I.J. Image, Graphics and Signal Processing, **2023, 4, 44-58**
Published Online on August 8, 2023 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2023.04.04

Mammogram Pre-processing Using filtering methods for Breast Cancer Diagnosis

Shah Hemali
Institute of Technology, Nirma University, Ahmedabad, Gujarat, 382481, Ahmedabad, Gujarat, India
Email: 21ptphde242@nirmauni.ac.in
ORCID iD: https://orcid.org/0000-0002-8203-0937

Agrawal Smita
Institute of Technology, Nirma University, Ahmedabad, Gujarat, 382481, Ahmedabad, Gujarat, India
Email: smita.agrawal@nirmauni.ac.in
ORCID iD: https://orcid.org/0000-0003-0486-8045

Parita Oza*
Institute of Technology, Nirma University, Ahmedabad, Gujarat, 382481, Ahmedabad, Gujarat, India
Email: parita.prajapati@nirmauni.ac.in
ORCID iD: https://orcid.org/0000-0001-9637-6157
*Corresponding Author

Sudeep Tanwar
Institute of Technology, Nirma University, Ahmedabad, Gujarat, 382481, Ahmedabad, Gujarat, India
Email: sudeep.tanwar@nirmauni.ac.in
ORCID iD: https://orcid.org/0000-0002-1776-4651

Ahmed Alkhayyat
College of Technical Engineering, the Islamic University, Najaf, 54001, Najaf, Iraq
Email: ahmedalkhayyat85@gmail.com
ORCID iD: https://orcid.org/0000-0002-0962-3453

Received: 15 July, 2022; Revised: 26 August, 2022; Accepted: 06 September, 2022; Published: 08 August, 2023

Abstract: Cancer is the second most found disease, and Breast cancer is the most common in women. Breast cancer is curable and can reduce mortality, but it needs to be identified early and treated accordingly. Radiologists use different modalities for the identification of Breast cancer. The superiority of Mammograms over other modalities is like minor radiation exposure and can identify different types of cancers. Therefore, mammograms are the most frequently used imaging modality for Breast Cancer Diagnosis. However, noise can be added while capturing the image, affecting the accuracy and analysis of the result. Therefore, using different filtering techniques to pre-process mammograms can enhance images and improve outcomes. For the study, the MIAS dataset has been used. This paper gives a comparative study on filters for Denoising and enhancement of mammograms. The study focuses on filters like Box Filter, Averaging filter, Gaussian Filter, Identical Filter, Convolutional 2D Filter, Median Filter, and Bilateral Filter. Performance measures used to compare these filters are Mean Squared Error (MSE), Structural Similarity Index Measure (SSIM), and Peak Signal-to-noise Ratio (PSNR). All Performance measures are evaluated for all images of MIAS dataset and compared accordingly. Results show that Gaussian Filter, Median Filter, and Bilateral Filter give better results than other filters.

Index Terms: Mammogram Pre-processing, Mammogram Enhancement, Breast Cancer Diagnosis, Filtering techniques, MIAS Filtering, MIAS pre-processing

1. Introduction

In the current era, Breast Cancer is the second-ranked majorly found disease in women [1]. According to WHO reports, "In 2020, there were 2.3 million women diagnosed with breast cancer and 685 000 deaths globally [2]. In cancer, the cells grow uncontrollably and make a tumor, also called a lump. There are two types of tumors: (1) Benign and (2) Malignant. In Benign, lumps are not cancerous and do not spread to other parts. In contrast, malignant lumps are cancer and spread to other body parts.

Early diagnosis of lesions affected and giving a treatment can lower the mortality rate. Breast cancer is curable if detected in an early phase. There are two ways to identify tumors. One is the self-examination, in which different symptoms can be seen and with regular observations can identify the disease. Symptoms may vary according to person, but commonly seen symptoms in patients are Rashes around the nipple, Redness of the skin, Inverted nipple, Change in the size of the nipple, and Armpit or breast pain. Another option is a medical checkup for tumor detection. Radiologists use imaging techniques to detect tumors. Different imaging modalities used by Radiologists for Breast cancer diagnosis are MRI, Mammograms, Ultrasound, Tomosynthesis, Thermography, and Biopsy.

Mammograms are commonly used imaging techniques because of their low cost, trivial exposure to radiations and ability to detect tumors in the early stage of a Breast cancer diagnosis. However, identifying affected lesions from different imaging scans depends on the Radiologist. CAD systems are used as a second opinion tool to avoid mistakes and identify affected areas from mammograms [36]. In collected mammograms, many abnormalities are misinterpreted for different reasons like pectoral muscles, poor image quality, artifacts, and noise—all of these results in a high False-positive rate [3]. Pre-processing improves the quality and enhances mammograms for interpretation, which results in improved visual characteristics like margin, contrast, and fine edges.

The first stage of any CAD system is image pre-processing. The noise and data that is not useful for the study will be removed by pre-processing the image. Pre-processing helps to focus on the required information from the whole image. For example, Micro-calcification and masses are vital signs of malignancy in the breast [4]. Micro-calcification is tiny spots of calcium spread in the breast. Calcification is very little, so it's hard to distinguish them from surrounding tissues in young women with dense breasts. Another lesion is mass, characterized using different attributes like shape (round, lobular, oval, and irregular), size, location, and margin (obscured, indistinct, and speculated). These attributes are used to classify as benign or malignant. Mass identification is more complicated than calcification because mass has low contrast compared to calcification. In addition, mass is more similar to the structure of normal breast tissue. All these attributes are crucial in correctly detecting and classifying the lesion. However, because of noise in the images, important information can be suppressed. As a result, the vital information which can help radiologists can't be identified. Therefore, preprocessing images and removing noise can help in a more informative view of data.

The paper's main objective is to give insight into different mammogram enhancement techniques, different kinds of noise in the mammogram, applying different filters, and comparing their results concerning different performance measures. So this can help to select a filter for preprocessing mammogram images before using them for classification.

By applying different filters and Enhancement techniques, mammogram quality can be improved and focus on more detailed data. The contribution of the paper is as follows.

1) MIAS database is used for analysis. The paper presents a comparative study highlighting the impotence of pre-processing mammograms for Breast cancer diagnosis.
2) Different filters like Average, Box, Gaussian, Median, and Bilateral filters are compared and applied to MIAS dataset for a comparative study of pre-processing the images.
3) Different types of noise in an image and their removal techniques are discussed.
4) Different filters were analyzed using performance measures like MSE, SSIM, and PSNR for the MIAS database. Results show that Median, Gaussian, and Bilateral filters give better results.

Filtering is one of the methods for pre-processing mammograms. Researchers use cropping, spatial and transform domain filters, CNN-based methods, sharpening, and smoothing for image enhancement.
In [13], author proposed anisotropic diffusion with Wavelet results in enhanced result. Different enhancement techniques result in different outputs as per the operations performed on the image. Filtering methods are most frequently used to remove noise and enhance image quality. Filters are conceptually simple with enhanced results, so they are used by many researchers for image pre-processing [28].

The paper reviews many articles from books and online scientific databases like Scopus, Web of Science, and IEEE Xplore. As shown in figure 1, the flow of the paper is as follows. The first section gives the study introduction. The following section Study Background, gives an overview of the areas related to work. Then, the Mammogram pre-processing section discusses different techniques which can be applied to enhance the image. Finally, the result and Discussion section shows the experiments performed on the MIAS dataset by applying different filters and comparing results using various performance measures.

