An Efficient Preprocessing Technique for Multimodality Breast Cancer Images

https://doi.org/10.3991/ijoe.v19i08.40043

Anupama Y.K.(✉), Amutha S., Ramesh Babu D.R.
Dayananda Sagar College of Engineering, Affiliated to VTU University, Bengaluru, Karnataka, India
anupama-cs@dayanandasagar.edu

Abstract—On average, one in every eight women is diagnosed with breast cancer during their lifetime, and accounts for 14% of cancers in women. Since early diagnosis could improve treatment outcomes and longer survival times for patients, it is absolutely necessary to develop techniques to classify lesions within breast cancer mammograms and ultrasound images. The main goal is to determine the class of tumor present within the image, which is pivotal in diagnosing breast cancer patients. In this paper, we propose an Sobel-Canny-Gabor (SCG) model, which is a hybrid model that implements three different edge detection filters; Sobel filter, Gabor filter, and Canny filter. This model is used to enhance the appearance of the mammogram and ultrasound images, which is then fed into a classification model. Through classification, there could be a potential improvement in the results of the overall classification. Post-classification, the model is then evaluated using the metric Peak Signal-to-Noise Ratio (PSNR), which measures the quality between the original image and the compressed image.

Keywords—SCG, Sobel, Canny, Gabor, PSNR

1 Introduction

Breast cancer is one of the prime reasons for women's death in the world. In India most common type of cancer affecting women is breast cancer. Account of 14% of breast cancer contributing to the mortality rate of women. Main motivation for this studies to early diagnosis, so that causalities can be reduced by helping radiologists to diagnosis effectively. Many modalities can be used to detect cancer such as mammograms, ultrasound, tomography, CT scan, etc…

Mammogram and ultrasound modalities are considered for study. Early mammographic and ultrasonic detection based on computer-aided detection (CAD) methods can improve treatment outcomes and yield longer survival times for the patients. Modalities such as mammograms and ultrasound help in determining the presence of tumors in breast masses. In all these modalities there might be different noises such as impulse noise, random noise, etc. So, it is required to preprocess these image modalities to remove the impurities present in them. Later on, obtained preprocessed images will be used for classifying. This in turn improves the diagnosis accuracy.

2 Related work

Yousif M.Y Abdallah et al., [1] proposed a methodology to improve the breast lesions diagnosis. Proposed methodology uses the wiener function to remove insignificant information from the image and the k-means algorithm has been used as a segmentation technique. Results were obtained, with an accuracy of 96.3 %. An extension of the work by using this methodology with different modalities, was proposed.

Dr. D. Devakumari and V. Punithavathi [2], explored existing processing methods of images to propose the best method for identifying breast cancer. In this study, the Support Vector machine (SVM) technique was proposed as the best method.

Tobias Chrisiian Cahoon et al. [3], proposed a technique based on the means of a window and standard deviation to reduce the incorrectly identified pixels in a breast image. This method is significantly better than using intensity as a feature for lessening the incorrectly identified pixels. A future extension involved including ratings from clinicians to evaluate the model outputs.

Farahnaz Sadoughi et al. [4], did a review on diagnosing breast abnormalities using different AI techniques. Accordingly, SVM provided best accuracy for mammography and ultrasound images. Future enhancement suggested reducing false positives, thereby increasing accuracy.

R. Ramani et al. [5], explored different preprocessing methods for mammography images. Various filters like average, adaptive median and wiener are used to remove the irrelevant information from the images. This study concluded that the Adaptive median filter is best suitable for removing irrelevant information from images of breasts.

Sushreetha Tripathy and Tripti Swarnkar [6], proposed a Contrast limited adaptive histogram equalization (CLAHE) to identify the edges of abnormalities of breast. Compared to existing methods, the proposed method provided better results in terms of Contrast improvement index (CII).

Ardalan Ghasemzadeh et al. [7], explored a methodology involving Gabor wavelet transform to get the feature vector of a mammography image. Then applied a classification technique which obtained an accuracy of 93.9%.

Vishnukumar K. Patel et al. [8] proposed a methodology to use image enhancement and sharpening processes on mammographic images. Resulting images are free from noise as a result of which the contrast is enhanced. This helped radiologists to classify in a better way compared to existing technologies. This methodology showed an improvement in signal to noise ratio.

