ISSN: 18158129 E-ISSN: 18151027
Volume 18, Issue 09, October, 2022

Improving Breast Mass Classification Performance by Dataset Enhancement with Deep Convolutional Generative Adversarial Network

Le Nhat Tan[1,2], Truong Thi Thanh Hang[1,2], Pham Tan Thi[1,2*]

Ho Chi Minh City University of Technology (HCMUT), 268 Ly Thuong Kiet, Ward 14, District 10, Ho Chi Minh City, Vietnam[1]
Vietnam National University Ho Chi Minh City (VNUHCM), Quarter 6, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam[2]

Corresponding author: 1,2*

ABSTRACT— Early detection and diagnosis of breast cancer is the key to controlling, curing this disease as well as reducing costs for the patient. AI-based computer-aided diagnosis (CAD) systems designed to help physicians make faster and more accurate decisions, convolutional neural network (CNN) has earned many achievements in the medical field. However, the performance of the CNN model is highly dependent on the quantity and quality of the input data sets, which is a big challenge for medical imaging because the image collection is very limited. To solve this problem, Deep Convolutional Generative Adversarial Network (DCGAN) is proposed to generate breast tumor images from the original breast ultrasound image BUSI dataset consisting of 437 benign masses and 210 malignant tumors. Then, in order to determine the performance of the proposed model, the newly created images combined with the images in the original data will be evaluated qualitatively, visually and quantitatively through the classification problem by feeding to 2 classification models: simple self-designed and Densenet to evaluate their quality and clinical value. The classification performance using only the original data yielded a sensitivity of 37% and an F1-score of 51% then increased to a sensitivity of 81% and an F1-score of 70%. The results show that breast tumor images generated from the DCGAN network can be used to significantly increase the efficiency thus has the potential to assist physicians in medical image reading task.

KEYWORDS: Data enhancement, Breast tumor classification, Deep convolutional Generative Adversarial Network, Neural Network.

1. INTRODUCTION

Breast cancer is a pathological condition in which mammary gland cells grow uncontrollably, producing malignant tumors, which have the ability to divide strongly, invade surrounding and distant metastases [1]. But fortunately, the incidence of breast cancer is increasing, but the mortality rate from breast cancer is on a downward trend, due to advances in cancer treatment and the proportion of patients diagnosed at an early stage increasing [2]. Therefore, screening to detect early signs of breast cancer for timely treatment is the key to preventing and curing the disease.

Breast ultrasound is a method of imaging breast cancer by building and reconstructing images of the internal structures of the breast and body. This imaging technology is an effective tool in screening and monitoring abnormalities in the mammary gland. Compared with mammography, ultrasound has no radiation harm, making it easier for the doctor to obtain a cross-section of breast tissue by manipulating the ultrasound

handheld device, so it has more flexibility than mammography image. A breast ultrasound can help doctor identify abnormalities in the breast. Ultrasound results can be printed and saved. Breast ultrasound is the best way to differentiate a solid from a cyst, which is usually a benign lesion [3]. Moreover, breast ultrasound is simple, effective and low cost.

In recent years, deep learning (DL) is one of the most important medical image analysis topics that takes huge attention from the research community. The diagnostic process is now possible with the help of Computer Aided Imaging Systems (CADs). Specifically, these systems allow the analysis and assessment of abnormalities from medical data in a short time. They can help improve the quality of medical images, highlight abnormal structures inside the body, and perform clinical measurements [4]. CAD systems are built on technologies including image processing, computer vision, and especially AI. The process of training these computers through Deep Learning (DL) algorithms is expected to become a useful assistant for doctors to support decisions such as helping to evaluate, classify, and further study the disease condition.

To perform these tasks requires a considerable amount of data for the computer to learn. This process is similar to how image recognition software learns through repeated inputs how to recognize a particular person or object in an image. Through a sufficient number of samples, the algorithm can collect the different characteristics of a benign tumor versus a malignant tumor and make an accurate decision. The biggest challenge in medical imaging is how to deal with small datasets and limited number of labeled samples, especially supervised machine learning algorithms that require data labeling and require a large scale training model. In the field of medical imaging, stickers are done by doctors with specialized knowledge of the data, which is time consuming. Unlike other fields, collecting large amounts of data in the healthcare industry is often not feasible and costly for technical and information security reasons. Traditional photo editing and birthing methods are time-consuming and limited, especially in areas where strict standards are required, such as the medical industry. Therefore, the task of improving the size of the medical image dataset has an important meaning in the quality of the training process for problems using artificial neural network models.

