Hybrid Data Mining Method For Early Detections Of Breast Cancer

M.S IRFAN AHMED¹, A.KRISHNAVENI²

Professor¹, Assistant professor²

¹,²Department of Computer Science Thassim Beevi Abdul Kader College for Women Kilakarai-623517, Ramanathapuram.

Abstract

Early detection of breast cancer is a very challenging task and difficulty faced by pathologists and researchers. High-quality breast images are needed for early detection of breast cancer without any difficulties required for an optimal image enhancement technique. In this paper, a new hybrid method is based on gamma correction (power-law) and morphological operations (erosion and dilation) are used. This method enhances the quality of an input image and applied two powerful techniques for the diagnosis of breast tumors with the help of the Region of Interest (ROI) method. Image quality metrics Peak Signal to Noise Ratio (PSNR), Root Mean Squared Error (RMSE), Standard Deviation (SD), and Contrast Improvement Index (CII) values are used for evaluation. The outcome of the image enhancement using the hybrid method indicates low RMSE, high PSNR, and SD values. This proves the efficiency of the hybrid method in the early detection of breast cancer.

Keywords: Gamma correction, Morphology, Mammogram, Hybrid, Optimization.

1. Introduction

Breast Tumor [BT] is a mixed set of diseases, which diverge from each other in the ordinary record, morphology, molecular phenotype, medical and radiological manifestations, and diagnosis [1]. Computer Vision and Artificial Intelligence are the diagnostic methodologies in medical image processing to simplify the stages of critical disease like breast cancer diagnosis [2]. Current
research statistics show that artificial intelligence is considered the main field and is frequently used by all researchers. However, accurate prediction of a tumor in the breast, brain, lung, and all other diseases is very challenging to all researchers [3]. In such cases, machine-learning algorithms Convolutional Neural Network (CNN) and Deep Belief Neural network (DBN) are used to improve the accuracy of cancer predicted results using Computer-Aided Diagnosis (CAD) systems, image enhancement methods, awareness programs, digital pathology, doctor’s guidelines, and digital mammograms [4-10].

The most predominant cancer in the world is breast cancer due to the high mortality rate as well as many women affected with breast cancer in a long period. Many of the women and men in the world are pretentious by breast cancer every year. The American Cancer Society's assessments for breast disease in the United States for 2021 are: the possibility that a lady will pass on from breast malignant growth is around 1 out of 39 (around 2.6%) and 43,600 ladies will bite the dust from the breast disease [11]. Generally, all the captured images have some noise. In medical images like breast images, have masses. To reduce the masses in mammogram images, enhancement techniques can be involved [12]. Hue, saturation, intensity, contrast adjustment, tuning the brightness are some of the methods in image enhancement [13]. The purposes of image improvement are listed below:

1) Poor contrast of an image can be improved using contrast enhancement methods such as histogram equalization that improves the quality of an image.

2) To highlight the important details/interesting details in an image.

3) To remove the noise from an image and make the image look more appealing.

Image processing methods whose inputs are images, the outputs are attributes extracted from those original images. Segmentation is another major step in this direction. The low contrast of an image affects segmentation processing and image analysis. Improving the contrast of an image is necessary to process in all real-time applications. Ganesh et al. (2010) introduced various techniques for mammogram image segmentation [22]. N.E.A.Khalid et al., (2011) reviewed many biologically inspired algorithms and their purposes in image processing applications. Medical image processing frequently used bio algorithms for finding diseases. Some of the bio-inspired algorithms but are not limited to genetic, memetic, lion, lizard, elephant, monkey, strawberry, duck algorithms. The study of both human beings and animal behavior is observed deeply [32]. Hajar et al., (2012) examined a comparative study about mammogram image enhancement using wavelet transform and morphological operations [23]. H.H.Aghdam et al., (2013) recommended a probabilistic method for boundary extraction using contour methods and thresholding methods. Image smoothening is done by using the active contour with the proposed method of 86% stability [35]. W.U. Shibin et al., (2013) applied a new procedure for improving the feature and contrast using laplacian filter and Contrast limited adaptive histogram equalization methods. Evaluation is

Das T.K and Chowdhary (2018) described the advantage of using median filtering for basic morphological operations and their needs in mammogram image processing like segmentation and classification. Binarization is the main process focused on using erosion and dilation. Region filling is a required method clearing the limitations in morphological methods [31]. J.Q.Dominguez et al., (2018) detected micro calcifications in breast images using morphological operations and k-means clustering methods. Accuracy and specificity measures are calculated and given in results [34].