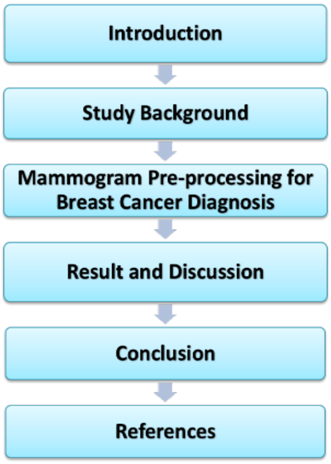


Fig. 1. Paper Organization

2. Study Background

Cancer is widespread these days, and Breast Cancer is the second leading disease found in women as per WHO. This paper's main objective is to study different types of noise in mammograms, filtering methods to enhance images, and compare the results of different filters based on performance measures for the MIAS dataset. This section gives an insight into the essential technologies related to the study of a Breast Cancer diagnosis. It describes areas like Medical Imaging, CAD systems, Breast Cancer, Mammogram Images, and Different types of noise in the image related to Breast Cancer Diagnosis.

As our study area is Breast Cancer Diagnosis, the paper discusses different medical imaging modalities used by Radiologists and highlights why mammograms have been chosen for study purposes. The basic steps of the CAD system are discussed here. The section also shows various mammogram data sets used by researchers for Breast cancer diagnosis. Different types of noise which can be added to an image while generating are highlighted here. This section background information of all these areas, which are required to get a view inside the study of mammogram pre-processing using filtering methods for Breast Cancer Diagnosis.

A. Medical Imaging

Radiologists use different Medical Imaging techniques for Breast cancer diagnosis. The noise can be added to these images while generating or processing the images [5]. Using other pre-processing techniques can remove noise and enhance images, resulting in more intractable and informative data. Different imaging techniques used for cancer diagnosis are Digital Mammography, Digital Tomosynthesis, Breast Ultrasound, Breast MRI, Biopsy, and Fine Needle Aspiration [6]. Of all these imaging methods, Mammography, Ultrasound, and Microwave imaging are comparatively low costs for regular screening. Most research work is done using Mammography because it is relatively low cost, uses intense radiation, and is more effective for diagnosing breast cancer in the early stage [34].

B. CAD System.

CAD tools can be used as a second opinion tool by a Radiologist for Breast Cancer Diagnosis. Different Medical images can be used as input to the CAD system for analysis. In our paper, mammograms are used for the diagnosis process. As shown in figure 2, for any CAD system, the basic steps are image pre-processing,

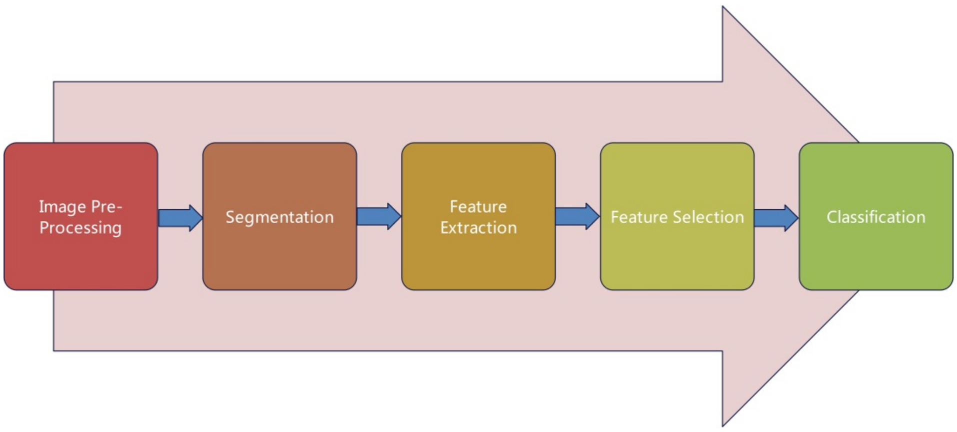


Fig. 2. Computer-Aided Diagnosis steps for any system.

Segmentation, feature extraction, selection, and classification. While capturing a digital image, noise is added to the original image. The CAD system is vulnerable to noise, generating problems locating important information in mammograms. We can improve an image by pre-processing and applying different filtering and enhancement techniques. Pre-processed mammograms can be used for cancer diagnosis, and the system can focus on important information related to Breast Cancer. Computer-aided detection (CADe) helps identify the area of interest and classify it as mass or non-mass, whereas CADx identifies this mass as benign or malignant [23].

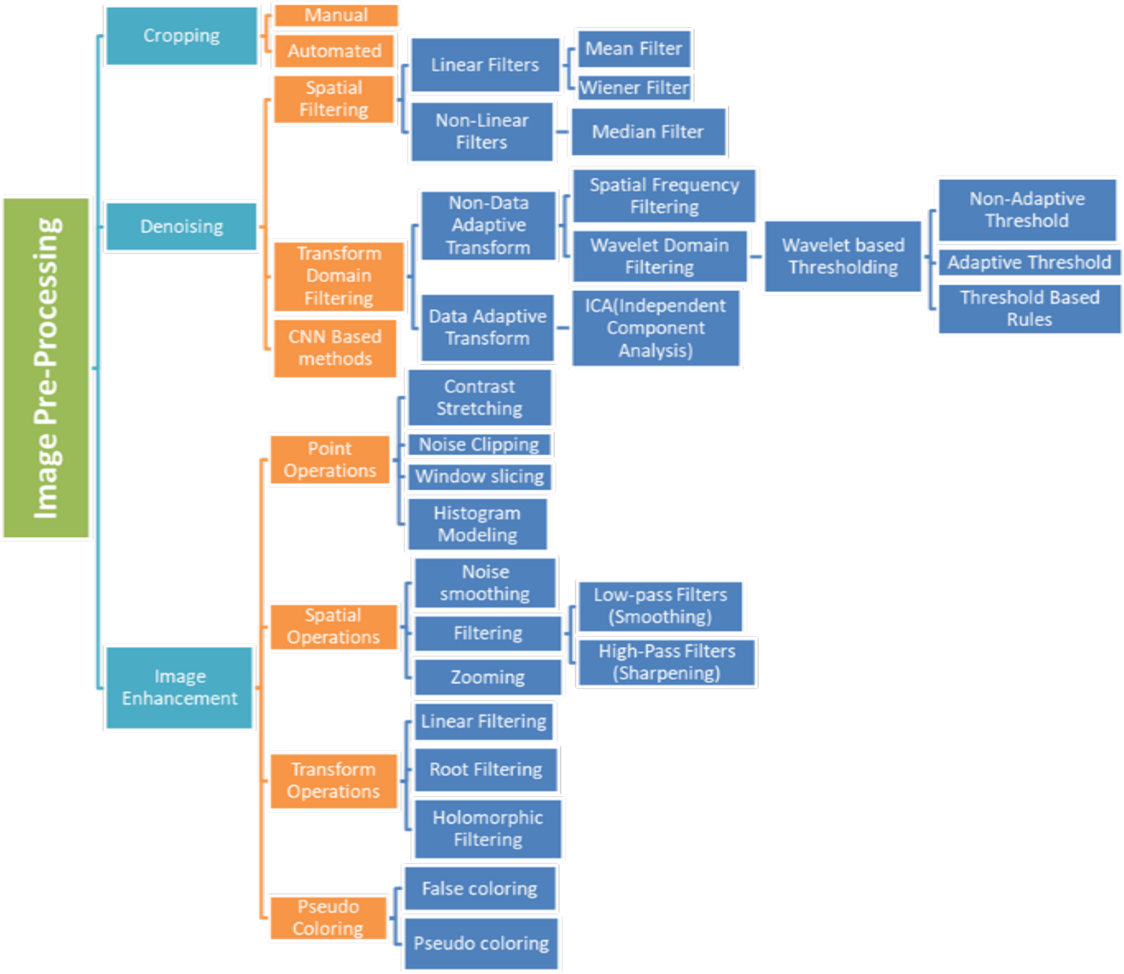


Fig. 3. Mammogram pre-processing techniques

C. *Breast Cancer*

Cancer is the top-ranked disease in the current era. Breast cancer is the second-ranked disease in women [1]. In cancer, cells grow uncontrollably and form a tumor. There are two types of tumor: Benign and malignant. Benign tumors are not cancerous and do not spread to other body parts. In contrast, the malignant tumor is cancerous and can grow in different body parts. Most commonly found Breast Cancers are Invasive Ductal Carcinoma (IDC) and Invasive Lobular Carcinoma (ILC). In diagnosis, 70-80 percent have Invasive Ductal Carcinoma (IDC) [29]. Diagnosis can be made in two ways: self or clinical examination. In self-examination, different symptoms are observed, like a new lump in the breast or underarm, thickening, swelling, irritation, and redness.