Deep convolutional generative Adversarial network (DCGAN) to enhance the existing BUSI breast tumor image data in order to improve the results for research papers that will use the dataset. to train. DCGAN is an improvement of the traditional GAN model and has proven its ability to synthesize convincing images and is applied in many other works in the medical field. Using the original open source dataset BUSI including 780 breast ultrasound images to train the network model, I processed and ran the model to create an enhanced dataset including images of benign breast tumors and malignancy. The images are compared with each other at each epoch to choose better image quality at the time of birth. These images are then mixed with the original data set to be included in the classification model to evaluate the quality of the generated images.

2. BACKGROUND/ RELATED WORK

Some techniques can be applied to overcome the problem of data shortage immediately such as Synthesis data generation [5] is applied to create new images that are difficult to compare with data collected in practice. Recent advances in techniques such as GANs (generative adversarial network), VA (variational autoencoders) or data augmentation (auto augmentation, random augmentation) partially solve the data problem. Transfer learning [6] is a technique that allows AI models to learn on a related task with lots of data, and then use this knowledge to solve problems with data sets. These methods reduce the time and effort of labeling huge amounts of data, helping AI models play their role in the medical field. These technologies have many advantages in terms of processing speed, generalizability and ease of use on many types of images, so they have attracted many researches, especially in medical imaging. There are three common methods in image synthesis from simple to complex: Self-encoder (AE), U-net, and GAN-generative adversarial network model.

The latter method may include the former.

The AE network [7] is the most basic network and can participate in the structure of more complex networks than U-net [8] and GAN. This network consists of a layer of deep neural networks that contain convolutional channels to uncover some hidden patterns in the data. 1 Basic AE consists of several convolutional layers interconnected from input to output; however, very few studies have used AE in this basic form. Instead, most of the studies evaluated used variations of the underlying AE architecture for better performance. For example, a residual neural network (ResNet) [9] has been chosen in a few studies because its short connections minimize one or more layers, making it easy to train further without additional parameters addition or increase computational complexity. AE and its variants are often used as a basic component in more advanced architectures.

The U-net model is an improvement of the AE network that usually has part encryption and part decryption. The two parts are connected via shortcuts on multiple layers. These shortcuts allow high-resolution detail from the encoder to be used as additional input in the decoding. Connections are used to connect the upper layers to the lower layers so that the lower layers can also learn the simple features documented in the original layers. Most research using U-net usually follows the above architecture, with many variations and improvements proposed and studied.

Recently, the medical field has begun to apply GAN-based methods – the most advanced method in image enhancement for image synthesis [10], [11]. GANs have been used effectively in the problem of data augmentation. [12], [13], using the GAN architecture in small datasets has achieved better results than traditional image enhancement techniques. In addition, Neff [14] proposes an image synthesis model by generating pairs of images and their corresponding segment masks to support UNet in the partitioning problem, which proves in the data sets whether there is a combination of DL-generated images and real data can perfectly compete with networks trained on rigorous real data using conventional DA.

1.3 Objectives and tasks of the project

The thesis improves the original data set, which is breast ultrasound images, in order to increase the number of images to create a balance for the dataset and put it into the CNN network model to train for diagnostic tasks for medical imaging.

- Applying the DCGAN model on breast ultrasound data, generates a set of images including benign and malignant breast tumors.
- Qualitative and quantitative assessments to verify the usefulness of synthesized images.

From this, it is possible to conclude the value of these images and solve the problem of data gaps in medical data and set the stage for diseases involving other imaging.

