

# Mammographic Cancer Detection Using Computer-Aided Diagnosis System

Dhanshree Navnath Kokare<sup>1</sup>, Prof. P.P. Gumaste<sup>2</sup>

Department of E&TC Engineering, JSPMs, Jayawantrao Sawant College of Engg, Hadapsar, Pune, India<sup>1,2</sup>

**Abstract:** Cancer is one of the common diseases occurring among the people all over the world. It can be due to various reasons such as different habitats, environmental disorders etc. Cancer being detected at early stages can save millions, if effective treatment is provided. It can cause damage to any part of body. Breast cancer occurs when breast cells divide rapidly to form a lump or mass known as a tumor. The detection of the breast cancer is a challenging problem, due to the structure of the cancer cells. This project presents a threshold method, for segmenting mammographic images to detect the Breast cancer in its early stages. The threshold will be determined by clustering an image based on row and column separation. The manual analysis of this samples are time consuming, inaccurate and requires intensive trained person to avoid diagnostic errors. The segmentation results will be used as a base for a Computer Aided Diagnosis system for early detection of cancer from mammographic images which will improve the chances of survival for the patient. Furthermore, the probability of the tumor to identify its type is also taken by us i.e. benign, suspicious or malignant.

**Keywords:** Biclustering, Image Processing, Benign, Malignant.

## I. INTRODUCTION

refers to the uncontrolled multiplication of a group of cells in a particular location of the body. A group of rapidly dividing cells may form a lump micro calcifications or architectural distortions which are usually referred to as tumors. 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. This method involves X-ray imaging of the breast. Screening mammography examinations are performed on asymptomatic women to detect early, clinically unsuspected breast cancer.[2] 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 230 000 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 microcalcifications, masses and architectural distortions 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[4]. 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 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. Classifier systems have been widely used in medical diagnosis.

In computer vision, image segmentation is the process of partitioning a digital image into multiple segments such as sets of pixels, also known as super pixels[4]-[7]. The goal of segmentation is to simplify and change the representation of an image into something that is more meaningful and easier to analyze. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture.

Biclustering was first used by Cheng and Church in gene expression data analysis. It belongs to a distinct class of clustering algorithms that perform synchronous row-column clustering[5]. Biclustering is an important technique in two way data analysis. Biclustering is an extremely useful data mining tool used for identifying patterns, where different genes are correlated based on the subset of conditions in the gene expression dataset. This methodology is effectively applied to extract finer details about the behavior of genes under certain experimental samples. Thus Biclustering can be very well used for detecting cancer.

## II. MATERIALS AND METHODS

A mammogram is an x-ray exam of the breast that's used to detect and evaluate breast changes.

X-rays were first used to examine breast tissue a century ago, by the German surgeon,

Albert Salomon. But modern mammography has only existed since the late 1960s, when special x-ray machines were designed and used just for breast imaging. Since then, the technology has advanced, and today's mammogram is very different even from those of the 1980s and 1990s.

Screening mammograms are x-ray exam of the breasts that are used for women who have no breast symptoms or signs of breast cancer (like a previous abnormal mammogram). The goal of a screening mammogram is to find breast cancer when it's too small to be felt by a woman or her doctor.

Any computer-aided diagnosis system is based on artificial intelligence (AI) techniques. The pipeline used in a CAD system for breast cancer detection is similar to any other AI-based system and consists of preprocessing, breast region segmentation, feature extraction and classification. A major difference between computer-aided detection of breast cancer and other AI-based technologies is that breast cancer detection using CAD systems requires human intervention for interpreting the final results.

