Approach for classification of medical image using deep learning technique

[1]Shweta Prafull Ajmera, [2]Prof Sarika B. Solanke

[1]M.Tech Student, [2]Professor
[1,2]Computer Department
Deogiri Institute of Engineering and Management, Aurangabad

***Abstract*: This paper tends to the issue of fragmenting an image into the segment. The magnetic resonance imaging (MRI) process is susceptible to a wide range of artifacts caused by various sources. In some cases, artifacts might be confused with pathology. In addition, state-of-the-art dynamic MR reconstruction algorithms are iterative in nature, causing longer reconstruction times. Recently, deep learning has been applied to MRI reconstruction and produces high quality images at high acceleration rates. Since deep learning highly depends on training data, the quality of training images must not be ignored. This article demonstrates how noisy images in the training data affect the quality of MR reconstruction. The proposed method modifies the loss function of the neural network to prefer higher quality target images by using a weighted loss function. In this paper mean squared error loss is used, but the approach can be extended to other types of loss function. Using still frames from cardiac MRI's, this approach is compared to existing approaches that discard noisy training data or ignore these quality differences. Even a basic weighting strategy improves the deep learning reconstruction quality over such methods. Our purpose is to develop a mammography-based DL breast cancer risk model that is more accurate than established clinical breast cancer risk models.**

***Index Terms*: MRI, Breast Cancer, Tumor, deep learning**

I. Introduction

Recent studies show that breast cancer is the most common cancer among women [1], which accounts for about one-third of all newly diagnosed cancers in the United States. [2] The death rate from breast cancer is also high Because it represents 17% of cancer-related deaths in general. [3] Early detection and evaluation of breast cancer are especially important when Reducing mortality Mammography is the most useful tool for detecting the general population. However, the correct examination and diagnosis of breast lesions alone depends on the results of the breast examination being difficult and depends on the experience of many radiologists, which leads to many false tests. And additional testing [4]. Computer-aided detection and diagnosis (CAD) systems have been used to provide important assistance in the radiological decision-making process. Such systems can reduce the amount of effort required for the evaluation of lesions in clinical practice while reducing the number of false-positive results that lead to unnecessary and inconvenient tissue cutting. Mammogram testing can handle two different tasks: detecting suspected lesions on a mammogram (CADe) and diagnosing a detected disease (CADx). Is the Classification of cancer or cancer.

Deep learning has been a technology that has made great progress in recent years, as it demonstrates superior performance in machine learning in the various machine learning tasks. The detection and classification of objects In contrast to the general method of machine learning, which requires a process to extract handmade features, which is a challenge, since it depends on the domain knowledge, the deep learning method learns the extraction process. Of the appropriate input characteristics About target output, This will eliminate the tedious process of engineering and inspection of the ability to distinguish characteristics while facilitating the repetition of methods. Since the appearance of deep learning, many works have been published, which make use of in-depth architecture. [5] The most common deep learning architecture is the neural network (CNN), Arévalo and the faculty. [6]] Test multiple CNNs and compare them with hand-made descriptors for overall diagnostic tasks. Their experiments were carried out in the BCDR-FM data set. They reported performance improvements with a combination of learning and handmade representations. However, the author has not tested the effectiveness of the previously trained networks and uses a simpler CNN architecture. Carneiro and the faculty [7] used the previously trained CNN, which was precisely adjusted using non-mammograms. Have registered and They evaluated the risk of developing breast cancer. According to BIRADS, they came to the conclusion that the trained model was randomly better than Huynh et al. [8] using AlexNet that has been Training beforehand [9] without many adaptive diagnoses They analyze the performance of the classification using the characteristics of multiple middle layers of networks that use SVM for classification. They compare their results with two methods: classifiers that work with handmade styles and a set of both using gentle voting, hybrids, and faculty. [10] The proposed scheme at the CNN that has been Previous training was done in a subset of the DDSM database and then extracted the characteristics of the mass of the various layers of this model. In this way, receive the "high level" and "medium" characteristics that correspond to different scales.