In clinical examination, radiologists use mammograms or other imaging techniques for diagnosis. Different breast abnormalities like mass, architectural distortion, calcification, and asymmetry are vital in tumor identification [16].

D. *Breast Cancer Data sets*

MIAS (Mammography Image Analysis Society): - It is a database for research purposes, and United Kingdom copyright laws protect it. In MIAS, data is images and labels that can be used for research. It is also called mini-MIAS and consists of 322 digitized films. The images are in.PGM format with each having 1024 x 1024-pixel resolutions. The MIAS database has seven columns or features like Reference number, Background tissue, Abnormality present, severity, coordinates, and Radius. In our study, we used the MIAS database [16]. Different filters are applied and analyzed; the result for both Benign and Malignant data is based on the performance measures like MSE, PSNR, and SSIM.

DDSM (Digital Database for Screening Mammography): - This project was a grant from the Breast Cancer Research Program of the U.S. Army Medical Research and Materiel Command. There are 2620 images stored in 16-bit PGM format and used for research purposes [25].

WBCD (Wisconsin Breast Cancer Database): - In this, the breast mass features are computed from a digitized image using a fine needle aspirate (FNA) method. Ten real-valued features are calculated for each cell nucleus: Radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension [26].

BreakHis (Breast Cancer Histopathological Database): - It has 9109 microscopic images of breast tumor tissue, which have been collected from different82 patients with different magnifying factors. It is divided into two main groups, Benign and Malignant [27].

E. Different types of noise in an image

Noise is the variation in the original image's color or information, which negatively affects the quality of an image. It might be added during the acquisition, processing, or transmission of mammograms, which degrades the quality of an image. The different types of noise added to mammograms are Gaussian, Salt-Paper, Speckle, Impulse, and Poisson noise [7]. Noise can be additive or multiplicative in equations-1 and 2 respectively.

$$AdditiveNoise: \ c\,(x,y) = a\,(x,y) + b\,(x,y) \quad (1)$$

$$MultiplicativeNoise: \ c\,(x,y) = a\,(x,y) * b\,(x,) \quad (2)$$

Gaussian Noise:

It is an additive, statistical noise. It is also known as Gaussian distribution because its probability density function is a normal distribution. The function g(x) shows in equation-3, the Gaussian distribution. It is also known as electronic noise because it is generated from amplifiers or decoders.

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\frac{-1}{2}\left(x-\mu^2\right) \quad (3)$$

Impulse noise:

It is generated because of a sudden change in the image signal. It is also known as spike noise, impulsive noise, or fat-tail distribution noise, which looks like black and white pixels in the image [9].

Salt noise:

It is added by the addition of random bright (255-pixel value) all over the image.

Paper noise:

It is added by the addition of random dark (0-pixel value) all over the image.

Salt and pepper noise: It is added by adding random bright (255-pixel) and random dark (0-pixel value) all over the image. If total noise density is p, then salt noise and paper noise have p/2 noise density, which can be represented mathematically in equation-4 as follows.

$$Y_{ij} = \begin{cases} 0 \\ 255 \end{cases} \quad (4)$$

Where Y_{ij} - Noisy image pixel, p-total noise density, p=p1+p2

Speckle noise:

It is generated by the multiplication of random pixels with original image pixels. Its multiplicative noise degrades image quality—incoherent imaging systems like a laser, RADAR, etc. [10].

Poisson noise:

It is seen because electromagnetic radiation is injected into the patient's body, and the resulting image will have fluctuation of photons arbitrarily. The mean and variance of Poisson distribution are identical, so it is assumed that noise has unity variance [8].

3. Mammogram Pre-processing for Breast Cancer Diagnosis

By pre-processing and applying different image enhancement techniques represent in figure-3, mammograms can be improved, which can help to identify calcification and mass in the breast. As shown in figure 3, there are mainly three types we can pre-process mammograms for improvement and extract more detailed information. First, Cropping can be manual or automated. By applying Cropping, we can focus on the Region of Interest and improve with results.

The second is noise removal, also called denoising techniques. We can use these techniques to remove noise from the image and identify more detailed information like calcification or structural distortions in mammograms. Different filters are used for denoising images. The third category is image enhancement using techniques like low and high pass filters or point and spatial operations. This section describes various methods to enhance mammograms.

A. Cropping

In mammogram images, background tissues and pectoral muscles occupy an extensive area. Only 30 percent of the site from the whole mammogram has breast tissue. Figure 4 describes mammogram with labels, artifacts, Pectoral muscle, and other parts of Breast. In the non-breast, tissues are firmly linked, affecting further diagnosis. So, removing unrelated content helps focus on the affected region called Region Of Interest (ROI) [35]. The area of interest can be extracted using cropping techniques. Cropping can be done in two ways: manual or automated cropping. Both manual and automatic cropping remove the following from mammogram images.

1) Removal of artifacts and Labels:

Mammograms are of two types—Screen-film and digital. Different artifacts on screen-film mammography are dust, pick-off, and processor roller, static and fogging artifacts. Commonly found artifacts in Digital mammography are as follows: patient-related artifacts: motion artifacts, antiperspirant artifacts, and thin breast artifacts. Hardware-related artifacts: field in-homogeneity, detector-associated artifacts, collimator misalignment, underexposure, grid lines, grid misplacement, vibration artifact. Software processing artifacts: “breast-within-a-breast” artifact, vertical processing bars, loss of edge, high-density artifacts [11]. In pre-processing of mammograms, it is advisable to remove labels and artifacts, which enhances image quality for identifying tumors in the breast.

2) Removal of Pectoral muscle:

Pectoral muscles are mass tissues used to support the breast body. It is essential to remove the Pectoral muscle because this will reduce the size of the image and save processing time and computation power [13]. It will also help to focus on the affected region.