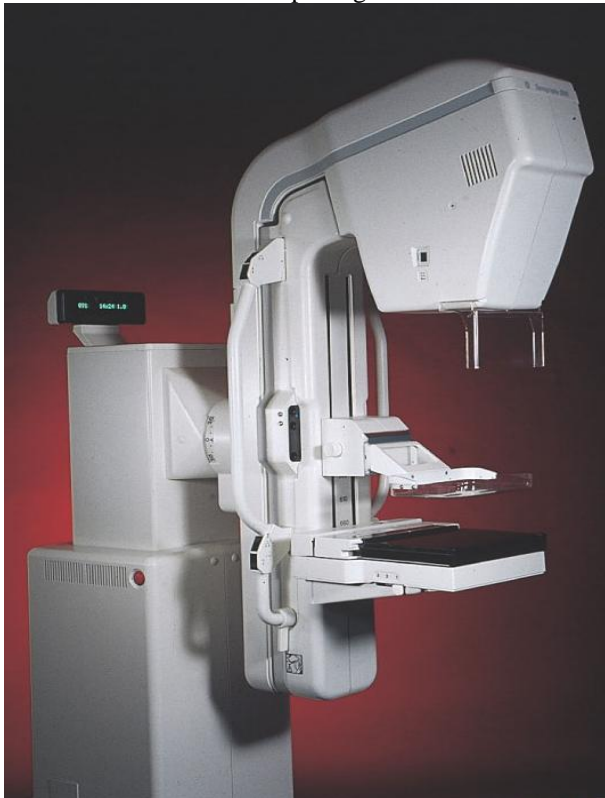


Fig.1.Mammography Unit

Fig1. Shows the mammography unit which is the Equipment C-arm SID is fixed at 24 – 26". It is dedicated units have high-frequency generators. It Provide more precise control of kVp, mA, and exposure time. This is specially designed to produce high-contrast and high-resolution images.

Preprocessing of mammograms is done to improve the contrast of mammograms which will be helpful in further stages of the detection pipeline. This step also includes denoising of the images. Segmenting the breast region from pectoral muscle and surrounding regions is carried out in order to make it easier to extract the suspicious tissues from breast segments. Feature extraction and classification steps are similar to other AI and pattern recognition systems with not much of a difference between commonly used methods.

## III. PROPOSED METHOD

In this the detail study of block diagram of proposed method is present. Fig.2. shows the typical block diagram of proposed method.

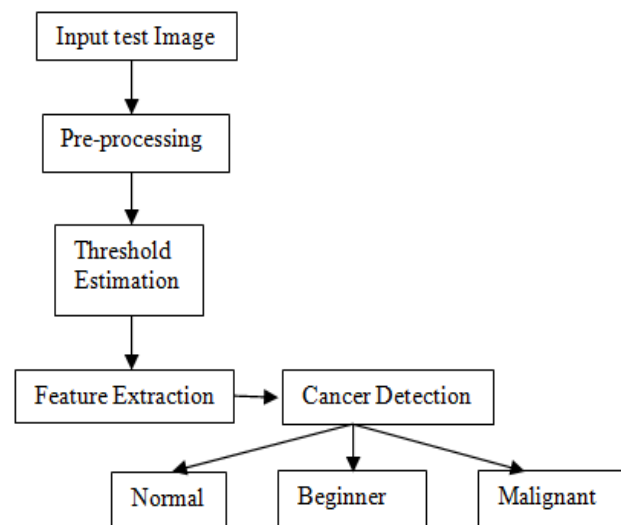


Fig.2. Block Diagram

### 1) *Input image:-*

Input Image is applied to the system which is in jpeg format from the database of the system.

### 2) *Preprocessing:-*

Image pre-processing is the term for operations on images at the lowest level of abstraction. These operations do not increase image information content but they decrease it if entropy is an information measure. The aim of pre-processing is an improvement of the image data that suppresses undesired distortions or enhances some image features relevant for further processing and analysis task. Image preprocessing use the redundancy in images. Neighboring pixels corresponding to one real object have the same or similar brightness value. If a distorted pixel can be picked out from the image, it can be restarted as an average value of neighboring pixels. Image pre-processing methods can be classified into categories according to the size of the pixel neighborhood that is used for the calculation of a new pixel brightness.

#### a. *Resizing:*

In the preprocessing, the region of interest (ROI) was selected from the digital mammograms images. The data used in our experiments were obtained from the Mammographic Image Analysis Society (MIAS) database [9]. It consists of 16 images belonging to normal, benign

and malignant classes. All images have a resolution of 1024×1024 pixels and 8-bit accuracy (gray level). They also include the locations of any abnormalities that may be present. ROI of size 512×512 pixels is extracted with breast cancer centered in the window.

*b. Cropping:*

The next step is image cropping. Some irrelevant parts of the image can be removed and the image region of interest is focused. This tool provides a user with the size information of the cropped image. Matlab function for image cropping realizes this operation interactively waiting for an user to specify the crop rectangle with the mouse and operates on the current axes. The output image is of the same class as the input image.