3. METHODOLOGY

In our experiments, generate images for each type of tumor: benign and malignant tumors with the same or larger number so that it can be convenient in choosing the appropriate images for each task and helping the dataset to be balanced. Testing set includes 194 images of which 131 are benign and 63 malignant. All images in this test set are actual tumor images from the original BUSI dataset. The purpose is to evaluate the ability to classify tumors on practical problems. The newly created images combined with the images in the original data will be evaluated qualitatively, apparently and quantitatively through the classification problem by feeding to 2 models: simple self-designed and Densenet to evaluate their quality and clinical value.

3.1 Dataset

Dataset of Breast Ultrasound Images (BUSI) [15] The project used included breast ultrasound images in women between the ages of 25 and 75 years. This data is open source published on the web for research purposes only. This data was collected in 2018 from Bahaya Cairo Hospital (Egypt) using LOGIQ - E9 Ultrasound System and LOGIQ E9 Agile Ultrasound. Probe 1-5 MHz on ML6-15-D Matrix linear probe. The number of patients was 600 female patients. The dataset consists of 780 images with an average image size of 500*500 pixels. The DICOM images in the dataset were converted to PNG format by the software RadiAnt DICOM Viewer and cut out the part containing the tumor information with the Fast Photo Crop software. The images are classified into three categories, which are normal, benign, and malignant. Specifically, there are 133 normal ultrasound images without tumors, 437 images of melanoma and 210 images of benign tumors. In this report, only images with tumors were used.

3.2 Generating images Procedure

The algorithm is as follows:

Step 1: From noise z, G generates a fake image G(z) of the same size as the real image (real image is x). At the first birth, G(z) was completely noisy, without any special content

Step 2: x and G(z) are both included in D with the label true and false. Train D to learn the ability to distinguish between real and fake images.

Step 3: Put G(z) into D, based on the response of D returned, G will improve the ability to fake images

Step 4: The above process will be repeated like this, D gradually improves the discriminant ability, G gradually improves the ability to fake images. As long as D can't distinguish which image is an image created by G and which is x, then the process stops.

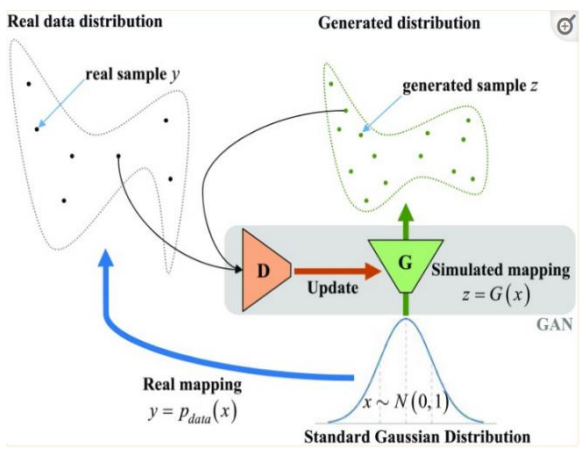


Figure 1: GAN algorithm for generating image procedure

3.3 Incorporating breast tumor image classification problem

At first epochs, the resulting image still looks like random noise. Continuing to increase the epoch later details will gradually appear and become more and more real. The images are comparable to the original data and then images are generated in batch at that epoch. Images generated from DCGAN and the images in the original data set will be combined together to put into the breast tumor image classification network to compare with the original data set.

Datasets

Training set	Set number 1	Set number 2
Benign	306 BUSI	306 BUSI+ 153 DCGAN

Malignant	147 BUSI	147 BUSI + 75 DCGAN

3.4 Evaluation

Qualitative Assessment Methods

From the generated images, they will be visually compared with the images in the original data through structures such as structure, general shape, shoreline, and other biological features.

Quantitative Evaluation Methods

Classification network structure

Simple model: use a simple classification model because it is simple, effective, and saves machine resources. Some characteristics of the network.

Complex model: Densenet model is a pretrained model. It is pre-trained on a dataset that will contain weights that are characteristic of that dataset. In DenseNet, each layer behind receives input from all previous layers and forwards it to all subsequent layers, so it carries higher computational efficiency and memory efficiency. Using a pre-trained model saves more time because parameters have been calculated to optimize the performance according to the purpose of the model.

