A novel approach to jointly address localization and classification of breast cancer using bio-inspired approach

Sushma S. J.1, S. C. Prasanna Kumar2
1Instrumentation Technology, Visvesvaraya Technological University, Belgavi, India
2Department of Instrumentation Technology, RVCE Bangalore, India

ABSTRACT
Localization of the cancerous region as well as classification of the type of the cancer is highly inter-linked with each other. However, investigation towards existing approaches depicts that these problems are always individually solved where there is still a big research gap for a generalized solution towards addressing both the problems. Therefore, the proposed manuscript presents a simple, novel, and less-iterative computational model that jointly address the localization-classification problems taking the case study of early diagnosis of breast cancer. The proposed study harnesses the potential of simple bio-inspired optimization technique in order to obtained better local and global best outcome to confirm the accuracy of the outcome. The study outcome of the proposed system exhibits that proposed system offers higher accuracy and lower response time in contrast with other existing classifiers that are frequently witnessed in existing approaches of classification in medical image process.

Keyword: Bio-inspired, Breast cancer, Classification, Detection, Localization, Optimization

1. INTRODUCTION
Detection and classification problems are highly connected issues with respect to medical image processing [1]. Usually this problem is investigated with respect to a specific case of clinical urgency whose information is usually captured in the form of medical images viz. ultrasound, Magnetic Resonance Imaging (MRI), x-ray, computer-aided diagnosis, etc. [2]. The proposed study considers the case of diagnosis of breast cancer where the input is usually obtained from mammograms. Usually, such forms of mammogram inputs could exhibit the presence of any abnormalities if it is visibly larger in size. From the medical image processing viewpoint, the captured image (that is usually a gray-scaled image) is investigated with respect to its specific proportion of contrast to display a presence of any lump, nodule, or tumor [3]. At present, there are good numbers of work being carried out toward studies towards facilitating the localization of the cancerous region e.g. [4]-[6], however, they don’t relate or address any form of classification problem that is next stage of the diagnosis. Similarly, there are also studies carried out toward medical image classification viz. [7]-[9], but they don’t initiate with typical localization problems and hence their own way to solve such classification problems. However, there are potential impediments towards the research work associated with diagnosis and classification problems e.g. i) the complete image is chosen to undergo detection or classification process which also involves some surplus region not required for any form of analysis, ii) lesser focus on preprocessing makes the classification process degrade in the later stage after detection process, iii) adoption of iterative or complex machine learning-based approach that offers good precision at the cost of computational complexity, iv) joint addressing of detection and classification problems are less visualized in
existing approaches, etc. It was also found that segmentation is always a common technique involved in both the process of localization and classification of cancer. Apart from these entire problems, the situation of classification turns most adverse if the classification/detection are supposed to be carried out in early stage of cancer where there is no significant definition of cancerous region for the given medical image. At present, majority of the analysis and investigation is carried out in by manually selecting the region bearing the clinical significance e.g. region of interest. Although adoption of region of interest offers good narrow down of the investigation toward finding the cancerous site but there is no denying the fact that it is highly manual process and is justified for only those images that requires special attention from the physician or radiologist. It is due to practical implementation of region-of-interest for diagnosis hundreds of medical image quite not possible in real-world scenario and this problem can be only solved if the system is capable of identifying the region of the image characterized by cancer. Hence, the practical application will always demand an automatic detection and classification process to perform diagnosis of breast cancer efficiently. The practical parameters to justify such performance in real time are always the response time and accuracy. The proposed manuscript introduces a novel optimization technique that harnesses the potential of bio-inspired algorithm. The contribution of the proposed study is that it offers solution by jointly addressing the problems of detection and classification of breast cancer. The study also implements a rule-set based approach in order to make a user-friendly classification of the breast cancer. Section 1.1 discusses about the existing literatures where different techniques are discussed for detection as well as classification schemes used in diagnosis of early stage of breast cancer followed by discussion of research problems associated with the existing system in Section 1.2 and proposed solution in 1.3. Section 2 discusses about algorithm implementation associated with the localization and classification process followed by discussion of result analysis with respect to visual and comparative analysis in Section 3 using standard performance parameters to assess the proposition. Finally, the conclusive remarks are provided in Section 4.