Two linear SVM classifiers are trained for one decision procedure for each group of their characteristics and predictions to be combined. Levy and Jain [11] classified the images using AlexNet and GoogleNet, they compare learning by transfer with training from the outset, finding that in the past, superior results were obtained. It is noteworthy that they examined the effects of the context of the data, concluding that cutting large-sized fixed-edge boxes around the incision are more effective when compared to proportioning cuts by Ting and the faculty [12] Create and train their nets for breast mass classification from the start The network consists of 28 layers and is fully connected and feeds from the ROI of the proposals detected by a single bullet detector. They conducted experiments in the MIAS Rampun database and faculty, [13] using a set of AlexNet trained and pre-tuned versions in CBIS-DDSM. During their inference, they chose all three versions as well. The best performance and combine their predictions.

Most modern jobs propose to use a network that has been trained before training from the start. However, the advanced network is designed and tested in a more diverse set of data in different ways and larger commands than the existing Mammogram data set. many As a result, the capacity, and complexity of such networks can exceed the needs of a small data set, which leads to a significant impact when training from the start. As a result, many works have appeared that the author offers training from the start.

Taking into account the above, in this document, we examine the performance of various networks We compare the performance of each network in two situations: the first is about starting a pre-trained weight training, and the second, a network beginning with random weights..

II. Related Work

AlexNet [9] is a neuron network. The first convolutional (CNN) that shows performance beyond state of the art in object detection and classification. As shown in Figure 1, the network has eight layers. The first five are convolutional. And the remaining three are completely connected The first layer of the network filters the input image (size 224 × 224) with 96 cores 11x11 with step 4 pixels. The depth of these image cores is equal to the number of input image channels. The second layer is used as the input to the output of the first layer. After normalization of the local response and the maximum grouping is applied, it is filtered with 5 × 5 × 96 256-sized nuclei. The third, fourth, and fifth layers are connected. There is no grouping or medium standard. The third layer has 384 grains of 3 × 3 × 256. The fourth layer has 384 grains of 3 × 3 × 384 grains and the fifth layer has 256 grains of 3 × 3 × 384 on the top. Of class convolutional, The layer connects completely with 4096 neurons each. The number of neurons in the third class that is fully connected is equal to the number of classes. However the above work has considerable limitation such as a false-negative mammogram looks normal even though breast cancer is present. A false-positive mammogram looks abnormal even though there's no cancer in the breast.

Along with the unique architecture of the network, the author [9] also introduced some new features that greatly helped the network's ability to learn and talk in general. The most important feature is that they replace the activation function of standard neurons. (Logistic and hyperbolic tangent functions) with the modified linear function f (x) = maximum (0, x). The neurons that use this activation function are called The modified linear unit (ReLU). The advantage of this activation function is that it charges unsaturated nonlinear values in contrast to saturated sigmoidal functions for large values. This will provide better gradations with better calculation performance. ReLU was created as a standard option of the activation function for CNN. The author also recommends the standardization model in Depth for each position of the map, features created by layers This type of response adjustment creates competition for large-scale activities between neuron results calculated using different cores while LRN is used and combined. Into many other network architectures. It was removed from AlexNet in a later publication. [14].

However this work has following limitation, The odds of a false-positive finding are highest for the first mammogram. Women who have past mammograms available for comparison reduce their odds of a false-positive finding by about 50%. An extremely important aspect of the training is the use of attrition [15] (with a 0.5 probability) for all three connected layers. This technique consists of zeroing the output of each hidden nerve cell with certain probabilities. The selected neurons do not contribute to moving forward or for retroactive diffusion. Therefore, in each practice, different architectural samples are sampled. The abandonment technique acts as a regulation to force the network to learn important features. But will increase training time.

