Segmentation and classification using image processing and supervising learning framework for mitosis detection in breast cancer mammographic images

P Krithika1, Sukruti Ramesh2, Padigi Reddy Satya3,7 and Dr C A S Deiva Preetha4,8

1Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus, No.1 Jawaharlal Nehru Road, Vadapalani, Chennai, Tamil Nadu, India
2Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus, No.1 Jawaharlal Nehru Road, Vadapalani, Chennai, Tamil Nadu, India
3Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus, No.1 Jawaharlal Nehru Road, Vadapalani, Chennai, Tamil Nadu, India
4Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus, No.1 Jawaharlal Nehru Road, Vadapalani, Chennai, Tamil Nadu, India

Mail id : 
5pm5288@srmist.edu.in 
6st1914@srmist.edu.in 
7pp2696@srmist.edu.in 
8deivaprc@srmist.edu.in

Abstract. Detection of the area withholding the mitotic cell growth is a vital marker in breast cancer detection. This paper aims to fabricate an automatic Computer-Aided Detection (CAD) [2] model that helps to locate the region of mitotic cell growth and signify the type of breast cancer found: benign or malignant. Contrary to the legacy literature [1], which uses partially supervised learning models applied to histopathology images, we devise a model which involves a fully supervised convolution model applied to mammographic images. The model trained with image datasets: benign and malignant breast cancer images. This model exploits the MIAS database and datasets collected privately from hospitals, consisting of mammographic images available as samples for breast cancer detection. Applying image segmentation techniques to the datasets, we highlight the region of interest, and thereby using classification methodologies, we separate the results as benign or malignant. The developed model equips us to yield results with an accuracy of 97.96% on the dataset.

Keywords: Mitosis Detection, Computer-Aided Detection, Segmentation, Classification, fully convolution network, mammographic images.

1. Introduction
Breast Cancer among the Indian woman population accounts for up to 14%. For every four minutes, one woman is being diagnosed with Breast Cancer. In the urban cities, 25-30% of women have succumbed to this. The survivorship for Breast Cancer patients is low due to their lack of mindfulness and hence late detection. In recent years, Clinical Breast Examination (CBE), also known as Mammogram, is interpolated in the overall women's master health check-up to increase awareness about the significance
of breast cancer detection during the earlier stages. Women have to take their tests annually or as advised by the doctor. Mammography consists of two types: Screening and Diagnostic. Screening mammography is to be performed annually or once in two years for asymptomatic patients. This mammographic methodology is for women with no visible symptoms. Diagnostic mammography's performed for symptomatic patients, where abnormalities are detected visibly. Mammography images are of types, bilateral craniocaudal (CC) and mediolateral oblique (MLO) views. Our model uses these image data sets to detect and classify the type of Breast cancer the patient is carrying, Benign or Malignant. The Computer-Aided Detection model we have designed accepts the mammographic images, performs the process of image segmentation, highlights the region of interest, and helps classify the tumor type. The image segmentation and classification techniques are performed by applying the CNN algorithm. MIAS is image and labels (or) annotations for the mammogram scans. The kernel shows how “Info.txt” and the Portable Gray Map (PGM) files are often parsed accurately. Due to a higher demand, the authentic MIAS database (digitized at fifty-micron pixel edge) has been lowered to two-hundred-micron pixel edge and clipped (or) padded so each image: 1024 × 1024 pixels. We were liberal to use the database in your research project however you must abide by the license agreement once victimization the imagery

Figure 1. Flow Chart explaining the proposed system

2. Proposed system
Breast cancer being one of the crucial threats in a female’s life. Earlier detection of tumour cells irresistibly aids to reduce the mortality rates. This paper proposes a novel Computer-Aided
Detection (CAD) model which reduces human error and automatically diagnoses the lumps from the mammogram screened images and indicates the existence of mitotic growth with its severity. The mammogram screened images are traditionally involved in the earlier screening stages due to relatively lower upkeep and are highly sensitive to minor lesions. The developed model improves the interpretation accuracy and distinguishing between benign and malignant tissues by utilizing basic morphological operations. The input Region of Interest (ROI) is identified manually using power-law transformation and exposed to further stages. After that, a CNN classifier is introduced to classify the required class of breast cancer.

