

Data Augmentation-aided Convolutional Neural Network for Detection of Abnormalities in Digital Mammography

Olaide N. Oyelade¹, Ahmed Aminu Sambo¹, Pam Bulus Dung², and Amina Hassan Abubakar¹
¹Department of Computer Science, Faculty of Physical Sciences, Ahmadu Bello University Zaria,
²Department of Computer Science, Federal College of Education, Pankshin
olaide_oyelade@yahoo.com, samboaminuahmed@yahoo.com, ahabubakar@abu.edu.ng,
victoryagain@yahoo.com

Abstract

Background: The use of data augmentation techniques to addressing the challenge of network overfitting and classification error is important in deep learning. Insufficient sample data for training have the tendency to bias the trained model so that it fails to generalize well. Several studies have proposed different augmentation techniques to solve this problem. But there are some peculiarities identified with the nature of datasets when applying augmentation methods. The subtle nature of some abnormalities in digital mammography often makes it difficult to transform such datasets into different form, while preserving the structure of the abnormality. **Aim:** To address this, this study aims to apply a combination of carefully selected data augmentation operations on digital mammography. **Method:** First, a convolutional neural network (CNN) is proposed suitable for feature extraction and abnormality detection from image samples. Secondly, transform operations are applied to the image data to generate more samples with different abnormalities to augment original datasets. **Results:** Using the image samples from the Mammographic Image Analysis Society (MIAS) with regions of interests (ROIs), Digital Database for Screening Mammography, Curated Breast Imaging Subset (DDSM+CBIS), INbreast, and whole images from MIAS, experimentation was carried out. The selected augmentation operations were applied to the datasets. These datasets contain samples presenting both bilateral craniocaudal (CC) and mediolateral oblique (MLO) views. Performance evaluation of the approach proposed in this study showed that classification accuracy of 90.62% was obtained with reduce loss values. The outcome of the study demonstrates the need to consider the structural and textural orientation of image samples when applying augmentation techniques to reduce overfitting.

Keywords: Breast cancer, convolutional neural network, deep learning, data augmentation, mammography,

1. Introduction

Cancer remains one of the most fatal non-communicable diseases with breast cancer reporting one of the highest rates of case identification (Oyelade, Obiniyi, & Junaidu, 2020). Several studies have approached the solution to the need for early detection of the disease using computational solutions. Methods including rule-based (Macek, et al., 2021), machine learning (Mangukiya, Vaghani, & Savani, 2022), deep learning (Oyelade, 2022), and several computer aided detection (CAD) approaches (Hassan, Hamad, & Mahar, 2022) have been reported. These methods have reported improved accuracy and recall with some values of 78% and 7.7% have been reported for both accuracy and recall rate respectively in some CAD-based systems (Freer & Ulisse, 2001). However, recent studies have shown that deep learning methods present more computational power to detecting the disease in medical images and identifying even very subtle abnormalities (Olaide & Ezugwu, 2022). Furthermore, the outstanding performance of deep learning method has motivated the use of CNN architectures to addressing the problem of classification of digital images for reduced false positive rate.

Considering the performance of deep learning methods in feature extraction and classification, research in the domain have now been advanced to improve the architectural configuration of the models and to aid their efficiency

through supportive techniques. Methods such as image preprocessing, data augmentation, regularization techniques, and architectural optimization have been researched. These methods have proven to further enhanced the performance of deep learning models so that abnormality detection, localization, and visualization are added to the task of feature classification. However, it remains a challenge to ensure that overfitting and underfitting are eliminated when reporting the performance of these models. The use of data augmentation has proven to yield increasing classification result so that other related task follows. Classical and artificial augmentation techniques have been used in recent studies with the latter often associated with generative adversarial networks (GANs). Whereas the classical method synthesizes new samples using transformational algorithms, the GANs-based methods approach image synthesis using two deep learning model positioned in an adversarial condition to generate and discriminate fake samples from real samples. In this study, we aim to use the transformational method to achieve out image augmentation task.

