Bangladesh University of Business and Technology (BUBT)

Course No. : **CSE - 465**

Course Title: **Machine Learning**

Semester: **Summer 2022 [Tri-Semester]**

Survey Paper Name: BreastNet18: A High Accuracy, Fine-Tuned VGG16 Model Evaluated Using an Ablation Study for Diagnosing Breast Cancer from Enhanced Mammography Images.

Submitted by -

Pallab Majumdar	(18192103050)
Al Ahad Sufian	(18192103056)
Joy Adhikary	(18192103062)
Mafuja Akter Mitu	(18192103068)
Habibullah	(18192103080)

Submitted to -

Khan Md. Hasib

Faculty Member, Department of CSE, BUBT

Summary of the Paper:

For the detection and treatment of breast cancer, mammography has become the most widely used imaging technique. Mammograms are used to detect cancerous lesions by examining the shape, size, density, and distortions of the mass as well as the size and location of the actual lesion.

A CBIS-DDSM dataset contains artifacts and noise in mammograms. This study proposes a fully automated and reliable deep learning model based on transfer learning and ablation studies, BreastNet18. To develop the model, the performance of six existing pre-trained and fine-tuned TL architectures, namely VGG16, VGG19, MobileNetV2, ResNet50, DenseNet201, and InceptionV3 is compared to determine which model reaches the highest accuracy level.

The aim of this study is to detect and classify breast cancer into benign calcifications (BC), benign masses (BM), malignant calcifications (MC), and malignant masses (MM) at an early stage in order to reduce the risk of death. Afterward, six fine-tuned convolution neural networks (CNN) were evaluated, and VGG16 yielded the highest performance. The model's accuracy is 98.02%. Results suggest that image pre-processing techniques, augmentation, and ablation studies can improve accuracy.

K-fold cross-validation is performed in order to investigate overfitting issues. A dataset of 11,536 pictures was produced by adding seven augmentation techniques, a total of 1442 preprocessed mammograms. With a training accuracy of 96.72%, validation accuracy of 97.91%, and test accuracy of 98.02%, the proposed BreastNet18 model fared best. Test accuracy was 96.24% for VGGNet19, 77.84% for MobileNetV2, 79.98% for ResNet50, 86.92% for DenseNet201, and 76.87% for InceptionV3.

Unique Contribution of the Paper:

Although using the DDSM for method evaluation presents significant challenges due to its size and other unique characteristics. DDSM is a database of 2,620 scanned film mammography studies. Contains normal, benign, and malignant cases with verified pathology information. The scope of the database together with the ground truth validation makes DDSM a useful tool in the development and testing of decision support systems despite the fact that the images are scanned as a film instead of full-field digital mammograms. Currently, there is no publicly available mammography database of this size.

- One of the main contributions of this study is the extensive examination of several standard image preprocessing techniques described in the literature for mammogram classification.

- Image processing algorithms namely binary masking, morphological opening, largest contour detection, in operation Range, Gabor filter, Gamma correction, CLAHE, and Green fire blue filter are used with the most appropriate parameter values.

- This study also shows how the performance of DCNN can be increased by ablation studies. Ablation studies include experimenting with different dose sizes, layer flattening, loss functions, optimizers, and learning rates.

- Although the loss-accuracy curve for the best-performing model confirms the absence of overfitting, we performed k-fold cross-validation with k values of 5 and 10 for a careful check.

- The most important contribution is the BreastNet18 model which is proposed to employ VGG16 as a foundational base, as VGG16 yields the highest accuracy.

- This resource will contribute to the advancement of mammography decision support systems research and provide standardized mammography data.

How the proposed model works in the paper:

Reference Model:

To find the best transfer learning model for the author's classification problem, they experimented with a total of six state of the art models to determine the optimal network based on accuracy. During the training phase, only the weights of the last FC layer of every model were modified and updated. They have created a transfer learning method by adding one flattened and dense layer at the bottom of the pre-trained models. The models are fine-tuned by adding a flattened layer right before the FC layers and changing the output layer with four neurons. The performance of a fine-tuned network is not much reduced even when the size of training data is quite small. Medical images may greatly differ from natural images , but pre-trained knowledge can be utilized to attain significant improvement in the model's performance as the pre-trained model's weight parameters have already been trained to detect primal features such as edges, corners, and textures. This method has proven to be very effective in dealing with small data sets.For pre-training, the weights of the ImageNet dataset are used. ImageNet is a wide-ranging dataset consisting of over 14 million 256 × 256 labeled natural images in more than 1000 classes which is publicly available.A total of six pre-trained models: InceptionV3, MobileNetV2, ResNet50, DeneNet201, VGG16, and VGG19 are trained on the training dataset and evaluated using the testing dataset to find the best performing model among them.MobileNetV2 consists of two types of blocks and each block contains three layers. In both blocks, the first and third layers are 1×1 convolutional layers with 32 filters and the second layer is a depthwise convolutional layer.Where ResNet50 is made up of a total of 48 convolutional layers along with a max pool layer and an average pool layer.InceptionV3 focuses on reducing the required computing power by altering prior Inception designs.In the case of DenseNet201, each layer accepts the feature maps of the previous layer as input, and the feature maps of the subsequent levels convey all the data by explicitly connecting all the layers in the

network.VGG16 model achieved 92.7% top 5 test accuracy in the ImageNet dataset and it won the Large-Scale Visual Recognition Challenge (ILSVRC) competition proposed by the Oxford Visual Geometry Group. Our authors, therefore, propose a model, called BreastNet18, which is built upon a fine-tuned VGG16 architecture where several experiments are carried out with our dataset to evaluate the performance.

