



Bangladesh University of Business and Technology (BUBT)

Course No. : **CSE - 465**

Course Title: **Machine Learning**

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

Survey Paper Name: A robust shallow convolutional neural network model performing ablation study on medical images assessing the effectiveness of applying optimal data augmentation technique.

Submitted by -

Submitted to -

Name: Md Shamiul Islam

ID: 18192103241

Name: Julker Naeem

ID: 18192103227

Name: Al Shah Riar Emon

ID: 18192103229

Name: Sheikh Rahmatulla
Sakib

ID: 18192103099

Name: Kamrun nahar sara

ID: 18192103127

Name: Rifatul Islam

ID: 18192103187

Khan Md. Hasib

Faculty Member

Department of CSE, BUBT

Summary of the Paper:

CAD interpretation is difficult due to the intricate structure of malignant lesions in multiple imaging modalities, similarity across interclasses, distinctive characteristics in intra-classes, a lack of medical data, and artifacts and noises. This paper overcomes these issues by building an ideal shallow CNN model, running ablation tests by altering the layer structure and hyper-parameters, and using an appropriate augmentation methodology. MNet-10, a low-complexity model, performs well on eight medical datasets with different modalities. Enhancement also uses photometry and geometry. Since mammograms strain patients the greatest, we used them for the ablation inquiry. Before modeling, the two strategies improve the dataset. First, a basic CNN model is built and deployed on enhanced and non-augmented mammography datasets. The photometric dataset was precise. Ablation defines the model's architecture and hyper-parameters using mammography photometric data. After training the model with the last seven datasets, network robustness and augmentation are analyzed. 97.34% accuracy on mammogram, 98.43% on skin cancer, 99.54% on brain tumor MRI, 97.29% on COVID chest X-ray, 96.31% on tympanic membrane, 99.82% on chest CT scan, 98.75% on breast cancer ultrasound datasets by photometric augmentation, and 96.76% on breast cancer microscopic biopsy dataset by geometric augmentation. Various methods of elastic deformation augmentation are investigated and compared using the model and datasets. VGG16, InceptionV3, and ResNet50 were trained on the highest-performing enhanced datasets. The findings may assist future researchers analyze augmentation and ablation data.

Unique Contribution of the Paper:

A shallow CNN model with the same parameters performs well on all eight datasets. We think the mammography dataset, one of the most complex imaging modalities, is the ideal way to do this (86). Ablation studies created MNet-10. Ablation improved mammography model classification from 89.55 to 97.34% (Figure 7). The seven remaining datasets are used to train an ideal model. The model's accuracy for multiple datasets is above 96%. Even without fine-tuning parameters with other datasets, extensive ablation trials employing the most demanding imaging modality can achieve optimal performance for all datasets. Intensive ablation studies on mammograms allowed the model to learn the smallest, most complex elements, leading to greater performance on datasets with less complex regions of interest (ear infection and skin cancer datasets). The literature review in Section "Dataset description" mentions trials to construct a model or preprocess a dataset, but not augmentation procedures. No work has evaluated CNN model performance using sickness and imaging benchmark datasets. Which imaging modality needs what enhancement strategy? Hue and saturation cannot be used to enhance grayscale photographs, despite being frequent. Table 1 shows that these approaches may impact critical RGB pixel information and yield unsatisfactory results. PSNR measurements reveal that fresh enhanced images generated using our chosen processes do not significantly change the original image's pixel intensity level. An optimum CNN has a shallow design, little training data, and low processing cost. Ablation studies set model parameters. Annotated medical datasets are usually too small to train a CNN. Data augmentation is used to enhance images. We obtained different picture enhancing results using eight datasets. An incorrect algorithm can lead to inaccurate medical image interpretation. The dataset should be tested before adopting any data augmentation strategy for medical images. This research suggests that a shallow CNN model combined with data augmentation improves medical image analysis. The shallow architecture increases accuracy to 97.34% and decreases parameters to 10 million. Augmented datasets perform better across all modalities. 66–88% accuracy for non-augmented datasets. Depending on augmentation approach, accuracy can reach 96–99%. Depending on data augmentation technique, accuracy varies between 3–7%. Data augmentation and

shallow networks aid with limited images, while shallow design decreases training time and complexity.

How the proposed model works in the paper:

In this part, the MNet-10 model is compared against VGG16, ResNet50, and Inception V3 on the best augmented datasets. They chose three models based on recent research on similar medical datasets. CNN models like VGG16 and ResNet (72), the most common transfer learning models for interpreting medical pictures, assist smart medicine's progress (73). These models can be used on datasets (40, 41, 72, 74, 75). Similar datasets (76–78) have utilized InceptionV3. These classic models are well-established and have been shown useful in many research studies. Due to their diverse CNN architectures (23 to 143 million parameters), these models perform differently with short and large medical datasets. As three state-of-the-art models, they can show their raw performance on eight medical datasets and compare to MNet-10. For comparison, they picked VGG16, ResNet50, and InceptionV3. Their suggested MNet-10 model is trained for 100 epochs using Nadam, a learning rate of 0.0007, and a batch size of 32. Mnet-10 outperforms the other three models across all medical datasets. Some datasets do well with VGG16, Inception V3, and ResNet50, while others perform poorly. VGG16 outperformed ResNet50 and InceptionV3 on datasets with small, complicated ROIs, such as mammography, COVID chest X-ray, brain tumor MRI, and chest CT scan datasets . InceptionV3 outperformed VGG16 in datasets with large ROIs, such as skin cancer, tympanic membrane, and breast cancer microscopic biopsy datasets. ResNet50's accuracy was below 80% in most cases. MNet-10 seems to beat all the models in terms of F1 score, specificity, and AUC values . The three CNN models (VGG16, Inception V3, and ResNet50) did not yield consistent F1 score, specificity, and AUC across all datasets. ResNet50 also lagged behind VGG16 and InceptionV3. This strengthens the model. A Wilcoxon signed-rank test (80) is also used to compare the proposed network's findings to those of other models in . A P-value below 0.05 is significant (81). Table 7 shows Wilcoxon signed-rank test results for F1 scores (Table 6). This test yields a P-value of 0.003 in all cases (Table 7), indicating that the performance difference between MNet-10 and other DL models is statistically significant. MNet-10 has 10 layers and six weighted layers and is a shallow CNN model with 10 million parameters; VGG16, InceptionV3, and ResNet50 have 143, 23, and 25 million parameters, which is high for real-world data. ResNet-50 overfits smaller datasets

(82). Deep CNNs are easier to train than models with more trainable parameters. Keeping this in mind, the number of model layers is kept low to reduce the number of trainable parameters for improved generalization on a short dataset. Ablation studies determine the ideal number of convolutional layers (four). MNet-10 uses PReLU instead of ReLU for fast convergence (83) and higher overall performance. Faster convergence enhances classifier performance and reduces computing complexity. Minor convolutional kernels can extract more low-level textural information and small details from datasets with tiny details. Complex ROI datasets (mammogram, chest X-ray, chest CT scan, Brain tumor MRI) benefit from MNet10's 3 3 filter. This boosts overall performance not just in small ROI datasets but also in large ROI datasets (Tympanic membrane and Skin cancer datasets), which improves the model's generalization across many datasets. Multiple FC layers can cause overfitting (84) for dense connections in MNet-10 (85). To prevent overfitting, a dropout layer randomly eliminates FC layer connections (85) utilized for feature generalization. Their suggested MNet-10 model has stable performance across all eight medical imaging modalities, with accuracies between 96 and 99.6%, adding to its effectiveness, consistency, and stability.

Advantages of the paper:

- Information and semantics of a picture can greatly vary with different images. In several cases, variability in shape, size, characteristics even with in the same modalities causes diagnostic challenges even to medical experts. Often the intensity range of cancerous region maybe similar to surrounding health tissues. So, this deep learning model gives us the solution of accuracy in detection complex structure of medical diseases and eliminates confusions of medical experts.
- MNet-10 using two types of augmentation technique. One is photometric, another one is geometric. Using photometric augmentations, changes pixel illumination, intensity, pigment and obviously geometry unaffected also benefited by detect complex structure, hidden characteristics. Because of these techniques, classification of various dataset images gives us the better accuracy and also images geometrical location and orientation of an irregular region is much clear. The traditional photometric and geometric augmentation

techniques seemed to outperform the elastic augmentation technique in most cases

- Human eye often cannot detect the loss of necessary pixels of images in medical datasets. That's why a solution is to derive peak signal-to-noise ratios (PSNR) for all augmentation as an effective quality measure comparing the original and transformed image.
- All the convolutional layers are equipped with the PReLU non-linear activation function where PReLU is 2 adjustable (parameterizable) linear functions within 2 different ranges joined together, while ReLU is just a single adjustable (parameterizable) linear function within half that range, so you require minimum 2 ReLU's to approximate a PReLU. After ward using softmax activation function as the final layer gives us the best output.
- The training curve converges smoothly from the first to the last epoch showing approximately no bumps and no evidence of overfitting is found.

Disadvantages of the paper:

- Even after using the data augmentation technique four to five times, even for small datasets with few samples, the dataset is still insufficient to train a large number of DCNN model parameters. Also, even though a lot of augmentation techniques are used to increase the number of images in small datasets to meet DCNN's minimum requirement, the number of original samples is still not enough to achieve optimal performance. On the other hand, reducing the number of parameters by developing a compact CNN (shallow CNN) can lower the requirement for larger datasets.
- In geometric augmentation alternation may not possible clinically, which might be an obvious reason acquiring poor performance.
- Conventional CNN models tend to have deeper architectures resulting in too many parameters creating issues regarding overall performance and increasing time complexity.
- The test accuracies of MNet-10 for the datasets of breast mammogram, COVID chest X-ray, and chest cancer CT scan were obtained with the elastic deformation data augmentation technique at 84.45, 87.31, and 91.55%,

respectively. The photometric augmentation technique achieved the highest accuracy, 97%-99%, while this elastic augmentation's performance is significantly lower than the accuracies achieved by the geometric augmentation technique.

Conclusion:

The primary objective of this study is to create the best CNN classification model for multiple disease-related medical image datasets. MNet-10 outperformed with optimal accuracy for all eight datasets and provide us the higher accuracy comparing to other CNN models. The main reason of this optimized classification is balanced dataset and photometric augmentation which is helpful to increases our accuracy for the base model is about 89.55 to 97.34% and also the number of parameters decreased from 66 to 10 million. It is possible to draw the conclusion that shallow architecture significantly reduces training time and time complexity while data augmentation and a shallow network work together to deal with a limited number of images.