Guangxing Guo & Navid Razmjooy [9] investigated interval analysis of Sobel filters to overcome intensity uncertainties. It's implemented on MIAS. The results of the proposed method were better compared to LoG, Prewitt and canny filters. Finale results showed that using the proposed method gives better achievement than the others by considering some kinds of uncertainties like Gaussian noise and salt and pepper noise.

In this paper, image preprocessing techniques for mammogram and ultrasound images are carried out. The rest of the paper is organized as follows: section two is about the material and method, section three is about the results and discussion, section four is about the conclusion.

3 Materials and methods

In this section materials such as data sets used for experiment are detailed. Then existing and proposed methodologies are discussed in detail.

3.1 Mammogram-data set

The dataset primarily used for the development of the proposed preprocessing technique is the Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM). The CBIS-DDSM collection includes a subset of the DDSM [10] data selected and curated by a trained mammographer. The images have been decompressed and converted to DICOM image format. The dataset is comprised of BUS images, which are converted from the dicom format to the png format before further processing.

Size of the dataset: 152 GB (Images in DICOM format)
Total number of images: 9979
Classes present in the dataset: Benign, Malignant, Benign without callback
Distribution of images in the above classes is as follows: Benign: 4006 images, Malignant: 4236 images, Benign without callback: 1737 images.
Details of the image after conversion to the PNG format: Dimensions: 3024×4808 431×515 681×549 145×113; Bit depth: 8

3.2 Ultrasound-data set

The ultrasound images [11] that have been taken into trial are from the Biomedical Engineering Unit of Sirindhorn International Institute of Technology, which has presented a database of ultrasound images of breast cancer. These have been provided by the Department of Radiology of Thammasat University and Queen Sirikit Center of Breast Cancer of Thailand. The database includes the various classes of ultrasound, Doppler and elasticity images along with the ground truth hand-drawn by leading radiologists of these centers.

Total number of ultrasound images which have been considered for the experiment includes around 296 images.

3.3 Existing methods

Image preprocessing involves cleaning the given image before feeding the image to any of the machine learning algorithm. Many image preprocessing techniques are available such as image resizing, thresholding, edge detection, morphology, etc. In this study mainly focused on edge detection techniques. Edge detection techniques can be of linear and non-linear filters. For detecting edges in an image, non-linear edge detector filters such as Robert, Prewitt and Laplacian are used.

Robert. The Robert filter is used to figure out the two dimensional slope of an image. High pitch rates with respect to edges are easily identified by this method. It takes a gray scale image and gives an output that consists of approximated absolute values of the slope of the input image at that point.

Figure 1 shows two 2×2 filters involved in operator used in Robert. One filter is counter part of the other filter.

+1	0
0	–1

0	+1
–1	0

Fig. 1. Filters

Advantage of the technique, is easy to compute operator. It only needs to input 4 pixels to know individual output pixel. Disadvantage of the technique, is sensitive to noise because of the small kernel size.

Prewitt. The Prewitt filter is used to compute an estimation of the slope of the image intensity function. Each pixel point of an image is obtained after applying the operator of prewitt, usually with respect to slope of the vector or derivation of the vector. It involves convolving an image in vertical and horizontal lines.

Prewitt mainly helps to find edges in vertical and horizontal direction respectively. Masks used for this as in Figure 2. Disadvantage of using the prewitt operator is mainly it is sensitive to noise and inaccurate.

–1	–1	–1
0	0	0
1	1	1

–1	0	1
–1	0	1
–1	0	1

Fig. 2. Vertical and horizontal masks

Laplacian. The Laplacian filter uses a second order derivative to find out edges. In other edge detection methods such as Robert, prewitt, etc…, derivatives of first order are used. Mammogram/ultrasound images were taken as input to non-linear edge detection methods in the existing method. Other techniques for detecting margins most of the time try to find the edges horizontal or in vertical direction or both. Laplacian detects outward or inward margins or both. Negative laplacian has a median peak of negative value looks as in Figure 3. Disadvantage of this technique, is incapable of recognizing the discontinuous or incomplete edges.

0	–1	0
–1	4	–1
0	–1	0

Fig. 3. Negative laplacian kernel

3.4 Proposed methodology

Proposed method involves comparison of existing edge-detection techniques with the proposed solution as in Figure 4.