Architecture Of Proposed Model

Our authors modified model BreastNet18 is generated by adding a flattened and dense layer after the fifth block of VGG16. By default, the input layer of the architecture requires that the size of the image, which is an RGB image, is 224 × 224 × 3, our input dimension for the first convolutional layer is 224 × 224 × 3. In this context, the first block contains two convolutional layers with 64 channels of 3 × 3 kernel size and the same padding followed by a 2 × 2 Maxpooling layer of stride 2 × 2. Similarly, the second block contains two convolution layers of 128 channels, with a kernel size of 3 × 3, followed by a Max pooling layer of stride 2 × 2 as in block one. The last three blocks contain three convolutional layers followed by a Maxpooling layer. The channel sizes of the three convolutional layers in blocks 3, 4, and 5 are respectively 256, 512, and 512, all having the same kernel size of 3×3. The initial input image is shrunken to half the size in each Maxpooling layer. After the stack of convolutional and Maxpooling layers, the outputted feature map from the last Maxpooling layer is of size 7 × 7 × 512. A flattened layer has been added to make a 1 × 25,088 feature vector. A dense layer has also been added which outputs four channels for the four classes. There is a Softmax activation function at the end which normalizes the classification vector obtained from FC. This is how the proposed model works

Advantages of the paper:

- Their proposed approach based on image processing, transfer learning, fine-tuning, and ablation studies has demonstrated a high correct breast cancer classification while dealing with a limited number of complex medical images.

- The proposed BreastNet18 model performed best with a training accuracy of 96.72%, a validating accuracy of 97.91%, and a test accuracy of 98.02%.

- Early detection of breast cancer and the classification of mammogram images are major areas of research. A recent study found that a pre-trained model provides a better performance on a DDSM dataset. Other authors combined various convolution and capsule features in a feature fusion method to derive an updated feature set. Mammographic mass detection is one of the most important areas of Computer Aided Diagnosis (CAD) and can be achieved by using DCNN. Hamid et al. proposed an architecture based on VGG16 and VGG19 models for the automatic classification of breast cancer from histopathology images.

- VGG16, VGG19, and ResNet50 are pre-trained networks known for their wide and deep architectures. They recorded an accuracy of 94.55% using a DenseNet-II classifier. Another new CAD system for the classification of breast masses was developed by Ragab et al.

- Paquin et al. developed a CNN model using mammograms on a simplified feature learning and fine-tuned network to distinguish cancer from normal mammograms. Their proposed architecture performed best on the DDSM. mammogram dataset by attaining accuracy, sensitivity, specificity, and precision of 92.84%, 95.30%, and 96.72%. To classify normal and abnormal mammograms, Cardezi et al. introduced a VGG-16 deep learning model with the Image Retrieval in Medical Applications (IRMA) dataset. Their proposed model resulted in an accuracy of 81.83% and 83.74% for the first and the second approaches, respectively.

Disadvantages of the paper:

- The proposed approach is carried out on a dataset containing a relatively small number of images. Though the number of images is increased by applying different data augmentation techniques, the performance of our proposed network could be further assessed using a larger dataset.

- It recorded a test accuracy of 98.02%. Fewer than 2% of the images are misclassified. From the confusion matrix, it is observed that among 2306 mammograms, a total of 48 mammograms from all four classes are misclassified. Among them, the highest correct predictions are achieved in the Benign Mass (BM) class, where only 5 images are misclassified. For the Malignant Mass (MM), a total of 11 images are misclassified, and for the Malignant Calc (MC), a total of 14 images are misclassified. The highest misclassification rate is noticed for Benign Calc, with a total of 18 misclassified mammograms.

- Some of the images of Benign Mass have similarities with Benign Calcification. On the other hand, the model sometimes classifies Benign Calcification images as Benign Mass or Malignant Mass as some of the images from these two classes have a strong resemblance to the Benign Calcification images. The opposite effect can be observed with malignant calcification and malignant mass images.

- Another reason could be that there is limited data diversity and the number of images. Though data augmentation increases the number of images, the model may not learn enough new features due to the lack of further data.

Conclusion:

The proposed approach based on image processing, transfer learning, fine-tuning, and ablation study has demonstrated a high degree of correct breast cancer classification while dealing with a limited number of complex medical images. In this study, a breast CAD system is proposed to classify mammograms into four classes. Images used in this experiment contained noise and artifacts, which were removed by using image pre-processing techniques, k fold cross-validation.

Six pre-trained and fine-tuned DCNN models were experimented with to determine which model performs with the highest accuracy. A BreastNet18 model is proposed, employing VGG16 as its foundational base, as VGG16 yields the highest accuracy.

Our proposed BreastNet18 model performed best with optimizer Adam and a learning rate of 0.0008, resulting in a training accuracy of 96.72%, a validation accuracy of 97.91%, and a test accuracy of 98.02%.

Transfer learning can be an appropriate approach in computer vision when working with a small number of images. To sustain the robustness of the model. The proposed approach may aid clinical specialists in diagnosing and treatment planning at an early stage.