The use of the data augmentation approach proposed in this study is motivated by the challenge of detecting abnormalities presenting subtle features, often making it difficult to detect their presence in digital images. In addition, we seek to address the limitation associated with the use of limited or biased image samples can lead to a model which fails to generalise given new samples (Oyelade & E.Ezugwu, 2021). Supported by the problem of limited data samples is the quality of the samples. Images containing noise and low contrast often reduce the capability of CNN model in detecting abnormality in the images. This study aims to address these problems by applying image augmentation to the datasets used for experimentation. The approach ensures that synthesised samples return with images with high contrast, with no structural defect sustained in the process. Meanwhile, we design a CNN architecture sufficient to understanding the classification problem associated with the datasets used in the study. The design yielded a good architecture whose hyperparameters were manually optimized and selected for configuration during training. Abnormalities including architectural distortion, macrocalcification, mass, and special asymmetry were classified in addition to benign and normal cases.

The remaining part of the paper is organized as follows: A review of related works is presented in Section 2; Section 3 presents the design of the proposed CNN model; In Section 4, results and discussion are presented; and conclusion on the findings of the study is done in Section 5.

2. Related Works

In this section, studies presenting the use of deep learning and data augmentation methods to addressing the challenge of classification of image samples are reviewed. For the use of data augmentation technique, we review studies which reported on using either the classical or GAN-based method. The presentation is made in chronological order.

Overfitting describes situations where a deep learning model suffers from poor learning of actual patterns due to memorization of certain aspect of a subset of data (training set). This usually results in poor performance of the learning model when another set of data with different pattern is supplied. The implication of this is that the deep learning model has not learnt to generalize. Training a CNN model requires a huge amount of data to overcome the problem of overfitting (Costa, et al., 2018), (Yamashita, Nishio, & Do, 2018). Other techniques for mitigating overfitting are regularization, batch normalization, reducing architecture complexity, increasing training data and use of transfer learning as in (Mendel, Li, Sheth, & Giger, 2018). We shall therefore review some literatures that applied the technique of data augmentation in tackling the challenge of overfitting.

In a study, authors applied both the classical-based and synthetic-based data augmentation techniques to help build a neural augmentation model – a model that uses a method to allow a neural net to learn augmentations that best improve the classifier (Perez & Wang, 2017). The classical data augmentation techniques adopted includes as cropping, rotating, and flipping input images, while the synthetic approach adopted GANs (CycleGAN) to generate

images of different styles. Although the research was based on the popular very large dataset known as ImageNet and MNIST, it was however artificially constrained to a small subset of the dataset. The approach of the GAN used to augment data was by transferring styles from images in the dataset to a fixed predetermined image such as Night/Day theme or Winter/Summer. Also, they proposed a different kind of augmentation where we combine neural nets that transfer style and classify so instead of standard augmentation tricks, the neural net learns augmentations that best reduce classification loss. The details of the dataset used are Tiny-imagenet-200 of 100k training, 10k validation, and 10k test images of dimensions 64x64x3, and the MNIST consists of 60k handwritten digits in the training set and 10k in the test set in grayscale with 10 classes with image dimensions of 28x28x1.

A study proposed a data augmentation approach applied to clinical image dataset to properly train a CNN (Costa, et al., 2018). The approach of increasing the data training set was based on using the classical standard techniques of image transformation. The image transformation operation was not carried out on whole images but on manually cropped ROIs from mammograms. Each cropped ROI was passed through vertical and/or horizontal flip; rotation at 90, 180, and 270 degrees; addition of Gaussian noise with zero mean and 0.02, 0.04 and 0.06 variance. The result of the data augmentation operation produced a new dataset consisting of 21600 samples of balanced classes. This data, in addition to original whole image data and the ROIs were used in training, validating, and testing the CNN (GoogleNet architecture) model and results using receiver operating characteristic analysis showed that with a very limited dataset GoogleNet model, adopted for the research, could be trained to detect architectural distortion in digital mammography with area under the curve (AUC) of 0.74.