*3) Threshold Estimation:*

The abnormal region will be detected by finding desired threshold from significant coarse details obtained from two directional decomposition. The threshold is defined as,

$$T = 2^{\log(\text{abs}(C_{\max}))} \tag{1}$$

Where,

$C_{\max}$  – Maximum coefficient of coarse details

The segmentation will be done by,

$$\text{Seg} = \begin{cases} 255 & \text{if } |D| > T \\ 0 & \text{otherwise} \end{cases}$$

$D$  – input data

*4) Significant Data Extraction:*

It is an algorithm used here to identify an abnormal region from an image. It belongs to a distinct class of clustering algorithms that perform synchronous row-column clustering. Biclustering algorithms have also been proposed and used in some application fields such as co-clustering, bi-dimensional clustering, two-mode clustering and subspace clustering. Biclustering is an important technique in two way data analysis. Biclustering is an extremely useful data mining tool used for identifying patterns, where different genes are correlated based on the subset of conditions in the gene expression dataset. This methodology is effectively applied to extract finer details about the behavior of genes under certain experimental samples. Thus Biclustering can be very well used for detecting cancer.

Clustering methods can be applied to either the rows or the columns of the feature matrix, separately whereas Biclustering methods perform clustering in the two dimensions simultaneously. This means that clustering methods derive a global model while Biclustering algorithms produce a more effective local model.

The goal of a statistical pattern recognition technique is to choose those features that allow pattern vectors belonging to different categories to occupy compact and disjoint regions in a  $d$ -dimensional feature space [38]. The effectiveness of the feature set is determined by how well patterns from different classes can be separated. The decision boundaries for separating patterns or features belonging to different classes are determined by the probability distributions of these patterns. As a result, it is essential to formulate the patterns in such a way that they can be classified in the most accurate and computationally efficient manner. A main step in any kind of pattern

recognition problem is the representation of data in a reduced number of dimensions. This is carried out for a number of reasons such as improved classification, stable representation or ease of computation. Geometrical methods such as principal component analysis and multidimensional scaling are used for representing data in a reduced dimension. These methods, apart from several others such as linear discriminant analysis and Karhunen-Loeve expansion, are collectively referred to as feature selection and feature extraction methods [39]. In any classification task, it is imperative to identify those variables that do not contribute to the classification. It is essential to get rid of these variables which do not contribute to the effectiveness of the classification.

To put it in perspective, for feature selection, the optimization is over the set of all possible subsets of size  $d$  of the  $p$  possible measurements  $x_1, x_2, \dots, x_p$ . Thus, we seek the  $X_d$  subset for which

$$J(\bar{X}_d) = \max_{X \in \chi_d} J(X) \tag{2}$$

In feature extraction, the class of transformation is usually specified and we seek the transformation  $A$  for which

$$J(\bar{A}) = \max_{A \in \Lambda} J(A(x)) \tag{3}$$

where  $\Lambda$  is the set of allowable transformations.

**IV. EXPERIMENTAL RESULTS**

This section describes the experimental results of proposed method.

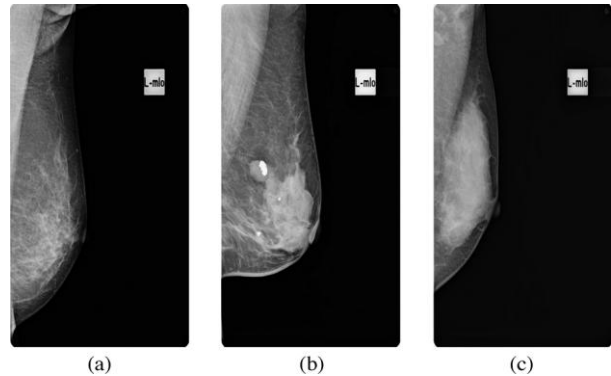


Fig.3. Typical mammogram images (a) normal, (b) benign, and (c) malignant.

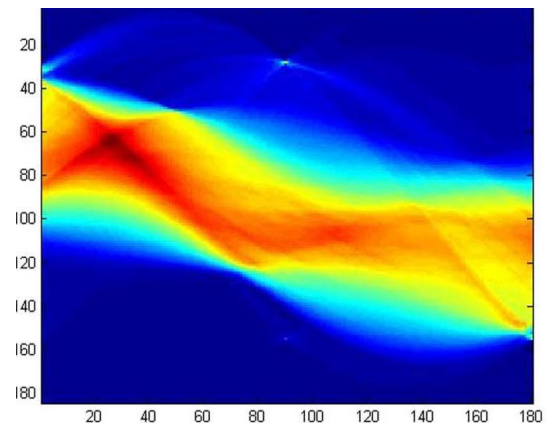


Fig. 4. Features extracted for mammogram images (a) normal.