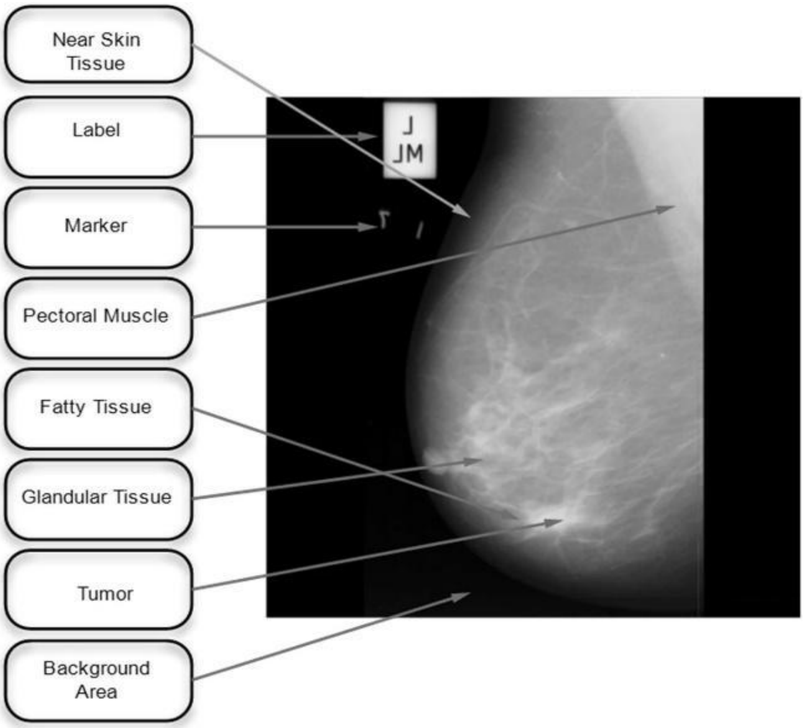


Fig. 4. Mammogram with labels, artifacts, Pectoral muscle, and other parts of Breast [41]

B. Denoising

Different techniques are used to remove noise or distortion from the image, which results in more informative data or images. Filtering can be used to denoise, modify, or enhance an image. Denoising techniques can be divided into three major categories: Spatial Domain Filters, Transform Domain Filters, and CNN-based.

1) Spatial Domain Filters:

Spatial Domain Filters directly work on pixels. There are two types: Linear Filters and Non-linear Filters. Filters are used for functions like re-sampling, blurring, interpolation, and noise reduction in the spatial domain [12]. Filter selection depends upon the type and amount of noise present in the image because different filters work efficiently on particular types of noise. There are two types of filters in the spatial domain: linear and non-linear.

1.1 Linear Filter:

They are also called average filters. Mean Filter and Wiener Filter are types of linear filters. They work well on blurred, sharp edges, corrupt lines, and other particular image information [40].

Mean Filter: They are the most detailed filters to apply and give smoothing to the image. The mean value of its neighbors replaces each value of a mammogram image. The problem with the mean filter is that a single pixel having a false value can affect the mean value of all pixels [9]. Arithmetic mean filters can be applied to an image using a given equation-5.

$$x = \frac{1}{n}(x_1 + \cdots + x_n) \tag{5}$$

Here x1, x_n are pixel values involved for the mean, and n is the number of pixels used for averaging. The mask specifies pixels for the mean filter.

Wiener Filter: They are used to limit the noise in the image, which is done by comparing it with the noise-less signal. It is used frequently for deconvolution and restoration. Inverse filters are used to recover the blurred image. Wiener filter provides noise smoothing, blurring as well as inverse filtering. The problem with the Wiener filter is that it destroys fine image details, sharp edges get blurred, and gives a degraded performance in the presence of noise in the mammogram image [10].

1.2 Non- Linear filter:

When non-linear features of the signal are more critical than the overall image, non-linear filters play a crucial role in noise reduction. Linear filters remove unusual noise from the image but with the cost of blurring it and making the edges invisible. Non-linear filters use memory very efficiently compared to linear filters. The median filter is the simplest non-linear filter [19].

Median Filter: It is a non-linear filter in which the corrupted pixel value is replaced by the median value of the corresponding mask window [37]. It gives the best results by vanishing random valued impulse noise in the image. It is also called an order-specific filter. The main characteristic of the median filter is that it preserves edges while removing noise [9]. It is superior in the reduction of salt-pepper noise and speckle noise. The problem with the Median filter is that it majorly shuffles the pixel intensity in neighborhoods which can be not required in many cases(see equation-6). It can also disarrange the fine details in the original image [13].

$$MdF(x_i) = Median(\| x_i \|)^2 \tag{6}$$

1.3 Other filters:

Like Spatial Median Filter (SMF) and Weighted Median Filter (WMF) are modifications over median filters. In the Median spatial filter, the depth of the middle pixel is calculated and checked for whether corrupted or not. If it is corrupted, then the only value is replaced. On the other hand, a weighted Median Filter uses weights and is easier to design and implement [18]. Figure 5 shows different filters and their effect for mdb010 mammograms of MIAS dataset.

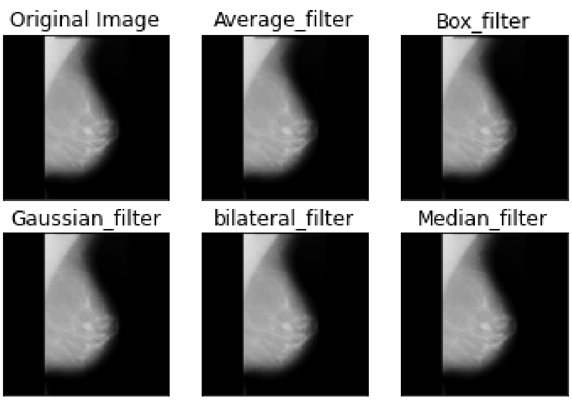


Fig. 5. Different Filters applied to image mdb010 of MIAS

2) Transform Domain Filters and Image Enhancement:

There are two types of Transfer Domain filtering. First is the Non-Data adaptive Transform, consisting of spatial filtering techniques and Wavelet Domain filtering. The second type is Data adaptive Transform, which works on Independent Component Analysis. Different Low-Pass and High-Pass filters are used to process and remove noise from the image [38]. Low-Pass filters remove higher frequency from the image and blur some image content. On the other hand, high-Pass filters remove lover frequency content from the image, eliminating overall brightness and sharpening the image [39]. Recently Independent Component Analysis (ICA) approach is gaining attraction for Denoising images by working on component analysis, factor analysis, and projection detection. In ICA signal is assumed to be Non-Gaussian. The disadvantage of ICA is its computational cost because it uses a sliding window [8].

3) CNN-based Methods: Different CNN models are used for Breast Cancer Diagnosis, and CNN shows better image pre-processing. It also improves the image resolution [24].

C. Image Enhancement:

Image quality can be improved by applying different image Enhancement techniques. The point, Spatial, and Transform operations are commonly used methods for image enhancement. Point operations are often used to change and enhance the contrast. Histogram equalization also plays a vital role in the Enhancement of mammograms. Most papers applied the CLAHE (Contrast-Limited Adaptive Histogram Equalization) technique for image enhancement and improving contrast. In [14], the author used CLAHE to enhance the contrast of the mammogram. Another is spatial operations, in which significant operations performed are noise smoothing using Low-pass and High-pass filters. Low-pass filters give a smoothing effect, while High-pass filters give a sharpening result. Other transform operations like Linear, Root and Homomorphic filtering can also apply for image enhancement.

4. Result & Discussion

This section discusses different work done in mammogram pre-processing, measuring the quality of output images using other measurements, and a comparative view of different methods for image Denoising. Here, further work done by researchers is also discussed in the related work section

A. Related Work

In [13], the author used a mini-MIAS dataset and enhanced mammograms by applying different filtering methods like Median, Weiner, anisotropic, and anisotropic. The author compared these methods using measures like Standard Deviation, PSNR, SNR, SSIM, IMAE, VAR, RMSE, and AMBE. Results show that anisotropic diffusion with wavelet filtering is best from the given techniques. The author Neha Shahare surveys different approaches which can be applied for image enhancement and pre-processing data. Different stages of image pre-processing and procedures related to that have been discussed.