Furtherance on the studies of using GAN and addressing the general limitation associated with it, a AugGAN is proposed by (Huang, et al., 2018). The aim of the study is to address the limitation of the basic GAN which are prone to fail at preserving image-objects and maintaining translation consistency, and reduces their practicality on tasks such as generating large-scale training data for different domains. The AugGAN architecture proposed in the research is a structure-aware image-to-image translation network, which is composed of encoders, generators, discriminators, and parsing nets in a unified framework. The authors argued that the proposed AugGAN network generates more visually plausible images compared to competing methods on different image-translation tasks. Their experimentation of the AugGAN was carried applied to training Faster-RCNN and YOLO using datasets generated from the image-translation. The result demonstrated significant improvement on the detection accuracies by using the proposed image-object preserving network - AugGAN.

The rapid improvement, high-resolution images and application of GAN-based models use to synthesize lesions onto healthy screening mammograms produces realistic images compared to the simple data augmentation techniques like horizontal flipping and others (Wu, Wu, Cox, & Lotter, 2018). This was achieved by training a class-conditional GAN to perform contextual in-filling to synthesize data used for experimentation with ResNet-50 classifier. The result of this augmented dataset to improve breast cancer classification performance produces a higher AUROC compared to the same model trained only on traditionally augmented data.

An application of generative adversarial network (GAN) to generate synthetic mammographic images from the digital database for screening mammography (DDSM) has been reported in (Guan & Loew, 2019). The research further extracted ROIs to train the GAN to generate new synthetic images. Experimentation was carried out to compare the performance of the GAN-based approach with the classical transformation augmentation methods, such as rotation, shifting, and scaling, by using them to train a CNN classifier from scratch. Results show that, to classify the normal ROIs and abnormal ROIs from DDSM, adding GAN-generated ROIs in the training data can help the classifier prevent overfitting. Furthermore, on validation accuracy, the GAN performs about 3.6% better than standard transformations on image augmentation. In another related work, authors designed MammoGAN to demonstrate that recent advances in generative adversarial networks (GANs) enable synthesis of realistic medical images that are hard to distinguish from real ones (Korkinof, Heindl, Rijken, Harvey, & Glocker, 2019). Their MammoGAN architecture was used to effectively synthesize mammograms at a resolution of 1280x1024 pixels, the highest so far.

Similarly, a study has also demonstrated the applicability of the classical image augmentation approach to the task of reducing overfitting in deep learning models. The study investigated different transformational operations in achieving image synthesis on-the-fly to augment real images during training. In a recent study, (Olaide & Ezugwu, 2022) proposed the application of GAN-based data-augmentation approach to experiment on the performance of a CNN architecture with a novel wavelet function. Using the combination of MIAS and DDSM+CBIS datasets, the study showed that the classification of the network model yielded better classification and recall rate as compared with using the model without the augmentation method.

This study therefore is aimed to advance research on using the classical approach with different transformational operations in synthesizing images for improved classification with deep learning model, emphasis on CNN. This approach and the proposed CNN model are presented in the next section.

3. Methodology

In this section, we present the design of the proposed CNN architecture. Also, discussion on approach for application of the architecture to the images from the four databases is presented.

The selection of the hyperparameters for the proposed CNN architectures was guided by the nature of the classification problem. This choice ranges from model configuration, the size of the dataset, depth of the architecture, filter size- filter count and other related hyperparameters. The Adam optimizer algorithm was applied for training the model in adjusting its weights while learning classification problem. The justification for the use of the optimizer is because it converges faster due to its use of both adaptive learning and momentum. We observed that using step decay rate while reducing learning rate by some percentage after a set number of training epochs increases the performance of characterization of abnormalities in mammograms.

The CNN architecture proposed in this paper consists of six (6) blocks of two (2) convolutional layers with each block appended with a max-pooling operation. Each convolutional layer is designed with a 3x3 filter and filter count of 32x32, 64x64, 128x128, 256x256, 512x512, and 1024x1024 for convo operation in each block in progressive manner up to the sixth block. Also, to regularize the model in achieving good performance, the use of dropout layer with drop rate of 0.5 was used at the end of each block. The fully connected layer is followed by a dense layer which applies a softmax function for the computation of the class probability scores, resulting in a volume of size [1 * 1 * number of classes].

