is achieved by classifier. Local adaptive threshold technique is used for the segmentation of cancer tissue. This examination aims at providing an outline about recent advances and developments in the field of breast cancer using mammograms, specifically focusing on the numerical aspects, aiming to act as a mathematical briefing for intermediates and experts in the field.

Key words: ACS (American Cancer Society), Sift (Scale Invariant Feature Transform), K-NN (K-Nearest Neighbor).

I. INTRODUCTION

CANCER refers to the uncontrolled multiplication of a group of cells in a particular location of the body. Breast cancer is any form of malignant tumor which develops from breast cells. Breast cancers are traditionally known to be one of the major causes of death among women. Mortality rates due to breast cancer have been reducing due to better diagnostic facilities and effective treatments. One of the leading methods for diagnosing breast cancer is screening mammography.

A mammogram is an x-ray of the breast. A diagnostic mammogram is used to diagnose breast disease in women who have breast symptoms or an abnormal result on a screening mammogram. Screening mammograms usually take 2 views (x-ray pictures taken from different angles) of each breast, while diagnostic mammograms may take more views of the breast. For some patients, such as women with breast implants, more pictures may be needed to include as much breast tissue as possible.

For a mammogram, the breast is pressed between 2 plates to flatten and spread the tissue. This may be uncomfortable for a moment, but it is necessary to produce a good, readable mammogram. The compression only lasts a few seconds. The entire procedure for a screening mammogram takes about 20 minutes. This procedure produces a black and white image of the breast tissue either on a large sheet of film or as a digital computer image that is read, or interpreted by a radiologist.

The need for early detection of breast cancer is highlighted by the fact that incidence rates for breast cancer is one of the highest among all cancers according to the American Cancer Society which quotes a morbidity of 230000 and a mortality of 40 000 according to the latest figures gathered for the American population. Important signs to look for in the case of breast cancer are clusters of micro calcifications, masses and architectural distortions. Following the results of screening mammography, a follow up study is made for patients according to the level of suspicion of the abnormality. This stage is referred to as diagnostic mammography. Both screening mammography and diagnostic mammography are performed by radiologists who visually inspect the mammograms. Early detection of breast cancer through screening and diagnostic mammography increases breast cancer treatment options and survival rates. Unfortunately, due to the human factor involved in the screening process, detection of suspicious abnormalities is prone to a high degree of error. Studies have shown that radiologists have an error rate between 10%−30% for detection of cancer in screening studies. Misinterpretation of breast cancer signs result in 52% of the errors and 43% of the errors are caused due to overlooking signs in abnormal scans. As a result of this error rate, biopsies are frequently performed on benign lesions, resulting in unwarranted expenditure and anxiety for the patient involved. The cost associated with errors due to misclassification of mammograms is considerable. This is because of the fact that false negatives are a huge problem in screening mammography as early detection can reduce treatment cost, time and effectiveness to a great extent. False negatives affect all three parameters as early detection is not an option with an incorrect diagnosis. A study by found that double reading of screening mammograms provided greater sensitivity than single reading without increasing recall rates. But, manpower is a major drawback with this approach. The number of radiologists required for double reading of mammograms will be huge. As a result, many nations might not be able to meet the manpower requirements for such an approach. A major reason for these errors is due to the fact that radiologists depend on visual inspection. During manual screening of a large number of mammograms, radiologists may get easily worn out, missing out vital clues while studying the scans. To offset these effects, tremendous effort is being made to automate the process of mammographic screening. Automated screening of mammograms or computer-aided diagnosis (CAD) of breast cancer is a vast field of research. Sampat *et al.*, and Rangayyan *et al.*, provide an extensive review on different stages of a CAD methodology for breast cancer. Classifier systems have been widely used in medical diagnosis. Though the most important factor in diagnosis is evaluation of data taken from patients by human experts, expert systems and various artificial intelligence techniques for

classification aid radiologists to a great extent. As yet, there is no definitive literature which focuses on an elaborate discussion on the feature extraction, feature selection and classification methodologies used in breast cancer detection. The current study aims at filling this gap by documenting developments in that aspect.

II. METHODOLOGY

- Preprocessing.
- Feature Extraction.
- Feature Selection.
- Classification.
- Segmentation.