In [15], the author used the MIAS dataset and the two stages Adaptive Histogram technique for image enhancement. In addition, the author used EME and entropy criteria for analysis and results in the discussion. The results show that TSAHE (Two-Stage Adaptive Histogram Equalization) gives better results than other methods. In [16], the author reviewed segmentation techniques like Region-based, Boundary based, Model-based, and deep learning. Other performances used for the measure are precision, recall, F1 score, specificity, FP/Image, accuracy, Dice similarity coefficient, and Jacquard Index. GLCM and GLRLM techniques are used for texture extraction. A Neural Network classifier using the Back-Propagation algorithm for classification. Results were compared using ROC analysis. In [18], the author presented applying CLAHE and thresholding methods on the MIAS database to detect Breast tumor boundaries. He also worked on image enhancement and segmentation using Thresholding, which resulted well. Finally, he used CII as a performance evaluation parameter. Here author worked on the system for automatic Mass detection and classification. The author used the CBIS-DDSM database. In [20], the first step of image preprocessing author used normalization and then converted the image into gray-scale. Now applied wavelet transform technique to denoise the image. Finally, using CLAHE for image enhancement.

In [21], to detect micro-calcification clusters, author used directional fractal filtering technique. It can also identify irregularly shaped micro-calcification. In the first step, detecting and locating micro calcification are done using the fractal mask, then zooming and enhancing an image at last applying segmentation using XOR, a logical operator. The author compared segmentation results with CLAHE, Atos, and Haar techniques. For result analysis, the author used measurement criteria like CII, PSNR, and ASNR.

Table 1. Comparative view of different methods for image Denoising

Ref	Dataset	Enhancement & Filtering	Segmentation	Feature Extraction	Classification	Pectoral Muscle removal	Quality Measures used
[17]	MIAS	Median, Wiener, Anisotropic with wavelet	no	no	no	no	Standard deviation, PSNR, SNR, IMAE
[2]	MIAS	Wiener, Median, CLAHE	yes	GLCM, SRDM, GRMF, SGLDM	no	no	SSIM, PSNR, SNR
[21]	MIAS	Two stage Adaptive Histogram	no	no	no	no	EME, Entropy
[8]	MIAS, DDSM, InBreast	Median filter, Morphological operators	yes	GLRLM, SRE, GLNU, RPERC	ANN	no	Accuracy, Precision, Recall F1 score, Jacquard & Dice index
[22]	DDSM	Gaussian Filter	Yes	GLCM, GLRCM	FFNN	No	Accuracy, Sensitivity, Specificity
[23]	MIAS	CLAHE	Yes	No	Yes	No	CII
[24]	MIAS	Non-linear filtering, Unsharp masking, Gradient minimization	yes	NLMUML oGMINAUTO	No	No	EME, AMEE, EMEE, AME, SDME, LogAME
[25]	CBIS-DDSM	Wavelet Transformation, Median Filter. CLAHE, Morphological operations	Yes	No	No	No	Sensitivity, Specificity, AUC
[26]	BCDR	Fractal mask, Exponential & log Transform, CLAHE	Yes	No	No	No	CII, PSNR, ASNR
[27]	MIAS	Median Filter, Morphological operations, GUTM	Yes	No	No	No	MSE, PSNR, SSIM
[1]	MIAS	Lookup tables	Yes	No	NN	No	MSE, PSNR, SSIM
[20]	MIAS	CLAHE	Yes	GLCM	SVM	Yes	Energy, Herlic contrast
[3]	MIAS	CLAHE	Yes	No	CNN	No	Accuracy, Sensitivity, Precision, F score
[12]	MIAS	Mean, Gaussian, Wiener, non-linear filters	No	No	No	No	PSNR, MSE, SSIM

The author [22] developed a hybrid denoising method that works in two steps in this paper—first, enhancing the image by applying different morphological operations. Then, he proposed a new technique called GUTM (Global Unsymmetrical Trimmed Median Filter) for image denoising. For this, the steps followed by the author are the first original image to be converted into a Gray-scale image. Now By intensity changing, the contrast of the image has been enhanced. Then Erosion is a morphological operation applied to an image. Finally, salt and paper noise is added, and applying the proposed GUTM method for image denoising has been compared to other filters like mean, median, and winner using criteria like MSE, SSIM, and PSNR. The results show that GUTM gives better results compared to other filters. In [1], the author proposed an effective preprocessing method that first removes background using Rolling Ball Algorithm. Then Petrol muscle is removed using Canny Edge Detection and Hough line transform technique. Adding noise without affecting the image makes the neural network system respond better. Furthermore, different Lookup tables are used for image enhancement, which helps outline the ROIs and regions.

In [29], author work for breast lesion segmentation and area calculation using region growing method. Active contour model is applied which make boundary clear. In [30], author work on enriching input images by generating pseudo colored images. To achieve the goal author preprocess images using CLAHE and pixel vise intensity adjustments. In [31], author proposed technique for micro calcification identification. First applied preprocessing techniques to enhance Image. After that fuzzy C-means clustering algorithm applied for suspicious region detection. In [32], author proposed hybrid approach for identification of Breast mass as benign or malignant. CBIS-DDSM dataset used for research work. In [33], author work for effective classification and segmentation technique for breast cancer detection. For preprocessing image, author used wavelet based enhancement method. Texture and statistical features are extracted from preprocessed image for further work.

B. Performance measures used and compared for MIAS pre-processing

The paper discussed different filters used to improve mammograms and noise removal. For quality assessment, various measures used for comparison in the article are MSE, SSIM, and PSNR. All the images of MIAS are applied to different filters and analyzed based on these measures. Table 1 highlights statistics concerning evaluation metrics after using other filters can be seen. Table 2 shows the resultant mammograms after applying different filters. Here the section highlights the filters like Median filter and Bilateral Filter having improved results.

Table 2. MIAS mammograms (randomly picked 4 images) results for different filters as per evaluation metrics

Evaluation Metrics	Filter	Mdb010	Mdb072	Mdb207	Mdb249
MSE	Average	8.97	8.56	19.91	20.34
	Box	8.97	8.56	19.91	20.34
	Gaussian	4	3.73	10.85	11.24
	Median	1.35	0.78	1.13	2.1
	Bilateral	2	1.26	1.39	2.1
SSIM	Average	0.98	0.99	0.98	0.98
	Box	0.98	0.99	0.98	0.98
	Gaussian	0.99	0.99	0.99	0.99
	Median	0.98	0.99	0.99	0.98
	Bilateral	0.98	0.99	0.98	0.98
PSNR	Average	38.6	38.81	35.14	35.05
	Box	38.6	38.81	35.14	35.05
	Gaussian	42.11	42.42	37.78	37.62
	Median	46.83	49.23	47.6	44.91
	Bilateral	45.12	47.14	46.72	44.9

1) MSE (Mean Squared Error):

It is the squared difference between each pixel of the original image and the processed image. MSE value near zero is upraised better (see equation-7). If MSE is zero means there is no noise in the image [1].

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (G(i,j) - P(i,j))^2 \tag{7}$$

Where, G is the original image, and P is the processed image. m, n are pixels of G(original image), i, j are rows of the pixels m and n, respectively.

In MIAS, out of 322 images, 61 images are of benign data, and 51 are Malignant. Now applied different filters are analyzed for the improvement using MSE for both benign and malignant images of MIAS database. As shown in figure 6 and figure 7, statistics after applying different filters can be seen. Here we can say that Bilateral Filters give the best result because, for MSE, a result close to zero is comparatively better.