We have chosen a 3x3 filter size because smaller filters collect as much local information as possible, while the bigger filters represent more global, high-level information. Our decision on this was reinforced by the common knowledge of the subtle nature in associated with some abnormalities such as asymmetry and architectural distortion in digital mammograms.

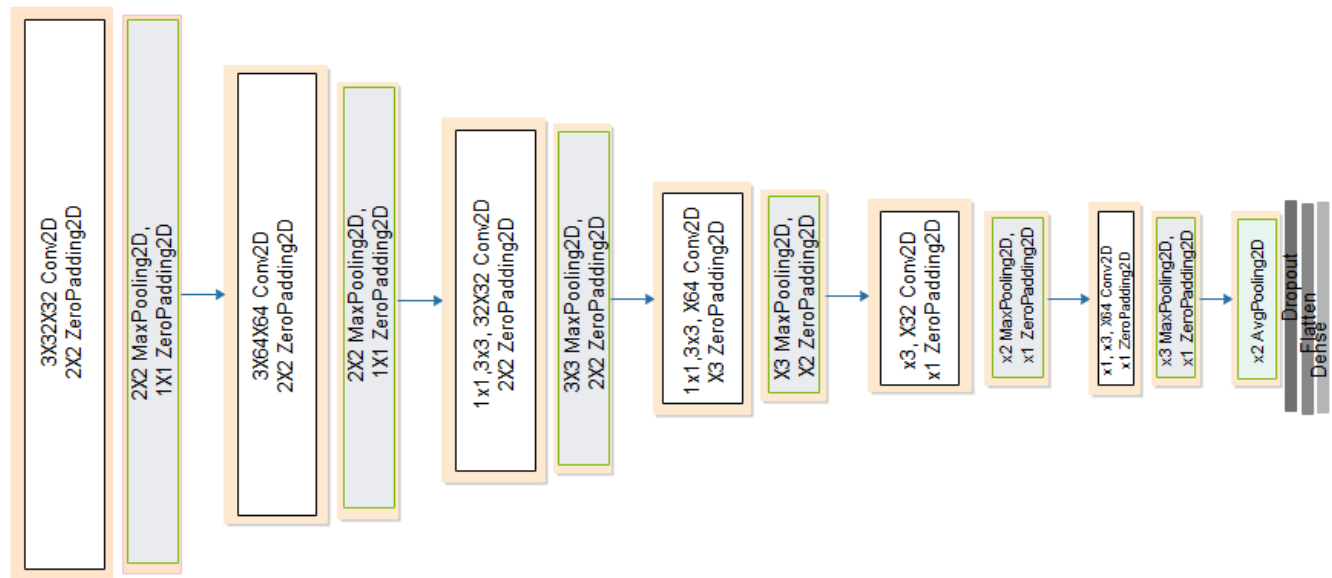


Figure 1: The proposed CNN architecture showing the input, blocks of convolution, max-pooling, fully connected layers and a three-class classifier

The architecture described in Figure 1 assumes the form of Conv-Conv-Pool-Drop-Conv-Conv-BatchNorm-Pool-Drop-Dense(relu)-BatchNorm-Drop, with number of filters modeled as 32(3, relu)-32(3, relu)-2(2)-64(3, relu)-64(3, relu). The main model is captured in Figure 3.

Inputs to the CNN model was first cleaned using some image pre-processing techniques. First, we applied to every sample from the four databases to denoising operation. This is to avoid interference which may arise from the noise thereby impairing on the performance of the CNN model. Secondly, segmentation technique was used to remove background area such as labels, artifacts, and pectoral muscle, tissues, and other muscles from the main breast tissue area. Again, this allows for elimination of background objects which are not required during the feature extraction and image classification procedure. In addition this paper uses a variant of adaptive histogram equalization (AHE), called contrast limited adaptive histogram equalization (CLAHE), to improve the contrast in samples.