III. RELATED WORK

A. Preprocessing

Preprocessing is the initial step for detecting the lung image. The preprocessing step is done in two steps. They are

- De-noising
- Median Filter

1) De-noising

Image de-noising algorithms may be the oldest in image processing. Many methods, regardless of implementation, share the same basic idea noise reduction through image blurring. Blurring can be done locally, as in the Gaussian smoothing model or in anisotropic filtering by calculating the variations of an image. White noise is one of the most common problems in image processing. Even a high resolution photo is bound to have some noise in it. For a high resolution photo a simple box blur may be sufficient, because even a tiny features like eyelashes or cloth texture will be represented by a large group of pixels. However, current direct class hardware allows us to implement high quality filters that run at acceptable frame rates. The main idea of any neighborhood filter is to calculate pixel weights depending on how similar their colors are. There are two such methods: the K Nearest Neighbors and Median filters.

The input image is a normal RGB image. The RGB image is converted into grey scale image because the RGB format is not supported in Mat lab. Then the gray scale image contains noise such as white noise, salt and pepper noise etc. This can be removed by using Median filter from the extracted breast image.

2) Median Filter

The median filter is a nonlinear digital filtering technique, often used to remove noise. Such noise reduction is a typical pre-processing step to improve edge detection on an image. Median filtering is very widely used in digital image processing under certain conditions. The sliding median filter of a pre-defined window size WxW centered at i= (i1, i2) moves uniformly over the noisy image, g, chooses the median, μ of the pixels within the spatial positions Ω^w around in to have g (i) replaced by μ. For the set of pixels within a square window WD x WD, centered at i= (i1, i2) and defined spatially by equation, the median, μ of the pixels in Ω_i^w

$$u(i) = \mu(i) = median\{g(i)/j \in \Omega_i^w\}$$

Thus the output of the median filter is that value θ which produces the least sum of absolute errors with all the pixels in the local neighborhood defined by them ask. The output of the median filter at spatial position can also be given as

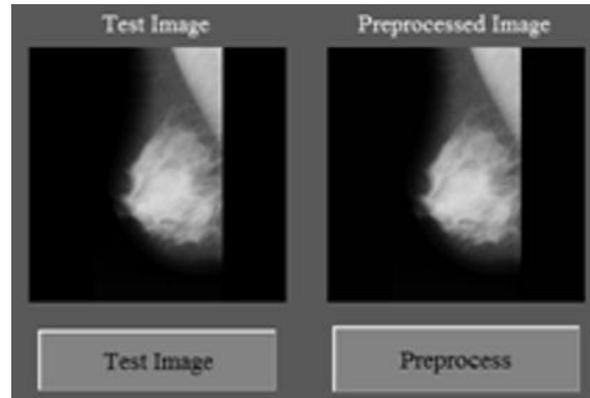$$u(i) = \mu(i) = \arg\min_{\theta} \sum_{r \in \Omega_i^w} |g(r) - \theta|$$

Fig. 1: preprocessed image

B. Feature Extraction

In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction.

When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (e.g. the same measurement in both feet and meters) then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input.
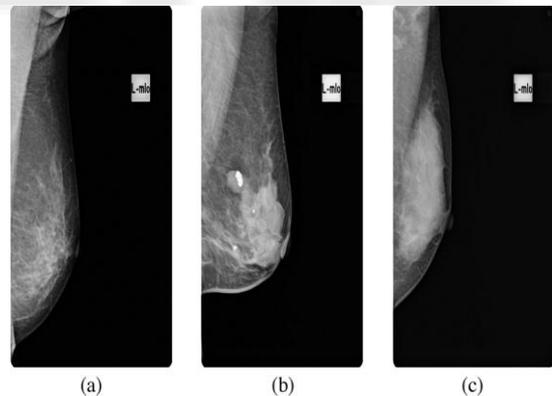

Fig. 2: (a) Normal (b) Benign (c) Malingnent

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy.

SHIFT algorithm is one of the emerging algorithms for feature extraction purpose. The detection and description of local image features can help in object recognition. The SIFT features are local and based on the appearance of the object at particular interest points, and are invariant to image scale and rotation. They are also robust to changes in illumination, noise, and minor changes in viewpoint. In addition to these properties, they are highly distinctive, relatively easy to extract and allow for correct object identification with low probability of mismatch. They are relatively easy to match against a (large) database of local features but however the high dimensionality can be an issue, and generally probabilistic algorithms such as k-d trees with best bin first search are used. Object description by set of SIFT features is also robust to partial occlusion; as few as 3 SIFT features from an object are enough to compute its location and pose. Recognition can be performed in close-to-real time, at least for small databases and on modern computer hardware.