Classifier network training

Method: use the same neural network model running on 2 datasets. First time: run on set number 1. Second time run on set number 2 (2:1 aspect ratio)

Parameter

Precision to represent the diagnostic accuracy of the network model used.

Sensitivity (recall) to calculate how many actual malignancies the model correctly identifies in the total malignancy classification.

The two parameters Accuracy and Sensitivity have each trade-off relationship. Then it is difficult to choose which is a good model because it is not known which parameter will be more suitable. Therefore, we will find a way to combine the two in a new metric, which is the F1 score.

F1-score is the harmonic average between accuracy and sensitivity. It therefore represents an assessment of accuracy on concurrency, used to test the balance between accuracy and sensitivity when the classes are not evenly distributed.

Type I and II error is a statistical concept that distinguishes two types of error based on its level of risk. Usually, the probability of making a type II mistake will have greater consequences because the patient is at a higher risk of not developing the right awareness of his condition for timely treatment, which will lead to the disease developing. Type I errors affect medical costs and patient psychology but do not affect health and the consequences are less serious.

ROC is a curve representing the classification ability of a classification model at thresholds representing the ratio of cases that are correctly classified as malignant to the total number of cases that are actually malignant. This index will evaluate the accuracy of the model's prediction on the melanoma grade. The higher its value, the better the predictive model on the group of interest in this case, melanoma.

4. EXPERIMENTAL RESULTS

The DCGAN model is trained on a dataset, the improved image is almost similar to the ultrasound image. Image post-processing steps: bring the image to the same size as the image in the dataset, remove the white edge in the generated image. The total number of images generated from the DCGAN model after running and selecting is 437 benign tumor images and 440 malignant tumors.

At first, there is a significant fluctuation in the loss of 2 networks G and D which is a good sign that the model is learning to adjust. During the run, the loss functions of the two functions tend to converge, but to a degree, almost 2 parallel lines. The network has found the optimal way and the learning process is just enough, the images can be used.

Images of benign tumors are generally easier to train, images are diverse in features, extract many different features. Many neoplasms are benign but produce a tumor that looks like a malignancy because the border is not smooth.

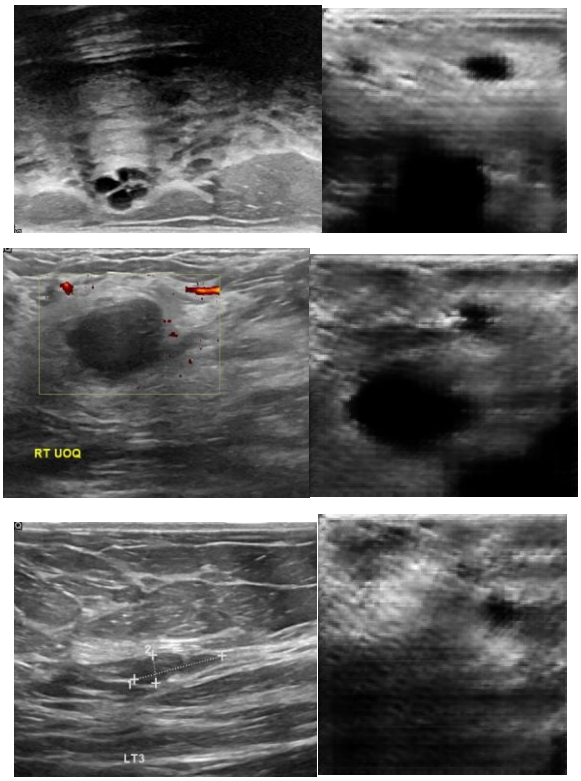
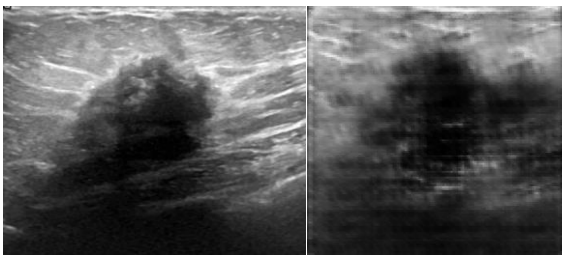


Figure 2: Example of the benign tumor generated from DCGAN model. The first column is several ultrasound benign tumor of the original data. The second shows the breast tumour image generated by DCGAN. The similarity of background tissue, colour and shape can be seen between original and generated ones.