The authors [16] have examined the effects of network depth while maintaining a small conversion filter, they show that significant improvements can be made by pushing the depth to the 16th floor. –19 Input to layers convolutional It is a 224 × 224 fixed-size image. The image is passed through the layer stack. Convolutional by activating ReLU by using filters with very small openings (3 × 3), moving forward with matching is still 1. The spatial combination is carried out by the highest profit consolidation layer. The five layers, which are implemented after some layers of transformations, as well as the AlexNet stack of three fully connected layers, will be at the top of the section. The advantage of VGG is that it adds layers. Convolutional Multi-layer with a small granularity, enabling the receiving field to increase network efficiency while reducing the number of parameters compared to using layers Convolutional With fewer seeds in the same open field The author tests configurations of various levels of different depths (9, 11, 16 and 19 layers). In one of the configurations, 1 × 1 filters are used, which can be seen as a linear transformation. Line of input channel this is a way to increase the nonlinearity of the decision function without affecting the open field of the layer. Convolutional one configuration also includes the LRN layer as reported in the paper.

GoogLeNet [17] is the first implementation using the Inception module. The main idea behind this module is based on the author's findings of how sparse structures in space can be estimated by dense components. Their goal is to find the best local structure and repeat it. Build a multi-layer network. The Inception module consists of four branches that receive the same input. The first branch filters the input with a 1 × 1 conversion, which acts as Linear conversion on input channels The second and third branches perform 1 × 1 kernel convolutions to reduce the dimensions followed by the 3 × 3 and 5 × 5 seed layers respectively. The fourth branch carries the maximum profit, followed by finally twisting with 1 × 1 seed. The results of each branch will be concatenated and entered as input to the next block. GoogLeNet is created by nesting nine Inception modules in the selected location. The highest aggregated layer will be placed between the initial modules. To reduce the dimension of the map, the GoogLeNet property worth observing is the combination of supplementary classifiers. Based on the assumption that the middle class of CNN should establish classification features, the author added simple classifiers. (Two fully connected and softmax layers) that work with properties created by the midpoint of the network Losses calculated by these classifier decisions will be used during the rear propagation process to calculate additional gradients that support the training of the relevant layers. At the time of inference, the extra classifier was canceled.

In the following publications, [18] a modified version of the registration module was presented with a slightly modified network architecture. The author proposed Batch Normalization (BN) and included in the Inception BN network. Techniques that normalize parts of an architectural model The author claims that BN helps with higher learning rates and easier startup techniques without experiencing side effects. According to BN, all images of the current mini-batch will be reduced in size so that Has an average of 0 and a variance of 1, so linear transformation is a parameter that is learned through the training process The network used in [18] is Inception-v2, a small modification to GoogLeNet. In addition to the integration of BN, the most significant change is the 5x5 layer of the registration module being replaced by two consecutive sequences. Limitations of this method is eliminated by way of combing neighboring tiles with the use of bilinear interpolation. Authos has referred to consequences of test to enhance digital mammogram using CLAHE method.

The remaining networks (ResNets) [19] consist of layers. convolutional Which has been reformed, which is currently learning the rest of the function with reference to input The author confirms that this type of network is easy to adjust and can increase the depth greatly. Operation of The "remaining blocks" as described in [19] are straightforward: for every Convolutional layer, a "Shortcut Connection" layer will be added. Layer results convolutional Will be added to the output of the branch shortcut, and the results will be propagated to subsequent blocks (Fig. 4). In addition to using the shortcut connection, the network architecture is inspired by the philosophy of the VGG network. Is essentially the Convolutional layer. All have a small seed size of 3 × 3 and follow two simple design rules: (i) for the size of the map, the output characteristics are the same (ii) when the feature mapping size is reduced by half (With the dual layer of stride 2), the number of filters is doubled to maintain the complexity of time per layer. The author tests architecture with different depths in the range of 34 to 152 layers.