C. Scale-space extrema detection

This is the stage where the interest points, which are called key points in the SIFT framework, are detected. For this, the image is convolved with Gaussian filters at different scales, and then the difference of successive Gaussian-blurred images is taken. Key points are then taken as maxima/minima of the Difference of Gaussians (DoG) that occur at multiple scales. Specifically, a DoG image $D(x, y, \sigma)$ is given by,

$$D(x, y, \sigma) = L(x, y, k_i\sigma) - L(x, y, k_j\sigma)$$

Where $L(x, y, k\sigma)$ is the convolution of the original image $I(x, y)$ with the Gaussian blurring $G(x, y, k\sigma)$ at scale $k\sigma$, i.e.

$$L(x, y, k\sigma) = G(x, y, k\sigma) * I(x, y)$$

Hence a DoG image between scales $k_i\sigma$ and $k_j\sigma$ is just the difference of the Gaussian-blurred images at scales $k_i\sigma$ and $k_j\sigma$. For scale space extrema detection in the SIFT algorithm, the image is first convolved with Gaussian-blurs at different scales. The convolved images are grouped by octave (an octave corresponds to doubling the value of σ), and the value of k_i is selected so that we obtain a fixed number of convolved images per octave. Then the Difference-of-Gaussian images are taken from adjacent Gaussian-blurred images per octave.

Once DoG images have been obtained, key points are identified as local minima/maxima of the DoG images across scales. This is done by comparing each pixel in the DoG images to its eight neighbors at the same scale and nine corresponding neighboring pixels in each of the neighboring scales. If the pixel value is the maximum or minimum among all compared pixels, it is selected as a candidate key point.

This key point detection step is a variation of one of the blob detection methods developed by Lindeberg by detecting scale-space extrema of the scale normalized Laplacian, that is detecting points that are local extrema with respect to both space and scale, in the discrete case by comparisons with the nearest 26 neighbour in a discretized scale-space volume. The difference of Gaussians operator can be seen as an approximation to the Laplacian, with the implicit normalization in the pyramid also constituting a discrete approximation of the scale-normalized Laplacian. Another real-time implementation of scale-space extrema of the Laplacian operator has been presented by Lindeberg and Bretzner based on a hybrid pyramid representation.

Interpolation of nearby data for accurate position

First, for each candidate key point, interpolation of nearby data is used to accurately determine its position. The initial approach was to just locate each key point at the location and scale of the candidate key point. The new approach calculates the interpolated location of the extremum, which substantially improves matching and stability. The interpolation is done using the quadratic Taylor expansion of the Difference-of-Gaussian scale-space function, $D(x, y, \sigma)$ with the candidate key point as the origin. This Taylor expansion is given by:

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}}\mathbf{x} + \frac{1}{2}\mathbf{x}^T\frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x}$$

Where D and its derivatives are evaluated at the candidate key point and $\mathbf{x} = (x, y, \sigma)$ is the offset from this point. The location of the extreme $\hat{\mathbf{x}}$ is determined by taking the derivative of this function with respect to $\mathbf{x}$ and setting it to zero.
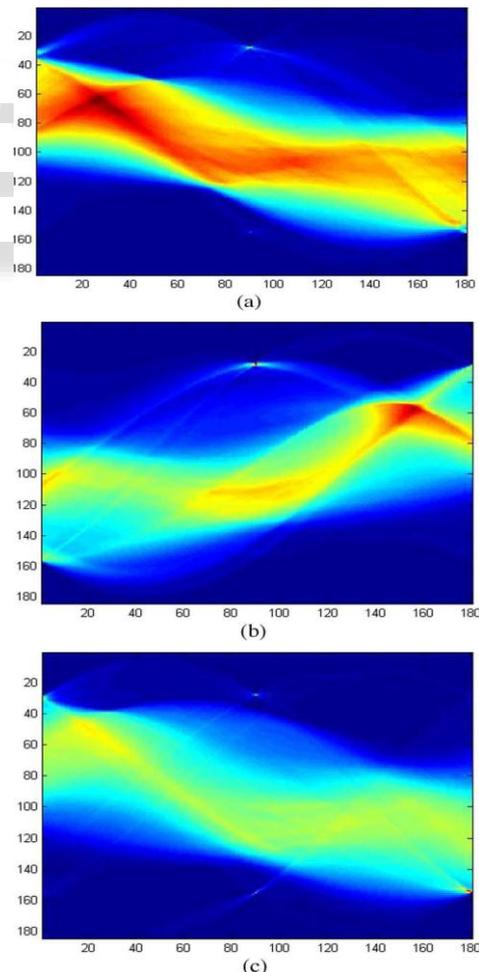

Fig. 3: Feature extracted from (a) Normal (b) Benign (c) Malignant

Fig 4: Feature value extraction by SIFT

If the offset $\hat{\mathbf{x}}$ is larger than 0.5 in any dimension, then that's an indication that the extreme lies closer to another candidate key point. In this case, the candidate key point is changed and the interpolation performed instead about that point. Otherwise the offset is added to its candidate key point to get the interpolated estimate for the location of the extreme. A similar sub-pixel determination of the locations of scale-space extrema is performed in the real-time implementation based on hybrid pyramids.