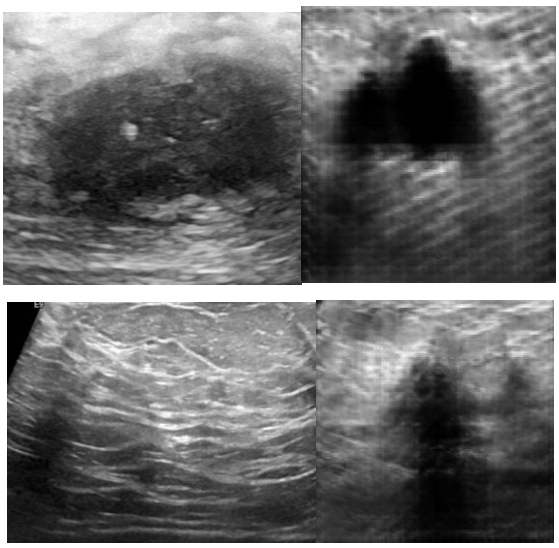


Figure 3: Example of the malignant tumor generated from DCGAN model. The first column is several ultrasound malignant tumor of the original data. The second shows the breast tumour image generated by DCGAN. The similarity of background tissue, colour and shape can be seen between original and generated ones.

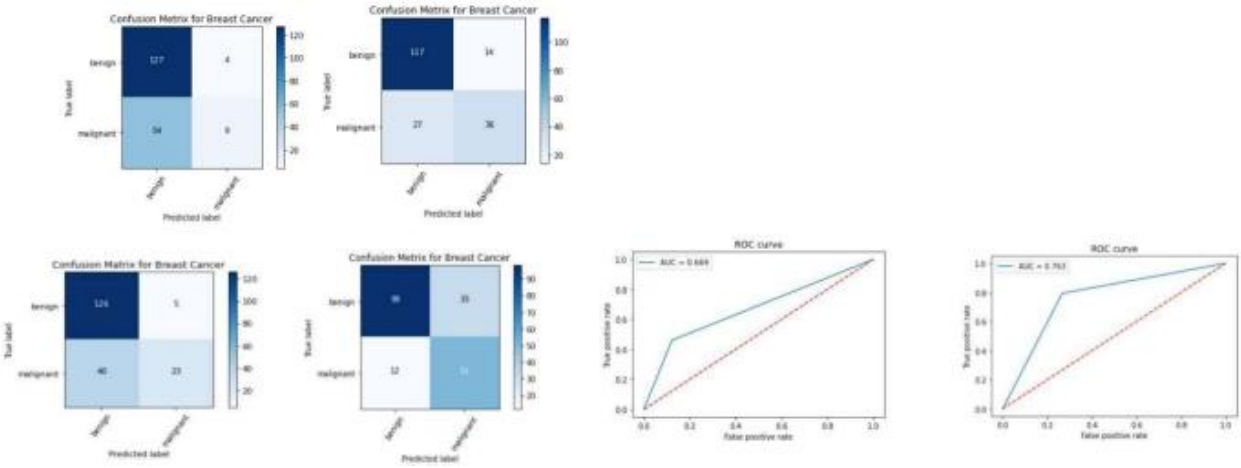


Figure 4: Confusion matrixes results running on simple classification (row 1) model and Densenet (row 2) on 2 datasets. The first column is running on BUSI. The second runs on combined dataset.