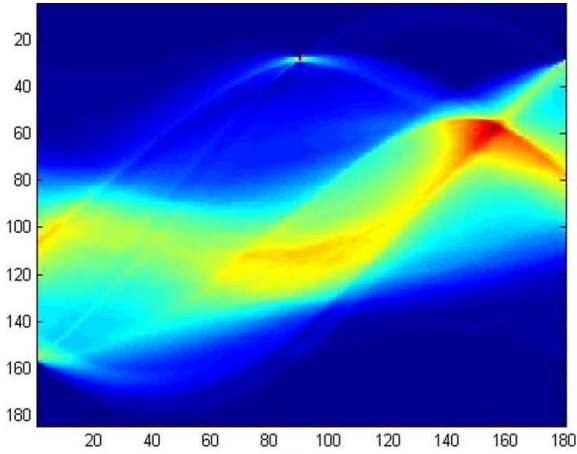


Fig. 4. Features extracted for mammogram images(b) benign.

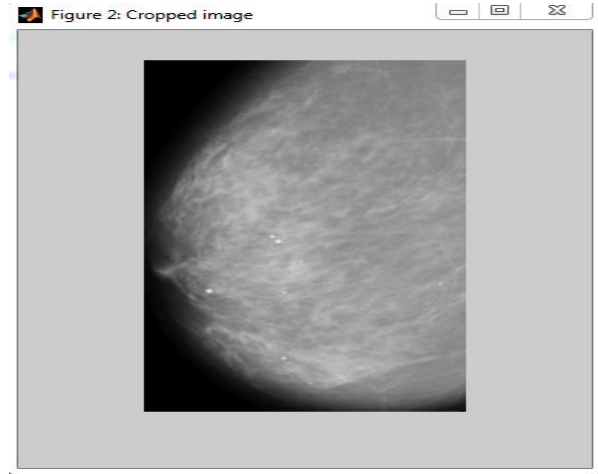


Fig.6. Cropped image

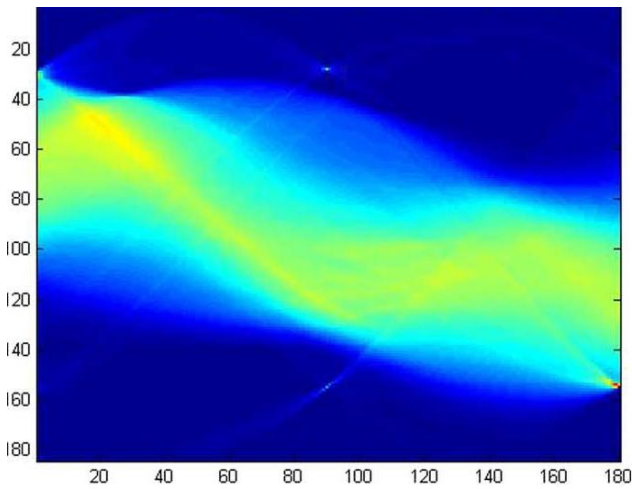


Fig. 4. Features extracted for mammogram images (c) malignant.

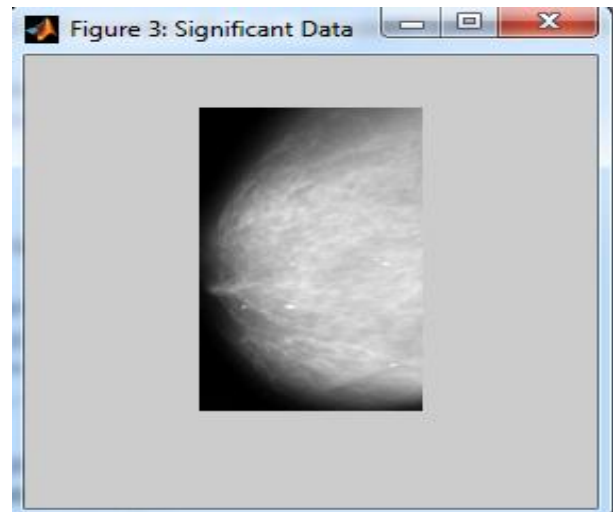


Fig.7. Output after vertical and horizontal decomposition

Fig. 4 shows an example of a feature extraction algorithm (trace transform) presenting good differentiation between the normal, benign and malignant mammograms.