3. Pre-processing
Pre-processing of the input images is one of the fundamental stages in the developed model. Pre-processing of the images has helped us reduce the noise and shift focus towards the regions of interest. The size of raw data pixels is usually large, giving room for redundancy. In this stage, the mammogram scanned images are taken in as the input, altered to fit the constraints (size, intensity, colour format). Images resized to fit the evaluation constraints, hence increasing the accuracy of the results. Images are also converted to Gray Scale from black-white to avoid the evaluation of unwanted regions.

4. Segmentation
The proposed model applies the concept of digital image processing - region-based segmentation to find the regions of interest. We have fed our algorithms with the required constraints that will help identify mitotic cell growth. The parameters involved are size, density, and symmetry. Recursive clustering and masking of the images performed to obtain the accurate regions. Basic morphological operations such as dilation, erosion, area opening and closing, box bounding, and border clearing, performed using the input before converting them to Gray Scale images. At the end of this stage, the input image will now have the region of interest highlighted.

5. CNN classification
The proposed model uses the convolution neural network (CNN) as the classifier. Using the convolutional neural network (CNN), we obtain the type of mitotic cell growth present in the given scans. Features similar to the visual cortex are the fundamental concepts of a CNN classifier. Computers perceive an image in the form of an array of pixels. Each pixel carries its unique feature, thus allowing...
the computers to differentiate. By using a CNN classifier, we first pass a pre-processed input image. The image further undergoes a series of convolution-nonlinear layers, pooling, and fully connected layers.

\[
s(t) = (x * w)(t) = \sum_{a=-\infty}^{a=\infty} x[a] w[a + t]
\]

**5.1. Image input layer**
The pre-processed input image fed into the classifier gets converted into an array of pixels. The size of the pixel array is 28x28x1. Channel size is one, as the input image is in grayscale. The "trainNetwork" function helps us train the networks.

\[
\text{net} = \text{trainNetwork}(\text{imdsTrain}, \text{layers}, \text{options})
\]

**5.2. Convolutional layer**
The pilot layer of the algorithm is the convolution layer. The layer breaks down the mammogram scans into smaller matrices, starting from the top left corner. "Filter-Size" is the first argument in the convolution layer. The argument (height, width, and density) for filter-size is the same as the training function. The filter size in the proposed model is three-by-three. "Num-Filters" is the second argument. This argument describes the number of neurons connected within the same region. This argument determines the number of feature maps.

\[
\text{layer} = \text{convolution2dLayer}(3,8,'Padding','same')
\]

The spatial output size must be the same as the input size. To achieve this, we have enabled "Same" padding, as the default stride for a convolution layer is one. Each hidden layer becomes the input for the upcoming hidden layer. An activation layer is introduced to bring a nonlinear property. The ReLU layer acts as the activation layer. The ReLU layers operate a threshold function. This function sets the input value as "0" for any input value lesser than zero and sets the input value as "x" otherwise.

**5.3. Batch normalization layer**
The Batch-normalization layer is between the convolution layer and the rectified linear unit layer (ReLU). By using this layer, we have decreased the sensitivity that occurs during network initialization. Also, the batch-normalization layer has helped fasten the training of the available network.

**5.4. Max pooling layer**
Reducing the volume of the image without affecting the vital values was one of our concerns. Downsampling of an image helped us achieve this by eliminating the supernumeraries. Max pooling layer selects the maximum element from the feature map. The output of a max-pooling layer is a feature map that consists of the most dominant features. The parameters of the max-pooling function are
poolSize, Name, and Value. The poolSize is [2,2], the value of the name-value pair is "Stride", this gives us a max-pooling layer of size [2,2] and stride [2 2].