W.Lestari and S.Sumarlinda (2019) discussed mathematical morphology for medical image field processing. Four operations such as opening, closing, dilation, erosion enhance the clarity of an image was described. The authors wish to further develop the morphological operations to enhance the quality of original mammograms [33]. Hyo-Eun Kim et al., (2020) used an artificial intelligence algorithm for mammogram image enhancement [36]. The traditional method for image enhancement technique is Histogram Equalization (HE), which uses the probability density function for picture elements. Gang Cao et al., (2021) planned the method of adaptive gamma correction method for contrast enhancement used in pattern recognition and other useful applications [14]. Dmitrii Bychkov et al., (2021) projected morphological operations using deep learning methods [17]. Prashengit (2021) planned for morphological operations based on mathematical morphology using gamma corrections for breast tumor detection [19].

Zeynab et al (2021) presented the FCM-GA algorithm for breast cancer detection [24]. R.D.Gilardi et al., (2021) utilized computer networks for mammogram subtypes and predict the cc that is cohesion and chromosome analysis. Healthy and unhealthy cells are classified in the group of inter classification of chromosomes and adjustable chromosomes. Gene expressions are observed very carefully in this network [27].J.Sun et al., (2021) analyzed the significance of (RAC 1) in mammogram tumor analysis. Alpha function in estrogen both positive and negative methods are very useful to find out the cancer data [28]. Pengfei Xu et al., (2021) described the significance of programmed cell death ligand 1 which is known as (PDL1) on immune cells for observing the activity of cells. The triple-negative breast cancer-detecting method is very useful to detect the breast tumor in an effective manner [29]. Lin Li et al., (2021) studied and signified the ectodermal-neural cortex 1, which is known as EC 1for breast cancer radiation analysis. This analysis was very helpful to found mRNA outstandingly through the breast cells functioning [30].
2. Materials and Methods

The intensity transformation function (gray level transformation) using power law is known as gamma correction [38].

\[ S = T(r) \quad (1.1) \]

- \( r \) → value of pixel before processing
- \( S \) → value of pixel after processing
- \( T \) → Transformation function

Here \( r \) is the pixel value of an input image and \( S \) is the pixel value of an output image. \( T \) is the transformation function that maps the value of \( r \) with the value of \( s \). Since the user is dealing with digital images the value of the transformation function is stored in a one-dimensional array, if the user says one-dimensional array, which is nothing but in matrix form. Therefore, the values of transformation functions are stored in matrix form. The mapping of value \( r \) with value \( S \) is implemented through a lookup table. If we consider an 8-bit environment, the lookup table is consists of 256 entities. Because \( 2^8=256 \) in that case. Users can have the transformation function from the value 0 to 255.

![Transformation functions using input and output intensity level](http://www.webology.org)

**Figure 1. Transformation functions using input and output intensity level**
In this graph x-axis, users are taking r which is the input pixel values, and the y-axis, users are taking S which is the output pixel values. Thereafter, the user represents that as intensity (grey) levels of input pixel and output pixel. For example, take the total grey level as $L$, so the user represents that $0$ to $L - 1.0$ denotes the black and $L - 1.0$ exemplifies the highest intensity level value white. Figure 1 shows that three basic transformation function that is frequently used for image enhancement are:

a) **Linear Transformation** function consists of image Negative and Identity. Image identity in linear transformation functions a very less helpful in digital image processing because the output image will be the same as the input image. In image negative, the intensity level is taken from $0$ to $L - 1.0$. The image negative can be represented by the function [38].

$$S = L - 1.0 - r \quad (1.2)$$

In image negative, the pixel value will be reversed to produce an equivalent photographic negative

b) **Logarithmic transformation** function consists of Log and Inverse Log. Logarithmic transformation is also known as log transformation. It can be represented as in equation 1.3 [38].