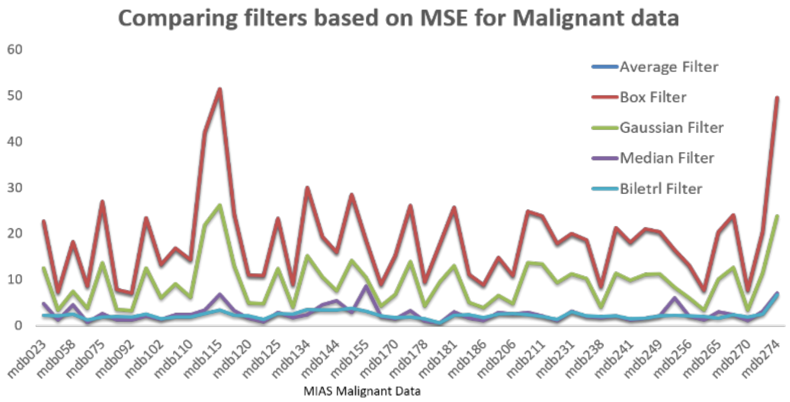


Fig. 6. MSE for different types of filters applied on malignant data of MIAS

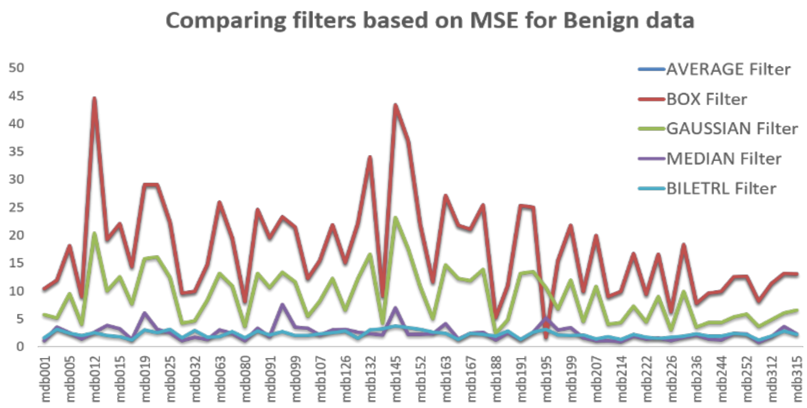


Fig. 7. MSE for different types of filters applied on benign data of MIAS

2) PSNR (Peak Signal to Noise Ratio):

It is the ratio between the maximum possible power of a signal and the power of the corrupting noise affecting the quality of the image. It is used as the quality measure between original and improved or filtered images. The high value of PSNR reflects the excellent quality of an image, and the low value reflects the lousy quality(see equation-8).

$$PSNR = 20(\frac{(MAX)}{\sqrt{MSE}}) \quad (8)$$

Where MAX is the maximum pixel value of the image means it's 255. For the 8-bit image, a PSNR value between 30 to 50 Db is considered a good quality image. While below 20 Db PSNR value images are unacceptable. As shown in figure 8 and figure 9, statistics after applying different filters can be seen. Here we can say that Bilateral Filters give the best result because, for PSNR, higher the result is comparatively better.

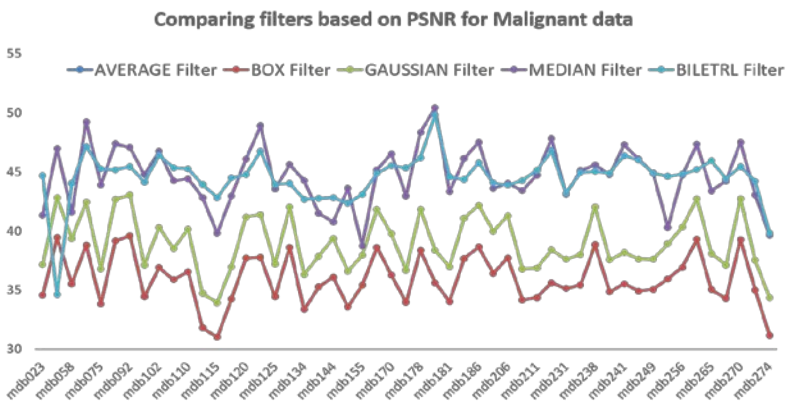


Fig. 8. PSNR for different types of filters applied on malignant data of MIAS

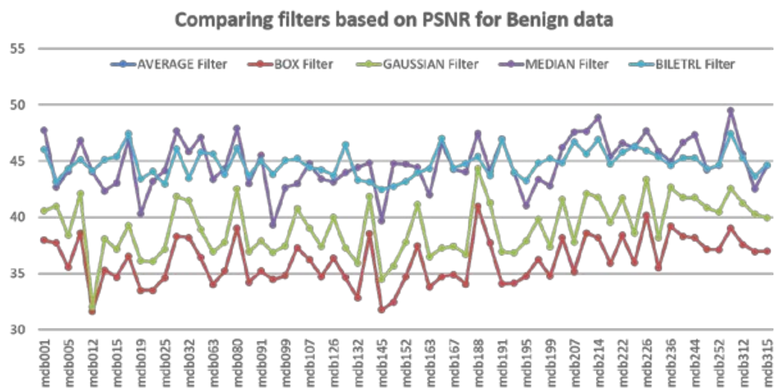


Fig. 9. PSNR for different types of filters applied on benign data of MIAS

3) SSIM (Structural Similarity Index Measure):

It is used to measure image quality which has been degraded due to processing. SSIM is calculated between original image and processed image(see equation-9). Output or Index value is in between -1 and 1. Where 1 indicates perfect structure similarity and 0 indicates no similarity between two images.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c1)(2\sigma_{xy} + c2)}{(\mu_{x^2} + \mu_{y^2} + c1)(\sigma_{x^2} + \sigma_{y^2} + c2)} \quad (9)$$

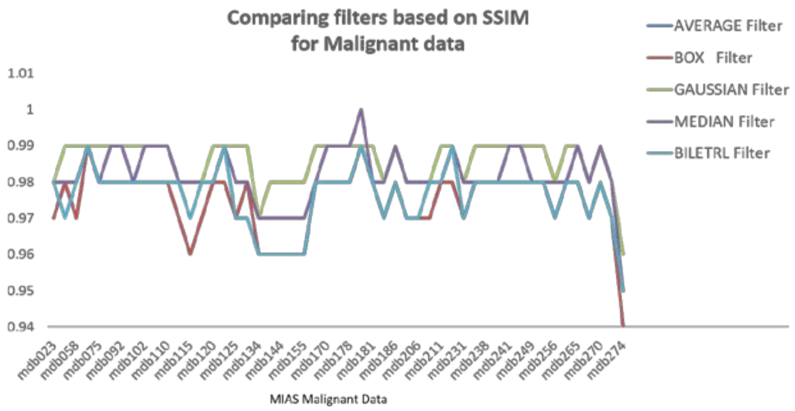


Fig. 10. SSIM for different types of filters applied on malignant data of MIAS

As shown in figure 10 and figure 11, statistics after applying different filters can be seen. Here we can say that Filters gives the best result because, for SSIM, values closer to 1 are comparatively better.