Once the images were pre-processed through the using techniques, all samples were applied to train the CNN model in the category of their databases. We avoided combining the images into a pool of samples to investigate the performance of our model on each category of datasets. As a result, the findings obtain from this study will provide a guide to further studies on the variation of the performance of DL models when applied to each database. In the section following, we provide a detail discussion on the four databases used and the statistics of image samples obtained for the experimentation phase of the study.

4. Results and Discussion

In this section, the result obtained for the experimentation are discussed and presented. The experimentation was carried out using the Floyd Server. Training for the CNN model was done with twenty (20) epochs so that all datasets including those of MIAS, DDSM+CBIS, and INbreast are subjected to the same environmental configuration for the experiments. The best performing setup for the experiments showed a classification of accuracy of 90.02%.

In Figure 2 and 3, we present a graphical chart of the training process for accuracy and loss function values obtained over the 20 epochs.

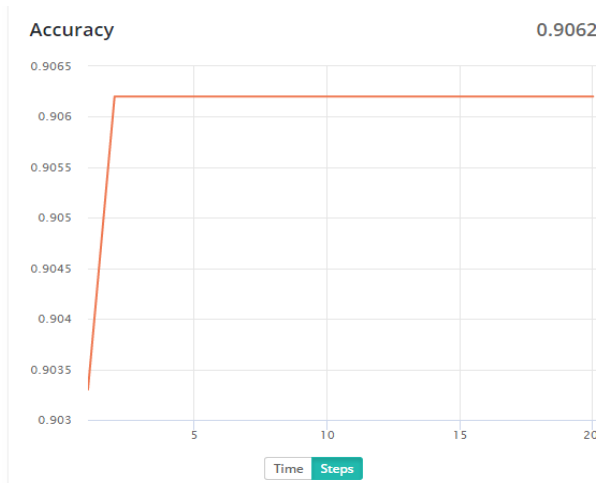


Figure 2. A graph for training accuracy using MIAS datasets

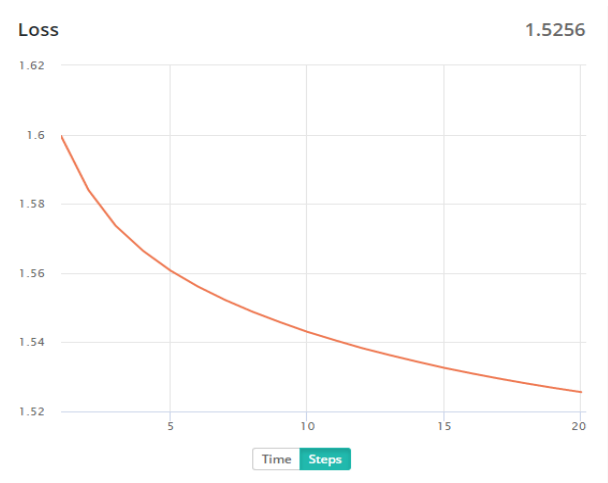


Figure 3. A graph showing the loss function values for training of MIAS datasets with image size 299x299

Furthermore, experimentation was done for the whole image datasets using MIAS database. The samples used in this second experiment differ from that reported in figure 2 and 3 which uses ROIs. Result obtained confirmed that the performance of the CNN model is 65% accuracy as compared to the 90% accuracy when ROIs samples were applied. The implication of this finding demonstrates the need to preprocess inputs to allow for only regions with susceptible abnormality are retained. This allows for the feature extraction process to eliminate false positives rates with regions presenting false abnormality. Meanwhile the input using the whole image samples are of size 1024X1024 pixel, and with classification accuracy of 64.29% and loss value of 5.9789. these results are shown in Figures 4 and 5, which presents the learning curve of the model with respect to the samples supplied as input.