1.1. Background

This section is a continuation of our prior review work towards approaches of breast cancer detection [10]. Beevi et al. [11] have presented a classifier design using deep belief network for assisting in segmentation and classification of a typical stage of mitosis in cancer progress stage with approximately 85% of accuracy performance. Similar adoption of advanced machine learning was witnessed in the work of Carneiro et al. [12] who have used deep learning approach in order to perform classification along with segmentation of lesions on breast image. Classification problem with respect to mass is also addressed in the work of Chokri and Farida [13] where multi-layer perceptron is utilized. Duraisamy and Emperumal [14] have used deep learning approach in order to perform classification for a given mammogram. The authors have also used convolutional neural network in order to carry out learning process. Elmoufidi et al. [15] have implemented a multiple-instance learning method for facilitating segmentation from pixel-level as well as classification from image-level using region-of-interest. Study towards classifier design was implemented by Manivannan et al. [16] as well as Mercan et al. [17] using learning-based method over multiple instances in order to perform classification. Nizam et al. [18] have carried out spectral methods in order to perform estimation of the spacing from the images obtained from the ultrasound. Rabidas et al. [19] have carried out analysis of classification problem with the help of Ripple-IT transformation technique by quantifying the textural features. Reis et al. [20] have used region-of-interest scheme as well as feature extraction using multiscale-based approach. Saha and Chakraborty [21] addressed the classification problem using deep learning approach along with a segmentation being carried out using semantics. Usage of Fisher vector towards facilitating classification of image is carried out by Song et al. [22]. However, the process of classification potential depends upon how strong is the detection process. There are certain studies carried out towards detections for ensuring better classification process. Strackx et al. [23] have introduced a hardware-based approach for implementing a unique subsampling process for facilitating identification of breast cancer. Investigation of cancer using breast phantoms using microwave imagery was carried out by Wang et al. [24], [25] where the authors have considered time-domain analysis. Yin et al. [26] have implemented a correlation-based method for enhancing the image analysis for breast cancer detection. Hossain and Mohan [27] have used an analytical technique along with consideration of time-domain to find efficient detection of cancer over microwave imaging. Jalilvand et al. [28] have used specific design of bowtie antenna in order to perform detection of the breast cancer. The work carried out by Kwon et al [29] have used Gaussian band-pass filtering in order to perform detection of cancer using three dimensional image. Adoption of S-transforms is reported to upgrade the classification process as discussed by Beura et al. [30]. The authors have also implemented AdaBoost algorithm along with random forest to enhance the classification process. The next section discusses about the research problems associated with existing techniques. Sakthi et al. [31] The Unit Commitment (UC) issue has been prepared by incorporates wind energy generators along with thermal power method.
Hamaine et al. [32] demonstrates the proposed method precisely differentiate standard brain images from the irregular ones and benign lesions from malignant tumors. Similarly Bangare et al. [33] illustrated and used to look for the targeted significance along with revealing the best-focused graphic location by way of aliasing search technique included with novel “Neuroendoscopy Adapter Module (NAM)” method.

1.2. The problem
The significant research problems are as follows:
1) Existing approaches doesn’t offer equal emphasis on jointly addressing the problems of detection and classification of breast cancer.
2) There are less benchmarked model to prove efficiency of classification approach with respect to simplistic and cost effective computational modeling.
3) Adoption of all existing machine learning offer increase precision but at the cost of resource and training dependencies thereby minimizing the practical utility.
4) Majority of the existing mechanism has manual selection of observation area and there are less involuntary techniques to support this phenomenon.

Therefore, the problem statement of the proposed study can be stated as “Developing a cost effective computational modeling for jointly addressing the localization and classification problems associated with breast cancer diagnosis is still an open challenge”. The next section outlines solution to this issue.