$$S = C \log (1 + r) \quad (1.3)$$

$C \to$ Constant and $r > 0$

With the help of logarithmic transformation and power-law transformation, images can compress the gray level or expand the grey level. The logarithmic transformation will produce high contrast image and the inverse log will provide the low contrast image.

c) **Power-law Transformation** function consists of $n^{th}$power and $n^{th}$ root. It can be represented as in equation 1.4 [38].

$$S = C r^{Y} \quad (1.4)$$

$Y > 1$ denotes $n^{th}$ power and $Y < 1$ denotes $n^{th}$ the root. The power law is also known as gamma correction (nonlinear transformation) since here using constant gamma $Y$ value and this transformation is similar to the logarithmic transformation but different values of gamma. Input has a low pixel value but output has more contrast value in nth power gamma value at different values of gammas like $Y=0.67$ and $Y=0.20$.

d) **Bi-histogram equalization**
Grayscale image is also known as black and white image and each pixel has only one value which is the intensity of the pixel value. The value may range from 0 to 255. Histogram equalization is used to be image enhancement. It is a graphical representation of any data related to the digital image, which represents the frequency of occurrence of various grey levels. Input mammogram images are taken from the MIAS database. All mammogram images in MIAS contain three classifications such as normal, benign, and malignant. Mammogram image names are like mdb262, mdb003, mdb261, mdb089, mdb048, mdb083, mdb156, mdb164, mdb217.
Figure 2. System architecture for an image enhancement using hybrid method
In Figure 2, System architecture for an image enhancement method is used to diagnose the tumor image. Gamma correction matters in the event that you have any interest in showing a picture precisely on a PC screen. Gamma remedy controls the general splendor of a picture. Gamma correction and morphological operations are used to increase the quality of the given input image. Desired tumor regions were also segmented using the Region of Interest (ROI) method.

**Morphological operations**

The study of objects in an image is known as morphological operations. Dilation, Erosion, Opening, and Closing are the four basic operations in morphological operations. To reduce the black pixel and increase the white pixel is in dilation. To increase the black pixel and decrease the white pixel is known as erosion. Erosion then dilation is called the opening. Dilation then erosion are a closing operation.

**Erosion and Dilation**

With X and B as sets in $\mathbb{Z}^2$, the erosion and dilation of X by B, denoted by the equations (1.5) and (1.6) in [39].

\[
X \oplus B = \{ p \in \mathbb{E}^2 : p = x + b, x \in X \text{ and } b \in B \} \\
X \ominus B = \{ p \in \mathbb{E}^2 : p = x + b \in X \ \forall b \in B \}
\]  

(1.5)  

(1.6)

In any image desired object finding is the necessary task for both segmentation (how the object) and classification (where the object) is localized. In mathematical set operations, fundamental morphological operations like erode and dilation interact through structuring elements called struct. The following equations are in [39].

\[
J \oplus K = \{ s \mid (K^c) s \cap J \text{ not equal to } \emptyset \} \\
J \ominus K = \{ s \mid ((K^c) s \cap J) \subseteq J \}
\]  

(1.7)  

(1.8)

**Quality metrics**

In this work, four types of image quality metrics CII, MSE, PSNR, and SD are used to evaluate the contrast between the input image and enhanced images.

**Contrast Improvement Index (CII)**
To assess the seriousness of the diverse differentiation improvement procedures, the most notable benchmark is the picture upgrade measure. It is utilized to think about the consequences of differentiation improvement strategies. Contrast improvement can be estimated utilizing CII as a proportion. CII is characterized as [41]:

$$\text{CII} = \frac{C_{\text{Proposed}}}{C_{\text{Original}}} \quad (1.9)$$

**Mean Squared Error (MSE)**

The MSE is acquired as a total of the square of the blunders between the picture got after equalization and the first picture. Bringing down the worth of MSE implies lower is the mistake [41]:

$$\text{MSE} = \frac{1}{\text{size}} \sum_{i=0}^{n-1} * \sum_{j=0}^{n-1} [f(i,j) - K(i,j)]^2 \quad (1.10)$$

**Peak Signal to Noise Ratio (PSNR)**

It is the proportion of pinnacle square worth of pixels by Mean Square Error (MSE). It is communicated in decibel (db). The PSNR is characterized as [41]:

$$\text{PSNR} = 10 \cdot \log_{10} \left( \frac{\text{MAX}^2}{\text{MSE}} \right) \quad (1.11)$$

**Standard Deviation (SD)**

The standard deviation estimates the spread of the information about the mean worth. It helps look at sets of information that might have a similar mean yet an alternate reach [41].

$$\text{SD} \sqrt{\frac{\sum|x-x|}{n}} \quad (1.12)$$

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a)</td>
<td>mdb26</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>b)</td>
<td>mdb003</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>c)</td>
<td>mdb083</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>d)</td>
<td>mdb175</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3. Input mammogram images from the MIAS database [42].