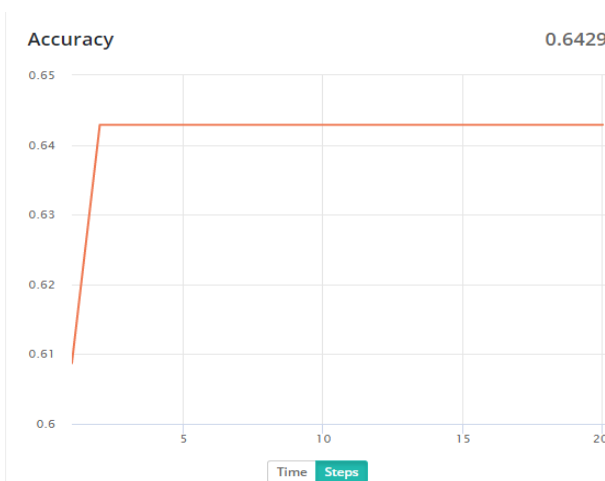


Figure 4. A graph for training accuracy using MIAS whole image dataset (1024X1024) dataset



Figure 5. A graph showing the loss function values for training of MIAS whole image dataset (1024X1024) dataset

In Figures 6, and 7, performance of the model on the DDSM+CBIs datasets are graphed to show how the model was able to learn the problem. The result obtained showed that although the loss function values progressively reduced during the learning phase, the classification accuracy performed well only after the model exceeded the 7th epoch. The staggering plunge in accuracy could be associated with the nature of the datasets which the model was able to appropriately understand the pattern of each class of samples reported in the dataset. Interestingly, we found that the performance of the model on this dataset showed something competitive with that ROI-based MIAS samples.

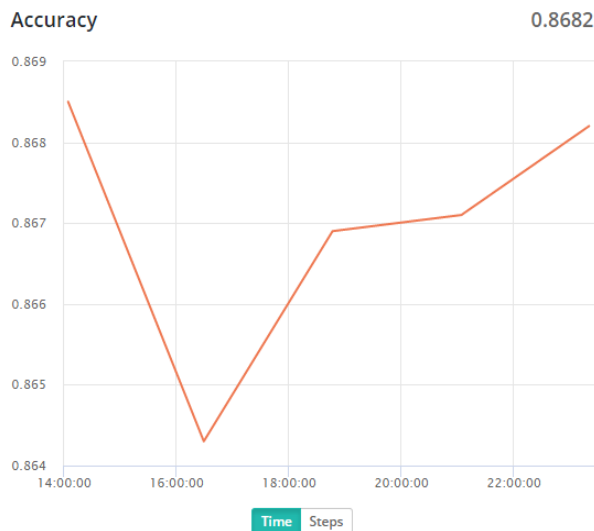


Figure 6. A graph for training accuracy using DDSM+CBIS image dataset (299x299) dataset

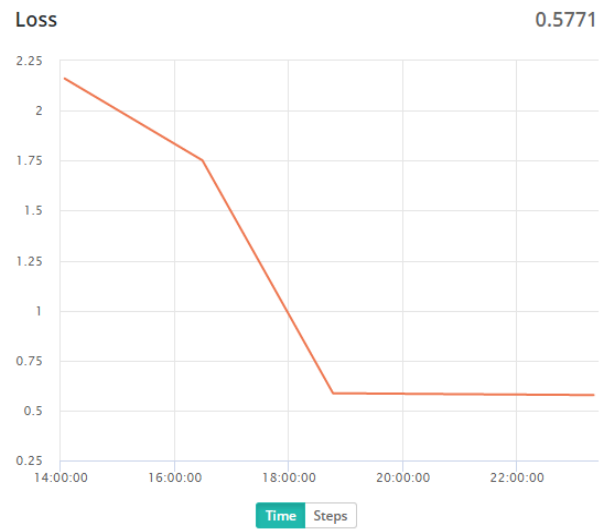


Figure 7. A graph showing the loss function values for training of DDSM+CBIS image dataset (299x299) dataset

We further investigated the performance of the proposed model on the INbreast dataset to know how the learning curve flows. Result obtained, which is captured in Figures 8 and 9 showed that the classification accuracy shares similarity with what is obtained when whole image size for MIAs were applied. Again, this further confirms that whole images are not suitable for shallow CNN models. And recall that the depth of CNN model in most cases do not guarantee improved performance. We therefore acknowledge that future studies might want to investigate the behavior of hyperparameters of the CNN architecture when some optimization of parameters is done. Whereas the classification accuracy reported during training lags behind those of MIAS and DDSM+CBIS datasets, we found the ability of the model to progressively learn the problem to be interesting as can be seen in the curve showing the loss function values.