1.3. Proposed solution
The proposed work is a continuation of our prior implementation [34] and [35]. In the present work, an integrated framework is modeled that is meant for addressing the joint problems associated with localization as well as classification in breast cancer. The implementation of the proposed system is carried out considering analytical research methodology. The schematic flow of the proposed system is as follows:

![Schematic flow of proposed system](image)

Figure 1. Schematic flow of proposed system

Figure 1 highlights that proposed system first address the localization problem and then the classification problem. In order to address localization process, a multi-layer enhancement is beign carried out with an aid of thresholding and using bio-inspired based implication of objective function. The classification problem is sorted by performing elimination of the surplus region followed by normalization and removal of the pectoral muscle. The outcome is further subjected to discrete wavelet transformation in order to extract decomposed wavelets as the feature. The process is than subjected to the bio-inspired based optimization principle that results in better selection of regions with most probability of features bearing cancerous region. The inferencing of the outcome is carried out by applying rule-set that significant assists in performing binary classification process. Therefore, the proposed system offers a progressive mechanism to address both the problems in order to offer better classification performance. The outcomes are made with respect to binarized classification in the form of malignant (abnormal) and benign (normal) state of breast cancer. The next section outlines algorithm implementation.
2. ALGORITHM IMPLEMENTATION

The primary function of the core algorithm is to ensure an effective detection followed by classification of the breast cancer from the captured medical image. The proposed algorithm uses bio-inspired algorithmic approach for designing the proposed algorithm. The complete operation of the algorithm is discussed with respect to algorithm design for localizing the area of breast cancer and algorithm for binarized classification of the breast cancer as following:

2.1. Algorithm design for localizing the area of breast cancer

This algorithm is responsible for localizing the exact position of the cancer in the breast cancer for a given medical image. Applying the method of simple bio-inspired technique, the algorithm takes the input of I (input image) and yield the outcome of Iloc (Output image) with identification of the cancerous region. The steps of the algorithm are as following:

Algorithm for Localizing the Area of Breast Cancer
Input: I (Input Image)
Output: Iloc (Output Image)
Start
1. \textbf{init I}
2. \(I_{\text{seg}} \rightarrow f_1(\text{Th}(I))\)
3. \(I_{\text{prim}} \rightarrow f_2(I_{\text{seg}})\)
4. \(I_{\text{sec}} \rightarrow f_3(I_{\text{prim}})\)
5. \(I_{\text{ter}} \rightarrow f_4(I_{\text{sec}})\)
6. \(I_{\text{op}} \rightarrow f_5(I_{\text{ter}})\)
End

The description of the algorithm is as follows: After taking the input of medical image (Line-1), it is subjected to different explicit functions to carry out different processing. The algorithm introduces a function \(f_1(x)\) that is essentially meant for carrying out involuntary segmentation process (Line-2). This function implements a cut-off operator \(\text{Th}\) over the input image \(I\) followed by obtaining binarized image in order to construct a suitable mask. The algorithm than onstricts the highest possible mask and continue labeling it followed by the concatenation of all the area. This operation leads to the maximum value of the concatenated area of the mask. The segmentation is carried out over original image as well on masked image ensuring that only one specification of mask is considered. The proposed algorithm also performs a simplified operation using another function \(f_2(x)\) that takes in the input of segmented image. The algorithm first performs local contrast modification considering input arguments of processed image and weight factor. (By processed image, it will mean applying unsigned integer of 8 bits on input image followed by altering the precision to double). The obtained image from the local contrast modification is then subjected to the entropy formulation followed by applying sobel operator to obtain prominent edges. An objective function is defined as:

\[
\alpha = \log(\log(\text{E})). \frac{\delta}{\text{MN}}
\]  

(1)

In the above expression (1), an empirical expression of objective function \(\alpha\) is constructed that considers the edge components \(\text{E}\) obtained from summation of all the edges using sobel operator of the area obtained after local contrast modification. Applying a simple bio-inspired approach, if \(\alpha>\text{gbest}\), than the original value of \(\alpha\) is considered as the gbest or else a probability value [0.1-1] is assigned to the pbest. The next step of the algorithm is to apply the secondary enhancement using function \(f_3(x)\) that is developed on the basis of threshold optimization (Line-4). The algorithm computes the probability as well as histogram for the actual input image followed by initialization of mean and weight factor. Computation of variance is carried out and only the variance matching with threshold is considered for the further computation. This summation of maximum value of this new variance is used for obtaining the new threshold value. The outcome is then subjected to the tertiary enhancement using a function \(f_4(x)\). In this case, the outcome image Isec is subjected to binarization followed by checking the situation when the value of the binarized image is more than 10, which is only the case of either lump or nodule in the breast tissue. The final function \(f_5(x)\) is applied to ensure that the region infected with cancer is identified (Line-4 and Line-5). A slight amount of recursive function is designed to apply probability to ascertain that there is a regular update of the threshold parameter in order to ensure a better for of identification process of the region detected with the
Another interesting fact of the proposed algorithm is that it offers a significant insights of the higher contrastive region to be normal tissue or cancer-inflicted tissue in order to ensure that a simple and accuracy identification of carried out using bio-inspired approach.