Fig. 5 shows an example of an input image which is taken from the patient. Fig.6 shows the cropped image of the input image and fig.7 shows the output after vertical and horizontal decomposition. Finally the fig.8 shows the segmented output for cancer detection.

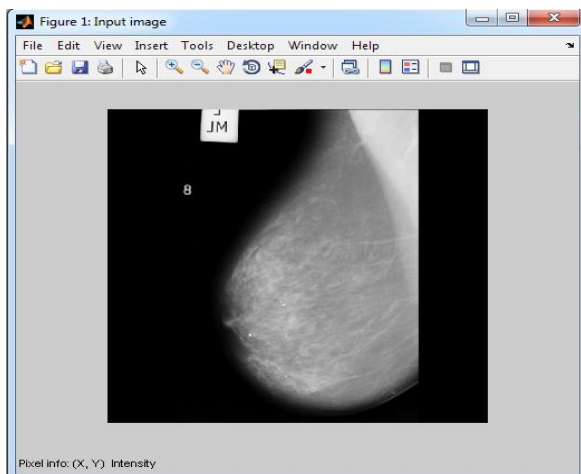


Fig.5.Input image

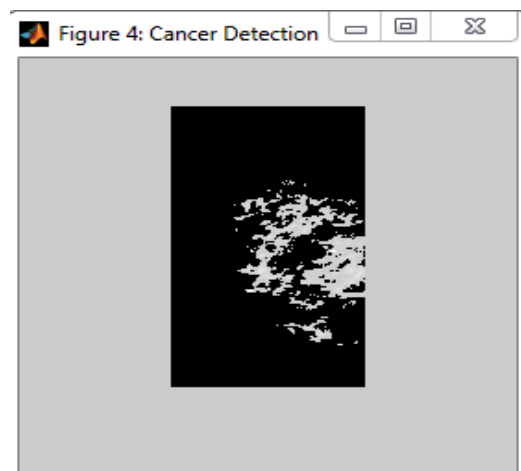


Fig.8. Segmented output

## V. CONCLUSION

Biclustering is a relatively young area & it has a great potential to make significant contributions to biology and to other fields. Proposed algorithm analyzes data and gives more accurate results. The result shows that the method implemented is an effective method for mammographic cancer detection. The experimental results demonstrate the cancer detection efficiency and effectiveness of the system.

Here, the best results obtained are around 90%, which is not sufficient enough for implementation in clinical trials. Though better results have been reported in literature, it is noted that results better than these are obtained on specific datasets which cannot be generalized to a wide array of data which can be seen in actual practice. It is also seen that the performance of conventional methods such as decision tree with texture features do not provide results as good as nonconventional techniques, such as Eigen faces approach or a model-based vision algorithm and region based edge-profile acutance measure. This can be due to the fact that conventional techniques are tuned to act on specific datasets while nonconventional techniques developed with the nature of mammographic datasets in mind can be well adapted to a wide array of mammographic data.