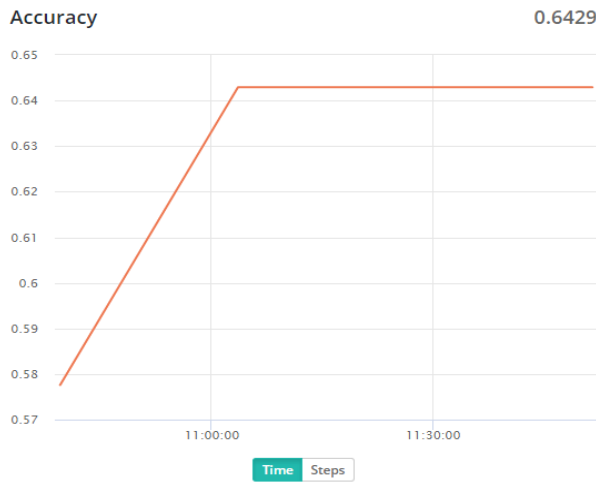


Figure 8. A graph for training accuracy using INBreast image dataset (2560x3328) dataset

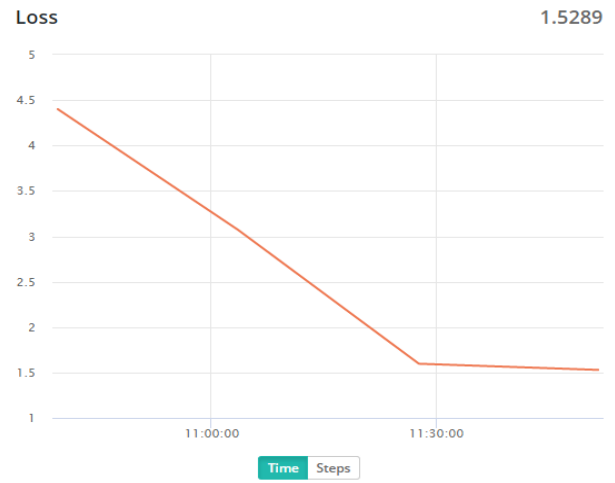


Figure 9. A graph showing the loss function values for training of INBreast image dataset (2560x3328) dataset

Table 1. Performance of The Proposed CNN Model on Four (4) Different Datasets

Model Description	Dataset	Accuracy
CNN model	MIAS	90.62%
	299X299 dataset	
	DDSM+CBS dataset	86.82%
CNN model + data augmentation	INbreast dataset	64.29%
	2560X3328	
	MIAS 1024X1024	64.29%

In Table 1, we present a summary of the performance of the proposed CNN architecture as applied to the four (4) different categories of sample images. The summary of the findings resulting from this study is as follows. Application well pre-processed images using state-of-the-art techniques proves beneficial to the classification of problem of detecting breast cancer in mammography images. Secondly, we discovered that reshaping input image in such a manner that allow for features to be dense within a smaller region presents a supportive mechanism for the CNN model to extract those features. The implication of this is that large-sized inputs may increase the computational cost on the CNN model and may impair on the process of accurately extracting features which will aid the classification process. Hence, we discovered that images resized to 299x299 yielded an impressive performance as compared with whole images.

Thirdly, ROIs-based input serves to improve performance than whole image. This is readily explainable from the results of DDSM+CBIS and that from the MIAS databases. The ROIs-based input provides the CNN model with a suggestive feature which support the process of extraction of such features and thereafter, classification. Whereas the whole images may contain unwanted regions which may interfere in the process of extraction of regions with features confirming abnormalities that is desired.