Existing methodology. Existing methodology algorithm involves following steps:

Input: Mammogram/Ultrasound Images
Output: Preprocessed Images
Procedure:
Step1: Ultrasound/mammogram images are processed using Robert method

Step2: Ultrasound/mammogram images are processed using prewitt method

Step3: Ultrasound/mammogram images are processed using Laplacian method

Step4: Preprocessed images obtained from Robert, Prewitt and Laplacian are analyzed using Peak to signal Noise Ratio (PSNR) metric.

Proposed methodology. Proposed methodology Algorithm involves following steps:

Input: Mammogram/Ultrasound Images

Output: Preprocessed Images

Procedure:

Step1: Histogram Equalization method is applied on Mammogram/ultrasound images, which magnifies the contrast of the image to improve the appearance of the image. After this, it is processed using Median Blur to eliminate random noise like pepper noise in the image.

Step2: Enhanced images are then taken as input to an SCG hybrid technique. (Sobel method used to detect high pitch frequency to find out the edges. Canny method is used to mine significant fundamental evidence from images to identify a wide range of edges. The Gabor method used to extract features consisting of high pitch frequency constituents.) The Hybrid technique reduces the possibility of the final image consisting of discontinuous or incomplete edges and helps in highlighting significant information.

Step3: Preprocessed images obtained from are analyzed and evaluated using Peak to signal Noise Ratio (PSNR) metric

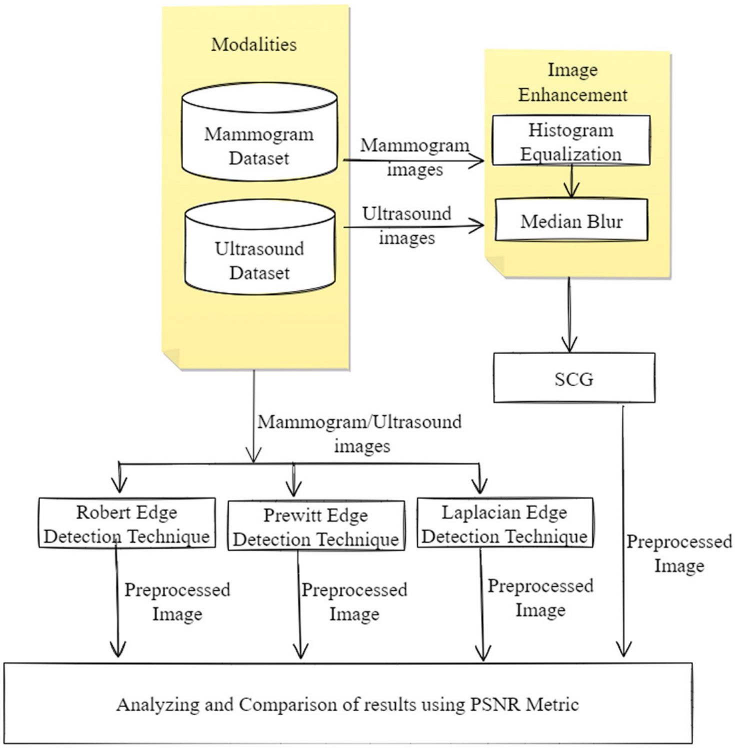

Fig. 4. Multimodality images preprocessing block diagram

4 Experiment results

PSNR is a mathematical term that represents the quality impacts of an image. Fundamentally, it is the proportion between the peak possible rate of signal and the value of altering misleads.

$$PSNR = 10log_{10}\frac{Peak}{Sqrt(M)} \tag{1}$$

Here, *Peak* is the peak possible rate of the image. *Sqrt*(*M*) is the square root of mean square error.

4.1 Mammogram experiment results

PSNR values of three mammogram images for SCG, Robert, Prewitt and Laplacian techniques are recorded as in Table 1. As per tabulation of results SCG outperforming in terms preprocessing compared to robert, prewitt and laplacian edge detection techniques.