5.5. Fully connected layer
The fully connected network (FCN) layer is the penultimate layer in a CNN classifier. FCN mimics a feed-forward neural network. The input for this layer is a flattened output vector of the max-pooling layer. The activation function carries the following mathematical formula as shown below.

\[
ReLU = g(Wx + b)
\]

Where
\( g \) => activation function  
\( W \) => weight of the matrix \([l,e]\); \( p \) stands for the number of neurons in the legacy layer, and \( n \) stands for the number of neurons in the existing layer.  
\( x \) => input vector with the dimension \([l,1]\)  
\( b \) => bias vector with the dimension \([l,1]\)

FCN applied in our proposed image feature detection is due to the following reasons: 1. Sparse interactions. Using FCN has enabled us to concentrate even on the smaller subsets of the array pixel. The efficiency is improved statistically. 2. Parameter Sharing. Instead of applying redundant parameters, we have shared the vital parameters throughout the classifier. Limited memory space and has lowered the costs of the proposed model.

5.6. Softmax-layer
In the end, we have applied Softmax Layer to obtain the probabilities of the mitotic growth in the mammogram scans. The softmax layer checks which class does the detected growth falls in. The output dimensions obtained using the following calculation:

\[
W2 = (W1 - F + 2P / S) + 1
\]

Where
\( W1 \) => Height (or) Width of the Input  
\( P \) => Padding in the layers  
\( S \) => Stride value  
\( W2 \) => Dimensions of the Output

At the end of this layer, the developed model determines the type of breast cancer the patient is carrying.

Figure 3. Full Convolutional Neural Network.
6. Conclusion
The proposed paper yields a breast cancer detection and classification model which applies region and texture-based features for its analysis. The system evaluated using a set consisting of mammogram image databases. A convolution network helps classify tumor cells in the mammogram images, and image processing detects mitotic cells. The developed model is efficient with increased scalability in the detection and classification of breast cancer.

Figure 4. The above pictures display the end result of the designed model.
7. References

[1] M. Sebai, T. Wang and S. A. Al-Fadhli, "PartMitosis: A Partially Supervised Deep Learning Framework for Mitosis Detection in Breast Cancer Histopathology Images," in IEEE Access, vol. 8, pp. 45133-45147, 2020, doi: 10.1109/ACCESS.2020.2978754.

[2] J.E.E. Oliveira, M.O. Gueld, A. de A. Araújo, B. Ott, T.M. Deserno, Towards a Standard Reference Database for Computer-Aided Mammography, Proceedings of SPIE, Vol. 6915, Paper ID 69151Y, 2008

[3] R. L. Siegel, K. D. Miller, S. A. Fedewa, D. J. Ahnen, R. G. S. Meester, A. B. M. PhD, and A. J. D. PhD, “Colorectal cancer statistics, 2017,” Ca A Cancer Journal for Clinicians, vol. 67, no. 3, pp. 177–193, 2017.

[4] J. B. Harford, “Breast-cancer early detection in low-income and middle-income countries: do what you can versus one size fits all,” Lancet Oncology, vol. 12, no. 3, pp. 306–312, 2011.

[5] C. Lerman, M. Daly, C. Sands, A. Balshem, E. Lustbader, T. Heggan, L. Goldstein, J. James, and P. Engstrom, “Mammography adherence and psychological distress among women at risk for breast cancer,” Journal of the National Cancer Institute, vol. 85, no. 13, pp. 1074–1080, 1993.

[6] P. T. Huynh, A. M. Jarolimek, and S. Daye, “The false-negative mammogram,” Radiographics, vol. 18, no. 5, pp. 1137–54, 1998.

[7] M. G. Ertosun and D. L. Rubin, “Probabilistic visual search for masses within mammography images using deep learning,” in IEEE International Conference on Bioinformatics and Biomedicine, 2015, pp. 1310–1315.

[8] S. D. Tzikopoulos, M. E. Mavroforakis, and H. V. Georgiou, “A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry,” Computer Methods and Programs in Biomedicine, vol. 102, no. 1, pp. 47–63, 2011.

[9] D. C. Pereira, R. P. Ramos, and M. Z. do Nascimento, “Segmentation and detection of breast cancer in mammograms combining wavelet analysis and genetic algorithm,” Computer Methods and Programs in Biomedicine, vol. 114, no. 1, pp. 88–101, 2014.