Author	**Summary**	**Dataset Used**	**Key Index Parameter**
A. Oliver, J. Freixenet, J. Martí, and E. Pérez	The performance of seven mass detection methods is compared using two different mammographic databases: a public digitised database and a local full-field digital database.	Breast mass Dataset Rio Medical	• FROC analysis • Receiver Operating Characteristic (ROC)
J. A. Schnabel, M. L. Giger, and N. Karssemeijer	Author has presented present a review of the current status of these task-based image analysis methods, which are being developed for the various image acquisition modalities of mammography, tomosynthesis, computed tomography, ultrasound, and magnetic resonance imaging.	Optical Breast Images	• Accuracy • Precision
J. A. Schnabel, M. L. Giger, and N. Karssemeijer	Developed a methodology to enable quantitative characterization of features in a suspicious region or tumor (those describing morphology or function. It can be to enable exploration of the complex relationships among image-based	used computer-extracted features of the tumor to assess its aggressiveness, as a ‘ virtual biopsy,’ which could be used with clinical biomarkers to	• Specificity • Accuracy

	tumor characteristics across large populations	help determine patient management.	
Deepa S. Deshpande, Archana M. Rajurkar and Ramchandra R. Manthalkar	Author has proposed an automatic classification system for breast cancer using Texture Based Associative Classifier (TBAC).	Mammography Image Analysis Society (MIAS)	• Accuracy • Certainty Factor • Completeness
B.-W. Hong and B.-S. Sohn,	Author presents a novel method for the segmentation of regions of interest in mammograms. The algorithm concurrently delineates the boundaries of the breast boundary, the pectoral muscle, as well as dense regions that include candidate masses.	multiscale isocontour maps images	• Specificity • Accuracy

Table 1 Literature Summarization

III. Proposed System

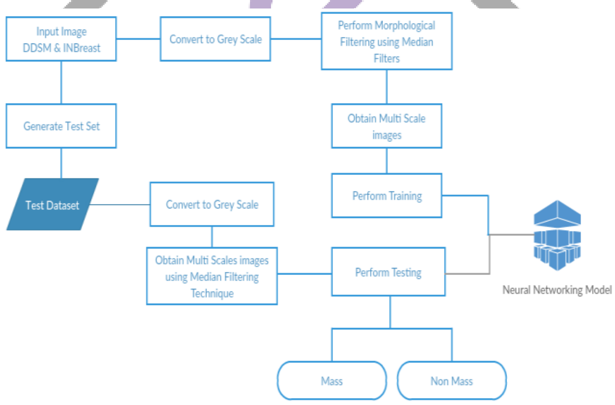

Figure Proposed System Flow

3.1 Data Augmentation

In deep learning techniques, NN models need to learn a lot of parameters. The chances of overfitting the training data are increased due to the complexity of the model. Adding data is an upright way to avoid this. [33] It creates new preview images using transformations such as flipping, rotating, and many other transformations with real data samples. For every image, we created a new preview image. Seven images using a 90-degree rotation and a combined 270 degrees and 270 degrees, so the resulting data set will have seven times more images than the original database

3.2. Enhancement of Digital Mammograms

Adaptive Histogram Equalization (CLAHE) [16] is used to increase the deterioration of sharpness in some mammogram images. The pixel intensity will be converted to value within the proportional display range of the pixel intensity rank in The CLAHE area intensity histogram is a special case of the Adaptive Histogram Equalization (AHE), in which images are adjusted by the user, determining the clip level is the local histogram height and the increase factor. The most obvious In this technique, improvements are made on very small discs, so the overhang due to noise or edge shadow effect is very low compared to AHE. The CLAHE method was originally developed to reduce the shadowing of edges and sounds that occur in homogeneous areas in the medical image. [18] This method is used for improving digital mammograms [1–7] and showing improvements. Good with the quality of mammograms images.

An input image I with dimensions $M \times N$, is divided into small blocks. CLAHE is then used to enhance the contrast of each block. Finally the bilinear interpolation is used to combine the neighboring blocks back into whole images. The steps in CLAHE are described as below.