In Figure 3, Input mammogram images are taken from the MIAS database for diagnosis the breast cancer.

The majority of mammographic datasets are not accessible to the general public. The Mammographic Image Analysis Society (MIAS) database for Screening Mammography is the most easily available and thus most widely utilized database. Furthermore, a few new and old efforts are producing mammographic picture databases at the moment. The majority of mammographic datasets are not accessible to the general public. The Mammographic Image Analysis Society (MIAS) database for Screening Mammography is the most easily available and thus most widely utilized database. In this research, input images are taken from the MIAS database for detecting the tumor region with the help of the region of interest method.
Morphological Operation is an expansive arrangement of picture handling activities that process advanced pictures dependent on their shapes. In a morphological activity, each picture pixel is related to the worth of another pixel in its area.

Classification of Morphological activities:

**Enlargement:** Dilation includes pixels the item limits.

![Figure 3 a) Dilation](image)

**Disintegration:** Erosion eliminates pixels on object limits.

![Figure 3 b) Erosion](image)

**Open:** The initial activity dissolves a picture and afterward widens the disintegrated picture, utilizing the equivalent organizing component for the two tasks.

![Figure 3 c) Opening](image)
Close: The end activity enlarges a picture and afterward disintegrates the expanded picture, utilizing the equivalent organizing component for the two tasks. The quantity of pixels added or eliminated from the article in a picture relies upon the shape and size of the organizing component used to handle the picture.

![Figure 3 d) Closing](image)

In the morphological expansion and disintegration activities, the condition of some random pixel in the resulting picture is controlled by applying a standard to the relating pixel and its neighbors in the info picture. The standard used to handle the pixels characterizes the morphological activity as widening or disintegration.

**MATLAB code for morphological operations:**
Input image = imread (“mdb003.pgm”);
s subplot (2,3,1), imshow(Input image);
title (“Input image”);
on open image = imopen(Input image, se);
s subplot (2,3,2), imshow(open image);
title (“Opening”);
se = strel (“line”, 8,8);
dilate image=imdilate(Input image, se);
s subplot (2,3,3), imshow(dilate image);
title (“Dilation”);
se = strel (“line”, 6,6);
erode image = imerode(Input image, se);
s subplot (2,3,4), imshow(erase image);
title (“Erosion”)

http://www.webology.org
In Figure 4, the output image for breast cancer diagnosis is shown. Where a) Input image, b) Gamma corrected, c) Dilated image, d) Binary image and e) Diagnosis of breast tumor.

<table>
<thead>
<tr>
<th></th>
<th>a) Input image</th>
<th>b) Gamma corrected</th>
<th>c) Dilated image</th>
<th>d) Binary image</th>
<th>e) Diagnosis of breast tumor</th>
</tr>
</thead>
</table>

**Figure 4. Output images for breast cancer diagnosis**
Table 1. Image enhancement quality metrics evaluated by PSNR, RMSE, and SD

<table>
<thead>
<tr>
<th>Image Id</th>
<th>Gamma corrected morphology</th>
<th>Gamma corrected morphology with ROI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR</td>
<td>RMSE</td>
</tr>
<tr>
<td>mdb003</td>
<td>36.34</td>
<td>15.01</td>
</tr>
<tr>
<td>mdb036</td>
<td>33.31</td>
<td>23.64</td>
</tr>
<tr>
<td>mdb042</td>
<td>31.97</td>
<td>32.47</td>
</tr>
<tr>
<td>mdb058</td>
<td>30.71</td>
<td>24.48</td>
</tr>
<tr>
<td>mdb096</td>
<td>33.47</td>
<td>31.20</td>
</tr>
<tr>
<td>mdb110</td>
<td>32.44</td>
<td>38.12</td>
</tr>
<tr>
<td>mdb132</td>
<td>37.14</td>
<td>20.12</td>
</tr>
<tr>
<td>mdb210</td>
<td>40.12</td>
<td>18.54</td>
</tr>
<tr>
<td>mdb232</td>
<td>37.15</td>
<td>26.44</td>
</tr>
<tr>
<td>mdb278</td>
<td>39.86</td>
<td>17.75</td>
</tr>
</tbody>
</table>