In addition, the structure of the CNN proposed in this study presents a foundation for modifying future CNN model aimed for solving the classification problem considered in this study. This is based on the performance obtained based on the choice and combination of hyperparameters used in the design of CNN architecture.

5. Conclusion

This paper proposes a CNN architecture which was applied to inputs images drawn from the MIAS, DDSM+CBIS, INBreast and Mini-MIAS databases. The result obtained demonstrates that the architecture successfully detected and classified the architectural distortion, asymmetry, mass, and micro-calcification abnormalities at an accuracy of 90.62%. the limitation of the study is discovered when whole images applied to the proposed CNN failed to yield an impressive result. As a result, in future, we propose that the samples drawn from the Mini-MIAS and INBreast datasets be cropped into smaller sizes of 299x299 so that only ROIs containing desired abnormalities are retained.

References

- Costa, A., Oliveira, H., Catani, J., Barros, N. d., Melo, C., & Vieira, M. (2018). Data augmentation for detection of architectural distortion in digital mammography using deep learning approach. *Computer Vision. Pattern Recognition*.
- Freer, T., & Ulissey, M. J. (2001). Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology*, 220(3), 781-786.
- Guan, S., & Loew, M. (2019). Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks. *J Med Imaging (Bellingham)*, 6(3).
- Hassan, N. M., Hamad, S., & Mahar, K. (2022). Mammogram breast cancer CAD systems for mass detection and classification: a review. *Multimedia Tools and Applications*, 20043–20075.
- Huang, S., Lin, C., Chen, S., Wu, Y., Hsu, P., & Lai, S. (2018). AugGAN: Cross Domain Adaptation withGAN-based Data Augmentation. *The ECCV 2018, by Computer Vision Foundation*.
- Korkinof, D., Heindl, A., Rijken, T., Harvey, H., & Glocker, B. (2019). MammoGAN:High-Resolution Synthesis of Realistic Mammograms. *Proceedings of Machine Learning Research*, 1–4.
- Macek, A. J., D.Kirschenbaum, J., J.Ricklan, S., Schreiber-Stainthorp, W., C.Omene, B., & Conderino, S. (2021). Validation of rule-based algorithms to determine colorectal, breast, and cervical cancer screening status using electronic health record data from an urban healthcare system in New York City. *Preventive Medicine Reports*.

- Mangukiya, M., Vaghani, A., & Savani, M. (2022). Breast Cancer Detection with Machine Learning. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*.
- Mendel, K., Li, H., Sheth, D., & Giger, M. (2018). Transfer Learning from Convolutional Neural Networks for Computer-Aided Diagnosis: A Comparison of Digital Breast Tomosynthesis and Full-Field Digital Mammography. *Academic Radiologists*.
- Olaide, O., & Ezugwu, A. E.-S. (2022). A novel wavelet decomposition and transformation convolutional neural network with data augmentation for breast cancer detection using digital mammogram. *Scientific Reports*.
- Oyelade. (2022). A comparative performance study of random-grid model for hyperparameters selection in detection of abnormalities in digital breast images. *Concurrency and Computation Practice and Experience*.
- Oyelade, O. N., & E.Ezugwu, A. (2021). A deep learning model using data augmentation for detection of architectural distortion in whole and patches of images. *Biomedical Signal Processing and Control*.
- Oyelade, O. N., Obiniyi, A., & Junaidu, S. (2020). ONCODIAG Select and Test (ST) Algorithm: An Approximate Clinical Reasoning Model for Diagnosing and Monitoring Breast Cancer. *Journal of Current Research in Bioinformatics (JCRB) - Science Publications, DOI: 10.3844/ajbsp, 1-17*.
- Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv:1712.04621*.
- Wu, E., Wu, K., Cox, D., & Lotter, W. (2018). Conditional Infilling GANs for Data Augmentation in Mammogram Classification. *MICCAI 2018, Breast Image Analysis Workshop*.
- Yamashita, R., Nishio, M., & Do, R. (2018). *Insights Imaging*. 9(611).