D. Feature selection

From the extracted features best features are selected using DEFS. Best features are selected from the extracted features. This process will reduce the time taken to detect the cancer in the Mammographic Images.

E. Classification

In pattern recognition, the *k*-nearest neighbor algorithm (*k*-NN) is a non-parametric method for classifying objects based on closest training examples in the feature space. In *k*-NN the function is only approximated locally and all computation is deferred until classification. The *k*-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its *k* nearest neighbors (*k* is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of that single nearest neighbor.

The same method can be used for regression, by simply assigning the property value for the object to be the average of the values of its *k* nearest neighbors. It can be useful to weight the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. (A common weighting scheme is to give each neighbor a weight of $1/d$, where *d* is the distance to the neighbor. This scheme is a generalization of linear interpolation.)
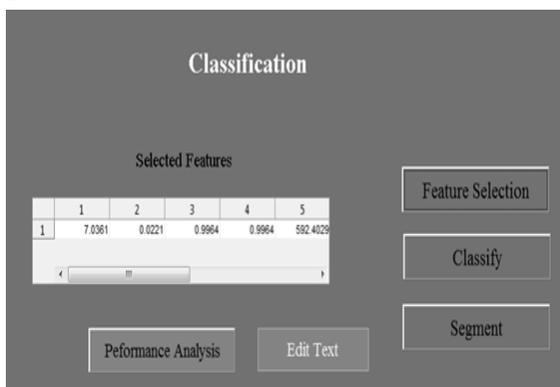

Fig 5: feature selection for classification.

The neighbors are taken from a set of objects for which the correct classification (or, in the case of regression, the value of the property) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. The *k*-nearest neighbor algorithm is sensitive to the local structure of the data.

Nearest neighbor rules in effect implicitly compute the decision boundary. It is also possible to compute the decision boundary explicitly, and to do so efficiently, so that the computational complexity is a function of the boundary complexity.

F. Image segmentation

It is important to segment the cancer tissue from the breast image for identifying the location. The segmentation is done for benign and malignant tumor not for the normal image. It is due to the normal image has the pixel value which is equal to the original image. For the abnormal image the segmentation is done according to the threshold value which has been set for the image.

Local and adaptive threshold segmentation will be carried out. In which the threshold value will be sets as 200. The image pixel value is compared with the original normal image then if the value is below threshold value set the pixel value as 1 otherwise pixel value as 0. For the pixel value 1 then it makes replacement as the original image pixel. By which the tumor part will be segmented.
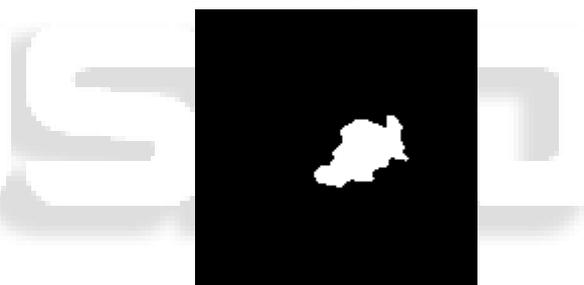

Fig 6: Segmented benign tumor cell

IV. System Architecture

A. Training Phase

In the training phase the network will be trained for some set images. In which the true label of the mammography images has to be extracted. Also the features from the images have been extracted for identifying the abnormal cells. After which the network will be trained for the feature those are extracted from the sample set of breast image. Then the required features will be selected for testing purpose it is done in feature selection block.
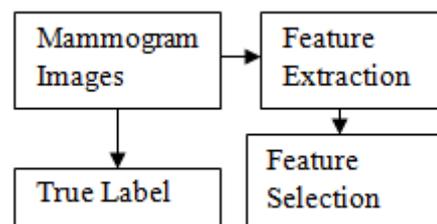