2.2. Algorithm for binarized classification of the breast cancer

The prior algorithm contributes in carrying out localization of the region infected with breast cancer while this algorithm assist to carry out a simple classification technique by evolving up with a simple and novel bio-inspired approach. This algorithm is essential for implementing a novel bio-inspired approach for carrying out identification followed by binarized classification of the outcome. The algorithm takes in the input of input image that after processing results in CC (Center of cluster) and Iout (classification outcome). The steps of the algorithm are as follows:

Algorithm for Binarized Classification of the Breast Cancer
Input: I (input image)
Output: CC (Center of cluster), Iout (classification outcome)
Start
1. \( I \rightarrow f_6(I) \)
2. \( I \rightarrow f_8(\text{prec}(f_7(I))) \)
3. If \( j > \text{c.maxI} \)
4. \((s_1, s_2) = (s_1, s_2) + (j, h(idx), h(idx)) \)
5. End
6. For \( k = 1:n \)
7. \([p_1, p_2] \rightarrow [Rw(idx_1), Rw(-idx_1)] \)
8. For \( l = x_l^T \)
9. idx = idx + 1
10. If \( l > T \)
11. \((s_{1b}, s_{2b}) = (s_{1b}, s_{2b}) + (l, h(idx), h(idx)) \)
12. End
13. If \( 1 < T \) &\& \( l > (T - c_{1\cdot\text{maxI}}) \)
14. \((s_{1b}, s_{2b}) = (s_{1b}, s_{2b}) + (l, h(idx), h(idx)) \)
15. End
16. End
17. \( I_{op}(x_{\text{win}-10:x_{\text{win}}+10, 1:A_y}) = I_{seg} \)
18. If \( \text{Fit} < P_{\text{best}} \)
19. \([\text{SIpart} P_{\text{best}}] \rightarrow \text{CC} \) & \( P_{\text{best}} = \text{Fit} \)
20. Else
21. store prior value of SIpart
22. If \( P_{\text{best}} < G_{\text{best}} \)
23. \( G_{\text{best}} \text{CC} = P_{\text{best}} \text{CC}, G_{\text{best}} = P_{\text{best}} \)
24. \( \text{CC} = \text{SIpart}(g_{\text{best}id, idvec(g_{\text{best}id})) \)
25. \( I_{\text{out}} \rightarrow \text{bin(CC, ‘Malignant’, ‘Benign’)} \)
End

The steps of the algorithm are as follows: The algorithm uses a function \( f_6(x) \) that is meant for assessing the left and right orientation of an image followed by correcting the orientation for enhancing the classification process (Line-1). The next step of the algorithm is to carry out segmentation to ensure that no unwanted region is selected for next process of analysis (Line-2). For this purpose, the segmentation is carried out by obtaining the binary image using two different explicit function \( f_7(x) \) and \( f_8(x) \). The next part of the algorithm is all about applying a bio-inspired algorithm in order to remove the unwanted tissue that creates an impediment towards identifying cancerous region (Line-3 to Line-24). The algorithm obtains histogram \( h \), index \( idx \), in order to obtain windows \( s_1 \) and \( s_2 \). The algorithm performs dual classification of the region viz. \( p_1 \) and \( p_2 \) followed by evaluation of the threshold value \( T \) that is equivalent to \( s_1/s_2 \). The algorithm further computes updated threshold followed by evaluating fitness value fit with respect to the pbest value. Likewise, the similar check is carried out towards assessing the comparative value of pbest with respect to gbest. This process is resumed by computing center of cluster that is considered to be the region of best outcome for the given frame of an image. The prime agenda behind designing the optimization technique is to filter out both pbest and gbest from the given problem space, where the fitness function is consistently updated if there is any form of change in the dimension of the problem space. For this purpose, if there is any form of slightest deviation for the images density (that occurs in different image samples),
it can easily identify the location. However, the significant benefit is that it checks for the complete region in order to avoid false positive while making decision. Finally, fuzzy inference system is utilized in order to further ascertain the classification outcome for stating whether it is malignant or benign state of breast cancer. The next section discusses about the outcomes obtained.