Table 1. Mammogram PSNR comparison

Class	SCG	Robert	Prewitt	Laplacian
Benign	30.21	7.35	7.59	7.25
Benign_Without Callback	31.41	11.08	11.65	10.74
Malignant	30.84	8.23	8.66	7.98

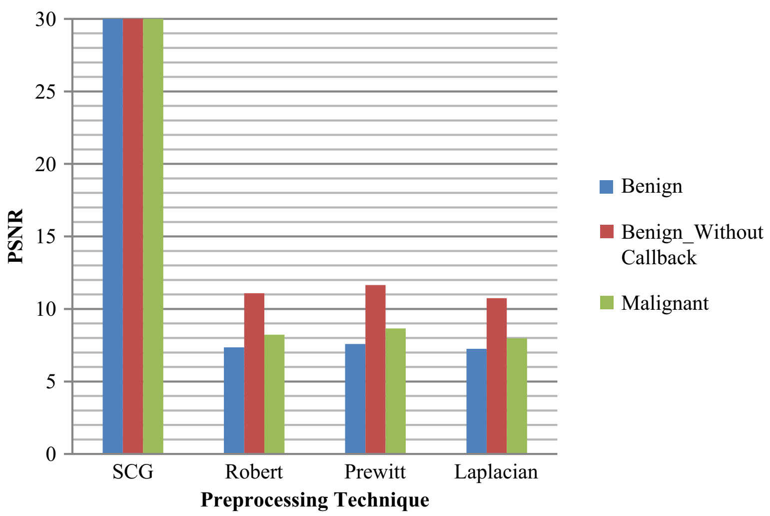

Fig. 5. Mammoram PSNR metric results

4.2 Ultrasound experiment results

PSNR values of three ultrasound images SCG, Robert, Prewitt and Laplacian techniques are recorded as in Table 2. As per tabulation of results SCG outperforming in terms preprocessing compared to robert, prewitt and laplacian edge detection techniques.

Table 2. Ultrasound PSNR comparison

Class	SCG	Robert	Prewitt	Laplacian
Benign	28.29	9.90	11.64	9.16
Malignant	27.93	8.59	11.46	7.81

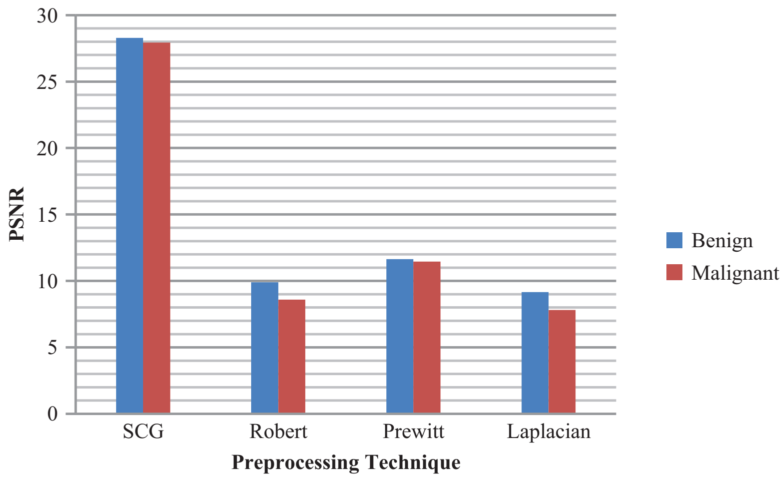

Fig. 6. Ultrasound PSNR metric results

5 Discussion

PSNR values of mammogram image labels benign, benign without callback and malignant are recorded as in Figure 5. As per graphs results SCG outperforming in terms PSNR metric compared to robert, prewitt and laplacian edge detection techniques.

PSNR values of ultrasound image labels benign and malignant are recorded as in Figure 6. As per graphs results SCG outperforming in terms PSNR metric compared to robert, prewitt and laplacian edge detection techniques.

SCG method performance indicates that, it is helping to remove the noise, capable of identifying discontinued edges or curves and significant edges. Quality of the image also improved. Preprocessed images obtained from this can be taken as input for any of the machine learning model. So, the performance of the model can be increased.

6 Conclusion

Proposed methodology SCG has better PSNR values compared to existing Robert, Prewitt and Laplacian methods. Using the proposed SCG method as an image preprocessing technique for multimodalities such as mammogram and ultrasound, improves the appearance of the images by denoising them effectively. So, SCG is a better preprocessing technique with respect to other techniques comparatively. As an extension for this work, SCG can be used with other modalities such as tomography, CT scan, etc.