## REFERENCES

- [1] Brijesh Verma and John Zakos" A Computer-Aided Diagnosis System for Digital Mammograms Based on Fuzzy-Neural and Feature Extraction Techniques" IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 5, NO. 1, MARCH 2001
- [2] Liyang Wei, Student Member, IEEE, Yongyi Yang\*, Senior Member, IEEE, Robert M. Nishikawa, Miles N. Wernick, Senior Member, IEEE, and Alexandra Edwards" Relevance Vector Machine for Automatic Detection of Clustered Microcalcifications" IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 24, NO. 10, OCTOBER 2005
- [3] Arnau oliver, Jordi freixenet, Robert mart '1, Josep pont, Elsa perez, erikare. Denton, and reyer zwigelaar" A Novel Breast Tissue Density Classification Methodology" iee transactions on information technology in biomedicine, vol. 12, no. 1, january 2008
- [4] Ji'ri Grim, Petr Somol, Michal Haindl, Senior Member, IEEE, and Jan Dane" Computer-Aided Evaluation Of Screening Mammograms Based On Local Texture Models" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 18, NO. 4, APRIL 2009
- [5] S.N. Deepa and B. Aruna Devi" A survey on artificial intelligence approaches for medical image classification" Indian Journal of Science and Technology Vol. 4 No. 11 (Nov 2011) ISSN: 0974- 6846
- [6] Uma Sahu, Antony John, Ancy Alphonso, Amit, Amiya Tripathy" Cancer Detection using Biclustering" 2013 International Conference on Computer Communication and Informatics (ICCCI - 2013), Jan. 09 – 11, 2013, Coimbatore, INDIA
- [7] H. Kobatake, M. Murakami, H. Takeo, and S. Nawano, "Computerized detection of malignant tumors on digital mammograms," *IEEE Trans. Med. Imag.*, vol. 18, no. 5, pp. 369–378, May 1999, 10.1109/42. 774164.
- [8] E. D. Pisano, E. B. Cole, B. M. Hemminger, M. J. Yaffe, S. R. Aylward, A. D. A. Maidment, R. E. Johnston, M. B. Williams, L. T. Niklason, E. F. Conant, L. L. Fajardo, D. B. Kopans, M. E. Brown, and S. M. Pizer, "Image processing algorithms for digital mammography: A pictorial essay," *Radiographics*, vol. 20, pp. 1479–1491, Sep. 2000.
- [9] N. Jamal, K. H. Ng, and D. McLean, "A study of mean glandular dose during diagnostic mammography in Malaysia and some of the factors affecting it," *Br. J. Radiol.*, vol. 76, pp. 238–245, 2003.
- [10] N. Jamal, K.-H. Ng, D. McLean, L.-M. Looi, and F. Moosa, "Mammographic breast glandularity in Malaysian women derived from radiographic data," *Amer. J. Roentgenol.*, vol. 182, pp. 713–717, 2004.
- [11] 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, 10.1109/34.824819.
- [12] A. R. Webb, *Statistical Pattern Recognition*, 2nd ed. New York, NY, USA: Wiley, 2002, 0470845147.
- [13] J. Scharcanski and C. R. Jung, "Denoising and enhancing digital mammographic images for visual screening," *Computerized Med. Imag. Graphics*, vol. 30, no. 4, pp. 243–254, Jun. 2006, 10.1016/j.compmedimag. 2006.05.002, 0895-6111.
- [14] N. Karssemeijer, "Adaptive noise equalization and recognition of microcalcification clusters in mammograms," *Int. J. Pattern Recog. Artif. Intell.*, vol. 7, pp. 135713–135776, 1993.
- [15] S. Theodoridis and K. Koutroubas, *Pattern Recognition*, 3rd ed. New Providence, NJ, USA: Elsevier, 2006, 0-12-369531-7.
- [16] A. Petrosian, H. P. Chan, M. A. Helvie, M. M. Goodsitt, and D. D. Adler, "Computer-aided diagnosis in mammography: Classification of mass and normal tissue by texture analysis," *Phys. Med. Biol.*, vol. 39, pp. 2273–2288, 1994.
- [17] R. M. Haralick, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. 3, pp. 610–621, Dec. 1973.
- [18] R. M. Rangayyan, N. M. El-Faramawy, J. E. L. Desautels, and O. A. Alim, "Measures of acutance and shape for classification of breast tumours," *IEEE Trans. Med. Imag.*, vol. 16, no. 12, pp. 799–810, Dec. 1997.
- [19] N. Karssemeijer and G. te Brake, "Detection of stellate distortions in mammograms," *IEEE Trans. Med. Imag.*, vol. 15, no. 10, pp. 611–619, Oct. 1996.
- [20] Guliato, R. M. Rangayyan, W. A. Carnielli, J. A. Zuffo, and J. E. L. Desautels, "Segmentation of breast tumors in mammograms by fuzzy region growing," in *Proc. 20th Annu. Int. Conf. IEEE Engineering Medicine Biology Soc.*, Hong Kong, Oct. 29–Nov. 1 1998, pp. II:1002–II:1004.
- [21] S. L. Kok, J. M. Brady, and L. Tarassenko, "The detection of abnormalities in mammograms," in *Proc. 2nd Int. Workshop Digital Mammography*, York, U.K., Jul. 10–12, 1994, pp. 261–270.
- [22] S. Timp, C. Varela, and N. Karssemeijer, "Temporal change analysis for characterization of mass lesions in mammography," *IEEE Trans. Med. Imag.*, vol. 26, no. 7, pp. 945–953, Jul. 2007, 10.1109/TMI.2007. 897392.
- [23] L. Ke, N. Mu, and Y. Kang, "Mass computer-aided diagnosis method in mammogram based on texture features," in *Proc. 3rd Int. Conf. Biomedical Engineering and Informatics (BMEI)*, Oct. 16–18, 2010, vol. 1, pp. 354–357, 10.1109/BMEI.2010.5639515.