3. RESULT ANALYSIS

From the discussion of algorithm implementation, it can be seen that proposed system performs localization as well as classification of the breast cancer from the medical dataset e.g. DDSM [36] and MIAS [37]. Hence, the analysis of the proposed system is carried out in two discrete way viz. visual assessment and numerical assessment. Following are the discussion of the outcomes.

![Visual outcomes of localization of breast cancer](image)

Figure 2 highlights the visual outcomes to show the stages of processing involved in identification of the breast cancer. A closer look into the outcome shows that after performing involuntary segmentation of the input image, there are three stages of bio-inspired image enhancement i.e. primary, secondary, and tertiary enhancement Figure 2(c) to Figure 2(e). It evidently shows that each progress rendered by the consecutive process of the proposed system entails the increase of the local contrast along with removal of the unwanted regions. The complete processing time in order to yield this outcome is approximately 0.26657 seconds in windows. Similarly, the proposed system also testifies the visual outcomes of the classification process as highlighted in Table 1. The visual outcome shows that input for both normal and abnormal images are initially assessed for any form of unwanted spaces that is not at all considered in the analysis. Hence, after removing all the unwanted regions, the proposed system performs normalization of the images. A closer look into this process of normalization will show that proposed system performs normalization in quite a different manner for both normal and abnormal images. Further the process of feature extraction is
carried out with an aid of decomposed wavelets obtained by applying discrete wavelet transformations. Further applying the novel bio-inspired algorithm, the actual region of breast cancer is finally localized and is now ready for classification. The proposed system applies rule-set based approaches using binary classification process, where the finally localized region is declared as benign or malignant stage of cancer after observing the normal or abnormal stages of cancer.

Table 1. Visual Outcome of Classification

<table>
<thead>
<tr>
<th>Input</th>
<th>Normal</th>
<th>Abnormal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Removal of unwanted region</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
</tr>
<tr>
<td>Normalize Image</td>
<td><img src="image3" alt="Image" /></td>
<td><img src="image4" alt="Image" /></td>
</tr>
<tr>
<td>2d-DWT</td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
</tr>
<tr>
<td>After applying bio-inspired</td>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
</tr>
<tr>
<td>Classification Outcomes</td>
<td><img src="image9" alt="Image" /></td>
<td><img src="image10" alt="Image" /></td>
</tr>
</tbody>
</table>
The proposed system also performs comparative analysis to evaluate the performance of the classification. The outcome of Figure 3 shows that mean and standard deviation of proposed system as well as existing classifiers e.g., Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Artificial Neural Network (ANN). Similarly, proposed system also offers reduced skewness and kurtosis value in Figure 4. The classification accuracy of the proposed system is significantly high compared to existing classifiers (in Figure 5). Apart from this, the proposed system also offers faster computational processing time to show that it is cost-effective approach to address the joint localization and classification of the breast cancer with a good balance between accuracy and faster response time in Figure 6.

![Figure 3. Comparative analysis of mean and deviation](image1)

![Figure 4. Comparative analysis of skewness & kurtosis](image2)

![Figure 5. Comparative analysis of accuracy](image3)

![Figure 6. Comparative analysis of response time](image4)

4. CONCLUSION

The proposed research work offers an insight that it is feasible to present a solution towards jointly addressing the problems associated with detection and classification problems associated with early stage of detection of breast cancer. The proposed system initially addresses localization problems by using a novel multi-layer enhancement using novel threshold-based approach along with simple bio-inspired optimization that allows its objective function to offer highly accurate outcome of localized region automatically. The second part of the implementation discusses about a novel classification approach that offers significant novelty over elimination of unwanted regions that hinders the classification of the breast cancer. A novel bio-inspired algorithm is implemented to ensure that it obtains both local and global outcome for ensuring highly correct classification process. The outcome is finally utilizing rule-set system in order to perform user-friendly inference of the criticality of cancer in the form of benign and malignant stage.

REFERENCES


A novel approach to jointly address localization and classification of breast cancer using… (Sushma S. J.)