7 References

[1] Yousif M.Y. Abdallah, Sami Elgak, Hosam Zain, Mohammed Rafiq, Elabbas A. Ebaid and Alaeldein A. Elnaema (2018), Breast cancer detection using image enhancement and segmentation algorithms. Biomedical Research, 29(20): 3732–3736, ISSN 0970-938X. https://doi.org/10.4066/biomedicalresearch.29-18-1106

[2] Dr. Devakumari, D. and Punithavathi, V. (2018), Study of breast cancer detection methods using image processing with data mining techniques. International Journal of Pure and Applied Mathematics, 118(18): 2867–2873, ISSN: 1311–8080 (printed version); ISSN: 1314–3395 (on-line version). http://www.ijpam.eu

[3] Tobias Chrisiian Cahoon, Melanie A. Suttor and James e. Bezdek (2000), Breast cancer detection using image processing techniques, 0-7803-5877-5/001 Q 2000 IEEE, 973–976.

[4] Farahnaz Sadoughi, Zahra Kazemy, Farahnaz Hamedan, Leila Owji, Meysam Rahmanikatigari and Tahere Talebi Azadboni (2018), Artificial intelligence methods for the diagnosis of breast cancer by image processing: A review. Breast Cancer – Targets and Therapy, 10: 219–230. https://doi.org/10.2147/BCTT.S175311

[5] Ramani, R., Dr. Suthanthira Vanitha, N. and Valarmathy, S. (2013), The Pre-Processing Techniques for Breast Cancer Detection in Mammography Images. International Journal of Image Graphics and Signal Processing, 5: 47–54, https://doi.org/10.5815/ijigsp.2013.05.06

[6] Sushreetha Tripathy and Tripti Swarnkar (2020), Unified preprocessing and enhancement technique for mammogram images. International Conference on Computational Intelligence and Data Science (ICCIDS 2019), Procedia Computer Science, 167: 285–292. https://doi.org/10.1016/j.procs.2020.03.223

[7] Ardalan Ghasemzadeh, Saeed Sarbazi Azad and Elham Esmaeili (2019), Breast cancer detection based on Gabor-wavelet transform and machine learning methods. International Journal of Machine Learning and Cybernetics, 10: 1603–1612. https://doi.org/10.1007/s13042-018-0837-2

[8] Vishnukumar K. Patel, Syed Uvaid and Suthar, A. C. (2012), Mammogram of breast cancer detection based using image enhancement algorithm. International Journal of Emerging Technology and Advanced Engineering, ISSN 2250–2459, 2(8).

[9] Guangxing Guo and Navid Razmjooy (2019), A new interval differential equation for edge detection and determining breast cancer regions in mammography images. Systems Science & Control Engineering: An Open Access Journal, 7(1): 346–356. https://doi.org/10.1080/21642583.2019.1681033

[10] CBIS-DDSM – The Cancer Imaging Archive (TCIA) Public Access – Cancer Imaging Archive Wiki.

[11] Rodtook, A., Kirimasthong, K., Lohitvisate, W. and Makhanov, S. S. (2018), "Automatic initialization of active contours and level set method in ultrasound images of breast abnormalities". Pattern Recognition, 79: 172–182. https://doi.org/10.1016/j.patcog.2018.01.032

8 Authors

Mrs. Anupama Y.K., Assistant Professor of Department of Computer Science and Engineering, Dayananda Sagar College of Engineering, Bengaluru. She obtained her M.Tech and B.E. degree from VTU University. She has published in 9 journals and conferences. Her research interests include data mining, digital image processing and computer vision (e-mail: anupama-cs@dayanandasagar.edu).

Dr. Amutha S., Professor of Department of Computer Science and Engineering, Dayananda Sagar College of Engineering, Bengaluru. She obtained her Ph.D. degree from VTU University. She has more than 24 years of teaching experience. She has published in 29 journals and conferences. Her research interests include Big Data, data mining, and cloud computing (e-mail: amutha-cs@dayanandasagar.edu).

Dr. Ramesh Babu D.R., HOD & Professor of Department of Computer Science and Engineering, Dayananda Sagar College of Engineering, Bengaluru. He obtained his Ph.D. degree from Mysore University. He has more than 20 years of teaching experience. He has published in 28 journals and 20 conferences. His research interests are data mining, digital image processing and computer vision (e-mail: bobrammysore@gmail.com).

Article submitted 2023-02-31. Resubmitted 2023-04-10. Final acceptance 2023-04-14. Final version published as submitted by the authors.