Fig 7: Training phase

B. Testing Phase

In this testing phase sequence of process will be handled to get the clear result. The image which is going to be process is taken from mammography. From that image it is must to preprocess it to eliminate the Gaussian noise and have to remove the pectoral muscle, label, articraft from the image. And also have to suppress the background noise to enhance the image quality. Therefore de-noise is done at preprocessing step.
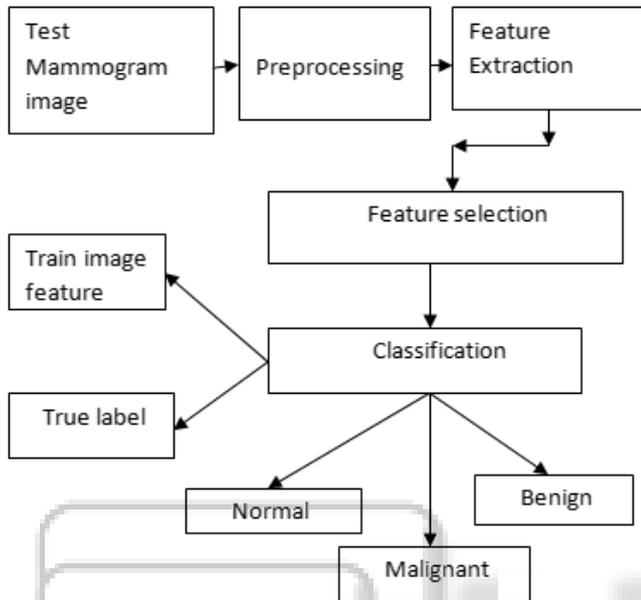

Fig. 8: Testing phase

Then the features of preprocessed image is extracted by the algorithm which is discussed above (SIFT). Then the feature selection is done by the DEFS algorithm in which the required feature is selected according to the difference evolution feature selection. After which the classification of the tumor cell is carried by classifier.

V. CONCLUSION

In this paper, we proposed a Computer Aided Diagnosis of Breast Cancer using the Mammographic Images. In this algorithm the input image is first preprocessed then features are extracted from the preprocessed image. From the extracted features by the SIFT algorithm, best features are selected using feature selection. The classifier identifies the normal and abnormal tissue using the feature values. The classifier accuracy is measured as 87%. The value shows us the efficiency of the algorithm while comparing with other algorithms such as wavelet algorithm, GLCM and so on. The wavelet has the efficiency of 80% and GLCM has 83%. The algorithm is much helpful in identifying the breast cancer earlier.

REFERENCES

[1] KarthikeyanGanesan, U.Rajendra Acharya, Chua Kuang Chua, Lim Choo Min, K. Thomas Abraham, and Kwan-Hoong Ng, "Computer-Aided Breast Cancer Detection Using Mammograms: A Review" IEEE reviews in biomedical engineering, vol. 6, 2013.

[2] Kother Mohideen1, ArumugaPerumal2, Krishnan 3 and Mohamed Sathik4, International Arab Journal of e-Technology, "Image Denoising and Enhancement Using Multiwavelet with Hard Threshold in Digital Mammographic Images", Vol. 2, No. 1, January 2011.

[3] Cheng-Lung Huang, Hung-Chang Liao, Mu-Chen Chenc, Expert Systems with Applications "Prediction model building and feature selection with support vector machines in breast cancer diagnosis",2008.

[4] M.P.Sampat, M.K.Markey, and A.C.Bovik, "Computer-Aided Detection and Diagnosis in Mammography, Handbook of Image and Video Processing". London, U.K.: Elsevier, 2003.

[5] F. Dehghan, H. Abrishami Moghaddam, M. Giti, "Automatic Detection of Clustered Micro calcifications in Digital Mammograms: Study on Applying AdaBoost with SVM-based Component Classifiers", International IEEE EMBS Conference Vancouver, British Columbia, Canada, August 20-24, 2008.

[6] Zheng and Andrew K. Chan*, Senior Member, Lei, IEEE, "An Artificial Intelligent Algorithm for Tumor Detection in Screening Mammogram", IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 20, NO. 7, JULY 2001.

[7] A.K.Jain, R.P.W.Duin, and J.C.Mao, "Statistical pattern recognition: A review," IEEE Trans. Pattern Anal. Machine Intell. vol. 22, no. 1, pp. 4–37, Jan. 2000, 10.1109/34.82481.

[8] Giuseppe Boccignone, Angelo Chianese, Antonio Picariello, "Computer aided detection of microcalcications in digital mammograms", Received 4 November 1999; accepted 13 April 2000.

[9] R. G. Blanks, M. G. Wallis, and S. M. Moss, "A comparison of cancer detection rates achieved by breast cancer screening programmes by number of readers, for one and two view mammography: Results from the UK National Health Service breast screening programme," J. Med. Screening, vol. 5, no. 4, pp. 195–201, 1998.

[10] Giuseppe Boccignone, Angelo Chianese, Antonio Picariello, "Mammographic Feature Enhancement by Multiscale Analysis", IEEE transactions on medical imaging, vol 13, no. 4, December 1994.