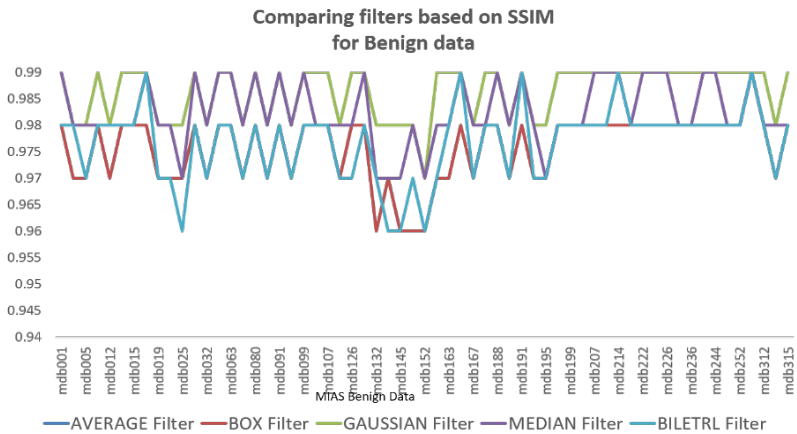


Fig. 11. SSIM for different types of filters applied on benign data of MIAS

The Paper discussed all the analyses done on the MIAS dataset using filters like Average filter, Box Filter, Gaussian Filter, Median Filter, and Bilateral Filter. As shown in table 3, we can say that the bilateral filter, Median filter, and Gaussian Filter give improved results for mammogram pre-processing compared to other filters.

Table 3. Filters having better results according to the measures used

MIAS Dataset	**Total Images**	**MSE**	**SSIM**	**PSNR**
Benign	61	Bilateral Filter	Median Filter	Bilateral Filter
Malignant	51	Bilateral Filter	Gaussian Filter	Bilateral Filter

As per the experimental results, we can say that applying different filters for noise removal can enhance mammograms. This is because different filters work differently per the image and noise.

5. Conclusion

Pre-processing of mammograms plays a crucial role in further diagnosis and extracting accurate results. Mammogram denoising and enhancement techniques improve the quality of an image and the perceptibility of abnormality in the image. In this paper, five different filters like Average, Box, Gaussian, Median, and Bilateral Filters are applied to the MIAS database. These filter results are compared using the performance parameters like MSE, PSNR, and SSIM. Data has been analyzed according to the benign or malignant data, and results show that different filters work better according to the data. From the experiments performed on MIAS, it is concluded that Median, Gaussian, and Bilateral filters work better than other filters. Different filters to the mammograms remove noise, and minor information or structures can be seen and analyzed. However, various filters give different results per the image content and noise. Therefore, pre-processing the mammogram is an important stage in Breast Cancer Diagnosis for more accurate results. Other imaging modalities can be used instead of mammograms for Breast cancer diagnosis using various filters and comparing the performance measures. In addition, other diseases like Corona, Brain tumors, Chest cancer, or disease where images are used for diagnosis can apply image enhancement techniques for improved results.

Conflict of Interest

The authors certify that they have no conflict of interest.

References

[1] Beeravolu, Abhijith Reddy, et al. "Preprocessing of breast cancer images to create datasets for deep-cnn." IEEE Access 9 (2021): 33438-33463.

[2] Breast Cancer who fact sheet (2021), https://www.who.int/news-room/fact-sheets/detail/breast-cancer, 26 March 2021

[3] Mahmood, Tariq, et al. "Breast lesions classifications of mammographic images using a deep convolutional neural network-based approach." Plos one 17.1 (2022): e0263126.

[4] Jalalian, Afsaneh, et al. "Foundation and methodologies in computer-aided diagnosis systems for breast cancer detection." EXCLI journal 16 (2017): 113.

[5] Breast Cancer who fact sheet, FDA-government, 2021, https://www.fda.gov/radiation-emitting-products/radiation-emitting-products-and-procedures/medical-imaging, 08/28/2018

[6] Bagchi, Susama, et al. "Image processing and machine learning techniques used in computer-aided detection system for mammogram screening-A review." International Journal of Electrical and Computer Engineering 10.3 (2020): 2336.

[7] Sarangi, Sunita, Nrusingha Prasad Rath, and Harish Kumar Sahoo. "A Comparative Study of Filters for Denoising Mammograms." 2021 International Conference on Intelligent Technologies (CONIT). IEEE, 2021.

[8] Swamy, Somashekhar, and P. K. Kulkarni. "A basic overview on image denoising techniques." Int. Res. J. Eng. Technol. 7.5 (2020): 850-857.

[9] Sarangi, Sunita, Nrusingha Prasad Rath, and Harish Kumar Sahoo. "A Comparative Study of Filters for Denoising Mammograms." 2021 International Conference on Intelligent Technologies (CONIT). IEEE, 2021.

[10] Joseph, Ardra Mariya, M. Grace John, and Anto Sahaya Dhas. "Mammogram image denoising filters: A comparative study." 2017 Conference on emerging devices and smart systems (ICEDSS). IEEE, 2017.

[11] Ayyala, Rama S., et al. "Digital mammographic artifacts on full-field systems: what are they and how do I fix them?." Radiographics 28.7 (2008): 1999-2008.

[12] Singh, Inderpreet, and Nirvair Neeru. "Performance comparison of various image denoising filters under spatial domain." International Journal of Computer Applications 96.19 (2014): 21-30.

[13] Ganvir, Neha N., and D. M. Yadav. "Filtering Method for Pre-processing Mammogram Images for Breast Cancer Detection." International Journal of Engineering and Advanced Technology. ISSN: 2249-8958.

[14] Mehmood, Mavra, et al. "Machine learning enabled early detection of breast cancer by structural analysis of mammograms." Comput. Mater. Contin 67 (2021): 641-657.

[15] Anand, S., and S. Gayathri. "Mammogram image enhancement by two-stage adaptive histogram equalization." Optik 126.21 (2015): 3150-3152.

[16] Justaniah, Eman, Areej Alhothali, and Ghadah Aldabbagh. "Mammogram Segmentation Techniques: A Review." International Journal of Advanced Computer Science and Applications 12.5 (2021).

[17] Punitha, S., A. Amuthan, and K. Suresh Joseph. "Benign and malignant breast cancer segmentation using optimized region growing technique." Future Computing and Informatics Journal 3.2 (2018): 348-358.

[18] Tripathy, Sushreeta, and Tripti Swarnkar. "Unified preprocessing and enhancement technique for mammogram images." Procedia Computer Science 167 (2020): 285-292.

[19] Christopher, Divya, and Philomina Simon. "A novel approach for mammogram enhancement using nonlinear Unsharp masking and L0 gradient minimization." Procedia Computer Science 171 (2020): 1848-1857.

[20] Christopher, Divya, and Philomina Simon. "A novel approach for mammogram enhancement using nonlinear Unsharp masking and L0 gradient minimization." Procedia Computer Science 171 (2020): 1848-1857.

[21] Sanchez-Montero, Rocio, et al. "Filtering of mammograms based on convolution with directional fractal masks to enhance microcalcifications." Applied Sciences 9.6 (2019): 1194.

[22] MehmoodGondal, Rashid, et al. "A hybrid de-noising method for mammogram images." Indonesian Journal of Electrical Engineering and Computer Science 21.3 (2021): 1435-1443.