(1) Images patches are divided into nonoverlapping blocks of size 8×8.

(2)The histogram of each block is calculated.

(3) For contrast enhancement of patches, a clip limit of histogram, $t = 0.001$, is set.

(4) After clipping the threshold value the histogram is redistributed.

(5) Every block histogram is modified by the following transformation function:

where pt(Ai) is the probability density function of the input patch image grayscale value at i and is define as
where mi is the gray scale value of input pixel i and m is the total number of pixels in a block.
(6) Bilinear interpolation is used to combine the neighboring blocks in each patch. The gray scale value of the patch is also changed according to the new histogram.

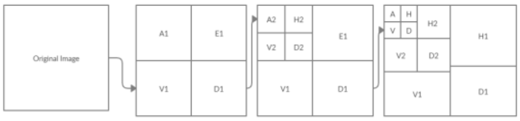

Figure 3

A two-dimensional DWT consists of down samplers and digital filter banks. The digital filter banks comprise low pass filter f(n) and high pass filter k(n). The number of banks depends upon desired resolution of the application [17]. As the mammogram images are two-dimensional signal, the DWT can be computed by separable wavelet functions. As shown in Figure 3, the columns and rows of the image are distinctly processed over the one-dimensional wavelet transform to establish the two-dimensional DWT. In frequency domain the enhanced image E is decomposed into subband images at resolution 2j+1. Ba is the approximation of the image. Bd, Bh, and Bv are three detailed subband images in diagonal, horizontal, and vertical, directions, respectively. As a result of wavelet decomposition the image I decomposed into four subband components like High-High (HH), High-Low (HL), Low-High (LH), and Low-Low (LL), which correspond to subimages that are Ba, Bd, Bv, and Bh, respectively.

3.3 Discrete Curvelet Transform

Discrete curvelet transform is an image representation technique used in computer vision. It was proposed by Candes and Donoho [21]. DCT codes image edges more efficiently than wavelet transform and it has useful geometric features that can be used as a feature vector in medical image processing. Eltoukhy et al.[22] have used DCT for the mammogram images. Let L be a function that has a discontinuity across a curve and is smooth otherwise, and consider approximating L from the best n-terms in the expansion.

In the next step we use CNN to learn features from the data set matrix M˜. CNN has proved its importance in classification of images by its significance results. CNN has a multilayered architecture, consisting of a convolution layer followed by a maximum pooling layer. The number of layers depends upon the designer. The output of final maximum pooling layer is fed to a fully connected layer that works like MLP which is further forwarded to softmax layer. The pooling layer is used for dimensionality reduction in the convolution layer. Mostly used pooling layer algorithms are average pooling, mean pooling, and maximum pooling. During the training, the dropout algorithm will be applied by randomly disabling the neurons, with a normally dropout ratio between 0.3 and 0.6. The final layer of CNN is a soft max layer that contains the output neuron according to the number of classes of the problem, which is assigned a confidence score. The two convolutional and max pooling layers will be used with a kernel size of 2 × 2. Convolutional layers have 16 kernels with size of 7 × 7 and the second layer uses kernel sized 5 × 5. Then, a fully connected neural layer is used. The dropout ratio in the experiment is 0.55. Softmax layer is used to train CNN for classification..

IV. CONCLUSION

B This evaluation will be very helpful for the medical people in detecting tumor in breast. Also we have discussed a methodology that can help rural people to find out the tumor occurrence in mammogram image. The system has potential of improving physician diagnostic performance mass detection systems and segmentation according to the division of multiple scales and randomly stacked forests were proposed and evaluated in public data sets by combining morphological filtration and grouping, the system can produce relatively accurate division of lesions..

REFERENCES

[1] A. C. SOCIETY, "GLOBAL CANCER FACTS & FIGURES, 3RD EDITION," AMERICAN CANCER SOCIETY ATLANTA,GEORGIA2015.