[10] S. A. Taghanaki, J. Kawahara, B. Miles, and G. Hamarneh, “Paretooptimal multi-objective dimensionality reduction deep auto-encoder for mammography classification,” Computer Methods Programs Biomed, vol. 145, pp. 85–93, 2017.

[11] X. W. Chen and X. Lin, “Big data deep learning: Challenges and perspectives,” IEEE Access, vol. 2, pp. 514–525, 2014.

[12] K. Ganesan, U. R. Acharya, C. K. Chua, L. C. Min, K. T. Abraham, and K. H. Ng, “Computer-aided breast cancer detection using mammograms: A review,” IEEE Reviews in Biomedical Engineering, vol. 6, pp. 77–98, 2013.

[13] X. Sun, W. Qian, and D. Song, “Ipsilateral-mammogram computer-aided detection of breast cancer,” Computerized Medical Imaging and Graphics the Official Journal of the Computerized Medical Imaging Society, vol. 28, no. 3, pp. 151–158, 2004.

[14] N. Saidin, U. K. Ngah, H. Sakim, and N. S. Ding, Density based breast segmentation for mammograms using graph cut and seed-based region growing techniques. IEEE Computer Society, 2010.
[15] S. Xu, H. Liu, and E. Song, “Marker-controlled watershed for lesion segmentation in mammograms,” Journal of Digital Imaging, vol. 24, no. 5, pp. 754–763, 2011.

[16] K. Hu, X. Gao, and F. Li, “Detection of suspicious lesions by adaptive thresholding based on multiresolution analysis in mammograms,” IEEE Transactions on Instrumentation and Measurement, vol. 60, no. 2, pp. 462–472, 2011.

[17] M. H. Yap, G. Pons, J. Marti, S. Ganau, M. Sentis, R. Zwiggelaar, A. K. Davison, and R. Marti, “Automated breast ultrasound lesions detection using convolutional neural networks,” IEEE J Biomed Health Inform, vol. 22, no. 4, pp. 1218–1226, 2017.

[18] K. C. Jr, L. M. Roberts, K. A. Shaffer, and P. Haddawy, “Construction of a bayesian network for mammographic diagnosis of breast cancer,” Computers in Biology and Medicine, vol. 27, no. 1, pp. 19–29, 1997.

[19] Z. Wang, G. Yu, Y. Kang, Y. Zhao, and Q. Qu, “Breast tumor detection in digital mammography based on extreme learning machine,” Neurocomputing, vol. 128, no. 5, pp. 175–184, 2014.

[20] Y. Qiu, Y. Wang, S. Yan, M. Tan, S. Cheng, H. Liu, and B. Zheng, “An initial investigation on developing a new method to predict short-term breast cancer risk based on deep learning technology,” in Medical Imaging 2016: Computer-Aided Diagnosis. International Society for Optics and Photonics, 2016.

[21] W. Sun, T. L. Tseng, B. Zheng, and W. Qian, A Preliminary Study on Breast Cancer Risk Analysis Using Deep Neural Network. Springer International Publishing, 2016.

[22] Z. Jiao, X. Gao, Y. Wang, and J. Li, “A deep feature-based framework for breast masses classification,” Neurocomputing, vol. 197, pp. 221–231, 2016.

[23] J. Arevalo, M. A. G. Lopez, and M. A. G. Lopez, “Representation learning for mammography mass lesion classification with convolutional neural networks,” Computer Methods and Programs in Biomedicine, vol. 127, pp. 248–257, 2016.

[24] G. Carneiro, J. Nascimento, and A. P. Bradley, “Automated analysis of unregistered multi-view mammograms with deep learning,” IEEE Transactions on Medical Imaging, vol. 36, no. 11, pp. 2355–2365, 2017.

[25] Y. Kumar, A. Aggarwal, S. Tiwari, and K. Singh, “An efficient and robust approach for biomedical image retrieval using zernike moments,” Biomedical Signal Processing and Control, vol. 39, pp. 459–473, 2018.