[23] Oza, Parita, et al. "A bottom-up review of image analysis methods for suspicious region detection in mammograms." Journal of Imaging 7.9 (2021): 190.
[24] Jiménez, Yuliana, et al. "Breast Mass Regions Preprocessing and Classification from Mammograms using Convolutional Neural Networks."
[25] Rani, Neha, and Deepak Kumar Gupta. "Performance Based Analysis of Classification Techniques in Breast Cancer Screening–A Review." International Journal of Research in Engineering, Science and Management 5.2 (2022): 130-135.
[26] Dua, Dheeru, and Casey Graff. "UCI machine learning repository." (2017).
[27] Spanhol, Fabio A., et al. "A dataset for breast cancer histopathological image classification." Ieee transactions on biomedical engineering 63.7 (2015): 1455-1462.
[28] Balafar, Mohammad. (2012). Review of Noise Reducing Algorithms for Brain MRI Images. ijtpe. 4. 54-59.
[29] Gökçen Çetinel, Sevda Gül, " Breast Lesion Segmentation and Area Calculation for MR Images", International Journal of Image, Graphics and Signal Processing, Vol.10, No.10, pp. 1-10, 2018.
[30] Samee, N. A., Alhussan, A. A., Ghoneim, V. F., Atteia, G., Alkanhel, R., Al-Antari, M. A., & Kadah, Y. M. (2022). A hybrid deep transfer learning of CNN-based LR-PCA for breast lesion diagnosis via medical breast mammograms. Sensors, 22(13) doi:10.3390/s22134938
[31] Hossain, M. S. (2022). Microc alcification segmentation using modified U-net segmentation network from mammogram images. Journal of King Saud University - Computer and Information Sciences, 34(2), 86-94. doi: 10.1016/j.jksuci.2019.10.014.
[32] Nazir, M. S., Khan, U. G., Mohiyuddin, A., Al Reshan, M. S., Shaikh, A., Rizwan, M., & Davidekova, M. (2022). A novel CNN-inception-V4-based hybrid approach for classification of breast cancer in mammogram images. Wireless Communications and Mobile Computing, 2022 doi:10.1155/2022/5089078
[33] Nagalakshmi, K., & Suriya, S. (2022). Performance analysis of breast cancer detection method using ANFIS classification approach. Computer Systems Science and Engineering, 44(1), 501-517. doi:10.32604/csse.2023.022687
[34] Dhar, Prashengit. "A Method to Detect Breast Cancer Based on Morphological Operation." International Journal of Education and Management Engineering, 11.2 (2021): 25-31.
[35] Kayode, Aderonke Anthonia. "Preparing Mammograms for Classification Task: Processing and Analysis of Mammograms." (2016).
[36] Fajila, M. N. F., and M. A. C. Jahan. "The effect of evolutionary algorithm in Gene subset selection for cancer classification." (2018).
[37] Sharma, Tejpal. "An efficient adaptive based median technique to de-noise colour and greyscale images". International Journal of Modern Education and Computer Science, 11.2 (2018): 48.
[38] Sharma, Tejpal. "Various Types of Image Noise and De-noising Algorithm", International Journal of Modern Education and Computer Science, 9.5 (2017).
[39] Maier, Andreas, et al., eds. "Medical imaging systems: An introductory guide", (2018).
[40] Gonzalez, Rafael C. Digital image processing. Pearson education India, 2009.

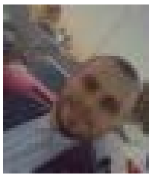
Authors' Profiles

Shah Hemali is working as an Assistant Professor at Computer Science Department since 2011.She has a teaching experience of over 10 years. She has also worked in BISAG (Bhaskaracharya Space Research & Geo Informatics) as Project Scientist for two years. She is currently pursuing PhD. From Nirma University. She has received M.tech degree in 2014.She worked in the area of Artificial Intelligence, Machine Learning and Big data. She has given one week training to the Indian Air force on GIS. She is involved in teaching courses at both under graduate and graduate level.

Dr Smita Agrawal is working as an Assistant Professor in Computer Science and Engineering Department since 2009. She has a teaching experience of over 14 years. She received her Master of Computer Applications degree from Gujarat Vidhyapith in 2004. She completed her doctoral studies in the field of Big Data Analytics from CHARUSAT University in 2019. She works in the area of Big Data Analytics, Parallel Processing, Web Development and IoT. She is a Senior Member of IEEE, Member of CSI, ACM, and ISTE. She has published several Scopus/SCIE-indexed research papers in national and international conferences and journals. Dr Smita has conducted ISTE approved short-term training Programme in the field of Web Services using PHP. She is involved in teaching courses at both undergraduate and postgraduate levels. She also serves as a member of the Programme committees and session chair for international conferences and as a reviewer for indexed international journals.

Prof. Parita Oza is working as an Assistant Professor in the Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad. She has pursued her MTech in Information and Communication Technology from Nirma University. She is involved in teaching courses at both undergraduate and postgraduate levels. She has published several research papers in national and international conferences and journals. She is pursuing her PhD. from Pandit Deendayal Energy University (Gandhinagar, India). Her research area includes Image Processing, Computer Vision and Medical Imaging. She has mentored many BTech and MTech projects. She also serves as a member of the programme committees and session chair for international conferences and as a reviewer for indexed international journals.

Dr Sudeep Tanwar is working as a Professor in Computer Science and Engineering Department. He has more than 19 years of teaching experience. Dr Tanwar received his BTech degree in Computer Engineering from Kurukshetra University, Kurukshetra, MTech (Hons) degree in Information Technology from the Guru Gobind Singh Indraprastha University, Delhi, and PhD degree in Computer Science and Engineering from Mewar University, India. Dr Tanwar's research interests include Blockchain Technology, Wireless Sensor Networks, Fog Computing, Smart Grid, and IoT. He has authored 02 books and edited 13 books, more than 200 technical papers, including top journals and top conferences, such as IEEE TNSE, TVT, TII, WCM, Networks, ICC, GLOBECOM, and INFOCOM. He has guided many students leading to ME/MTech and guiding students leading to PhD. He has been awarded as Outstanding Reviewer Award from FGCS, Elsevier, CEE, Elsevier, JNCA, Elsevier, DCN, Elsevier. (July 2018). Dr Tanwar initiated the research field of blockchain technology adoption in various verticals in 2017. His h-index is 38. Dr Tanwar actively serves his research communities in various roles. He is currently serving the editorial boards of Physical Communication, Computer Communications, International Journal of Communication System, and Security and Privacy. He has been awarded the best research paper awards from IEEE GLOBECOM 2018, IEEE ICC 2019, and Springer ICRIC-2019. He has served many international conferences as a member of the organizing committee, such as publication chair for FTNCT-2020, ICCIC 2020, WiMob2019, member of the advisory board for ICACCT-2021, ICACI 2020, workshop co-chair for CIS 2021, and general chair for IC4S 2019, 2020, ICCSDF 2020. Dr Tanwar is a final voting member for IEEE ComSoc Tactile Internet Committee in 2020. He is a Senior Member of IEEE, CSI, IAENG, ISTE, CSTA, and the member of the Technical Committee on Tactile Internet of IEEE Communication Society.



Ahmed Alkhayyat is associated with the Department of Computer Technical Engineering, College of Technical Engineering, Islamic University, 2976+72M, Najaf, Iraq. Dr. Ahmed Alkhayyat received the B.Sc. degree in electrical engineering from AL KUFA University, Najaf, Iraq, in 2007; the M.Sc. degree from the Dehradun Institute of Technology, Dehradun, India, in 2010; and Ph.D. in electronics and communication from Cankaya university, Ankara, Turkey, in 2015. He contributed in organizing several IEEE conferences, workshop, and special sessions. He is currently a dean of international relationship and manager of the world ranking in the Islamic University, Najaf, Iraq. To serve my community, I acted as a reviewer for several journals and conferences. His research interests include network coding, cognitive radio, efficient-energy routing algorithms and efficient-energy MAC protocol in cooperative wireless networks and wireless local area networks, as well as cross-layer designing for a self-organized network.

How to cite this paper: Shah Hemali, Agrawal Smita, Parita Oza, Sudeep Tanwar, Ahmed Alkhayyat, "Mammogram Pre-processing Using filtering methods for Breast Cancer Diagnosis", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.15, No.4, pp. 44-58, 2023. DOI:10.5815/ijigsp.2023.04.04