In Table 1, the computation of input of mammogram image quality is measured using PSNR, RMSE, and SD quality metrics. In comparison between Gamma corrected morphology and Gamma corrected morphology with ROI, the results showed that the Gamma corrected morphology with ROI seems to have high PSNR, high SD, and low RMSE.

Table 2. CII values for the proposed method

<table>
<thead>
<tr>
<th>Image Id</th>
<th>Gamma corrected morphology</th>
<th>Gamma corrected morphology with ROI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CII</td>
<td>CII</td>
</tr>
<tr>
<td>mdb003</td>
<td>12.1214</td>
<td>14.2458</td>
</tr>
<tr>
<td>mdb036</td>
<td>20.12</td>
<td>36.47</td>
</tr>
<tr>
<td>mdb042</td>
<td>18.6524</td>
<td>24.3114</td>
</tr>
</tbody>
</table>

In Table 2, the predicted Contrast Improvement Index (CII) values of GCM and GCM-ROI comparisons are given.
Figure 5. Comparison of CII value between GCM and GCM-ROI

In Figure 5, CII values of GCM and GCM-ROI comparison are plotted.

4. Conclusion

Medical image processing has very big challenges to detecting and curing diseases in earlier stages. Now the requirements are also faced as critical challenges to every researcher. Therefore, both pathologists and scientists are still have working in the field of finding diseases and caring for humans at an earlier stage. Many areas like image processing and biologically inspired computing are given their support through efficient algorithms. In this research work, gamma-corrected morphological operations with the region of interest method-based erosion and dilation are performed. The proposed method has high CII, PSNR, and SD values with low RMSE for image enhancement.

References

12. https://www.who.int/newsroom/fact-sheets/detail/breast-cancer
42. https://www.kaggle.com/kmader/mias-mammography


Authors Profile

Dr. Irfan Ahmed is currently working as the Director-Research, Industry-Institute Relations at Thassim Beevi Abdul Kader College for Women, Kilakarai, Ramnad, Tamilnadu. He received his MCA from Bharathidasan University in 1994. After that, he obtained his Ph.D in Computer Science with specialization in Trusted Networks from Algappa University, Karaikudi. Dr. Irfan’s research expertise includes Trustworthy Computing, Networking, and Data Mining, etc. He has authored 4 books and published more than 58 international and national journals along with several conference publications. Professor Ahmed has 27 years of experience including teaching, research,
and industry. He has supervised 12 Ph. D students and currently, 3 more are under his supervision. He is the reviewer of a number of journals all over the globe. He has given several invited and contributed presentations in multiple national and international conferences and symposia. He is the life member in IST (Indian Society for Technology Education), ACM (Association for Computing Machinery), Associate life member in IACSIT (International Association of Computer Science and Information Technology), a member in IAENG (International Association of Engineers).

A.Krishnaveni is currently working as an Assistant Professor in the Department of Computer Science at Thassim Beevi Abdul Kader College for Women, Kilakarai, Ramanathapuram, Tamilnadu. She was a Doctoral Candidate in Computer Science at the Chikkanna Government Arts College, Tirupur in 2019-2021. She was an M.Phil Research Scholar in Computer Science at Bharathiar University in 2013-2015. She obtained her M.Sc. in Computer Science in 2013, and B.Sc in Computer Science in 2011 from Bharathiar University. She was started her career as an Assistant Professor and had been on the faculty at the Terf’s Academy since 2015. After that, she continued her service at Park's College and totally has three and half years of teaching experience. She published more than 15 research papers in International Journals/ Springer Conferences. Her research concentrates on Bio-Inspired Computing and her research interests include Medical Image Processing, Data Mining, and Data Structure algorithms.