[2] A. Oliver, J. Freixenet, J. Martí, and E. Pérez, "A review of automatic mass detection and segmentation in mammographic images," *Medical image analysis*, vol. 14, no. 2, pp. 87-110, 2010.

[3] J. A. Schnabel, M. L. Giger, and N. Karssemeijer, "Breast Image Analysis for Risk Assessment, Detection, Diagnosis, and Treatment of Cancer," *Annual Review of Biomedical Engineering*, vol. 15, pp. 327-357, 2013.

[4] M. Gromet, "Comparison of computer-aided detection to double reading of screening mammograms: review of 231,221 mammograms," *American Journal of Roentgenology*, vol. 190, no. 4, pp. 854-859, 2008.

[5] B.-W. Hong and B.-S. Sohn, "Segmentation of regions of interest in mammograms in a topographic approach," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 1, pp. 129-139, 2010.

[6] N. Dhungel, G. Carneiro, and A. P. Bradley, "Automated Mass Detection in Mammograms using Cascaded Deep Learning and Random Forests," in *Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on*, 2015, pp. 1-8: IEEE.

[7] L. Liu, J. Li, and Y. Wang, "Breast mass detection with kernelized supervised hashing," in *2015 8th International Conference on Biomedical Engineering and Informatics (BMEI)*, 2015, pp. 79-84: IEEE.

[8] P. Kang and S. Cho, "EUS SVMs: Ensemble of under-sampled SVMs for data imbalance problems," in *International Conference on Neural Information Processing*, 2006, pp. 837-846: Springer.

[9] H. He and Y. Ma, *Imbalanced learning: foundations, algorithms, and applications*. John Wiley & Sons, 2013.

[10] A. Bria, N. Karssemeijer, and F. Tortorella, "Learning from unbalanced data: a cascade-based approach for detecting clustered microcalcifications," *Medical image analysis*, vol. 18, no. 2, pp. 241-252, 2014.

[11] S. N. Murthy, A. Kumar, and H. Sheshadri, "Mass Detection and Classification using Machine Learning Techniques in Digital Mammograms," *International Journal of Computer Applications*, vol. 76, no. 1, 2013.

[12] E. Kozegar, M. Soryani, B. Minaei, and I. Domingues, "Assessment of a novel mass detection algorithm in mammograms," *Journal of cancer research and therapeutics*, vol. 9, no. 4, p. 592, 2013.

[13] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: toward a full-field digital mammographic database," *Academic radiology*, vol. 19, no. 2, pp. 236-248, 2012.

[14] K. Bowyer *et al.*, "The digital database for screening mammography," in *Third international workshop on digital mammography*, 1996, vol. 58, p. 27.

[15] *DoD BCRP Spiculated Mass Detection Evaluation Data*. Available: http://marathon.csee.usf.edu/Mammography/DDSM/BCRP/bcrp_mass_01.html

[16] H. Li, Y. Wang, K. R. Liu, S.-C. Lo, and M. T. Freedman, "Computerized radiographic mass detection. I. Lesion site selection by morphological enhancement and contextual segmentation," *IEEE Transactions on Medical Imaging*, vol. 20, no. 4, pp. 289-301, 2001.

[17] Y. Wang, "Hierarchical Masses Detection Algorithms Based on SVM in Mammograms," Master's Thesis, Xidian University, China, 2006.

[18] J. Chu, H. Min, L. Liu, and W. Lu, "A novel computer aided breast mass detection scheme based on morphological enhancement and SLIC superpixel segmentation," *Medical physics*, vol. 42, no. 7, pp. 3859-3869, 2015.

[19] J. Serra, *Image analysis and mathematical morphology, v. 1*. Academic press, 1982.

[20] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274-2282, 2012.

[21] D. L. Donoho, "Orthonormal ridgelets and linear singularities*," tech. rep., Department of Statistics, Stanford University, 1998. Submitted for publication.*

[22] Eltoukhy MM, Faye I, Samir BB. *"A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram." Comp Biol Med 2010; 40: 384-391*