Table 1 Table of precision and sensitivity when running with simple classification model

	Precision	**Sensitivity (recall)**	**F1-score**
BUSI	**69**	**14**	**23**
BUSI+DCGAN	**72**	**57**	**64**

Table: Table of parameters for precision (Precision) and sensitivity (sensitivity) when running the model on Densenet

	Precision	**Sensitivity (recall)**	**F1-score**
BUSI dataset	**82**	**37**	**51**
BUSI + DCGAN dataset	**61**	**81**	**70**

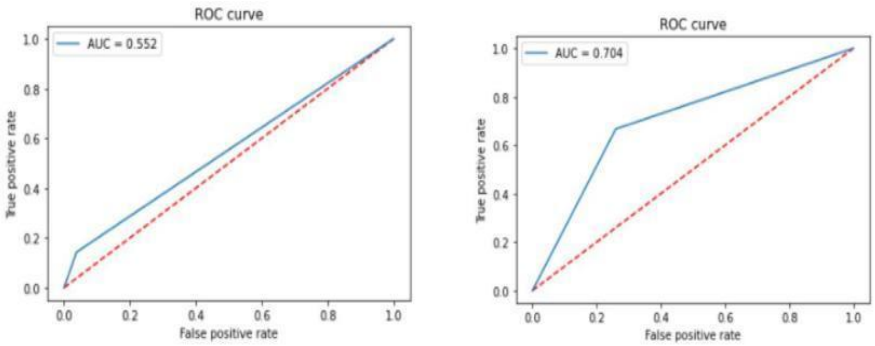


Figure 5: ROC curve image for 2 cases a) BUSI and b) BUSI combined. The first column is running on simple model. The second runs on Densenet.

Based on the parameters of Accuracy, Sensitivity, F1-score and AUC, we see that the enhancement of the original data set has been effective for the problem of image classification of benign/malignant breast tumors. The sensitivity parameter is significantly improved with an acceptable trade-off of accuracy reducing the type 2 error in misdiagnosing malignant tumors as benign. The AUC index also showed that the model was able to improve the classification ability on the subclass of melanoma.

With the results of the combined data set, we see that the performance of the classification model has improved in terms of sensitivity and the balance of the two parameters accuracy and sensitivity - shown by F1-score. Sensitivity to detect melanoma has increased, the number of cases classified as benign from 54 cases has been reduced to 27 cases by half. At the same time, the accuracy also increased from 69% to 72%, showing stability in the diagnosis of melanoma. The parameter F1-score has since also increased dramatically by nearly 3 times, showing that not only the other two parameters have improved but also increased the balance and reliability of judgment.

In the current model, the threshold set for the ROC curve is 50%. A good ROC curve is when it moves away from the set threshold and towards coordinates (1;1). Using the combined dataset BUSI and DCGAN gave better curve results and AUC increased from 0.552 to 0.704. This shows that the model has an improvement when classifying for the positive class or in this case, malignant tumor.

When using only images from the original data set, the results of correctly classifying benign masses are quite good, but the very low sensitivity shown at the sensitivity index of 37% causes many cases of melanoma to be classified as benign. The accuracy is reduced but not significantly, but because there is a trade-off relationship between the accuracy parameter and the sensitivity, this ratio is acceptable.

With the results of the combined dataset, we see an improvement in the performance of the classification model. The box that mistakenly classifies melanoma as benign is much lighter blue than the run with the original dataset, from 40 mistakenly reduce d to 12. The significantly improved sensitivity increased from 37% to 81%, indicating that the model significantly reduced the classification of melanoma to benign.

The ROC curve of the model after running on the combined dataset also shows that the classification ability on the interest-melanoma subclass is improved when the top of the ROC line tends to go closer to coordinates (1, 1) and AUC area increased from 0.669 to 0.763.

Overall, in both classification models, the use of enhanced images from the DCGAN dataset contributes to improved performance of the benign/malignant classification model. Specifically, it minimizes the type 2 error - classifying melanoma is benign - an important factor that increases the reliability of the model and helps avoid missing potentially malignant cases.

5. DISCUSSION

The DCGAN-generating adversarial network model used in this paper has proven to be an improvement over the original GAN. The stable running model can monitor the training progress and the resulting image quality. Stable running parameters suitable for imaging in the medical field. The fast model run time does not require large computer resources, enabling a large number of images to be generated in a short time and better quality images can be selected. Here are some image results during the training of the model.

The melanoma image has more specific features, which makes the model more time consuming to learn and image. Tumor imaging ranging from single-volume to multi-volume shows a limited mode-collapse problem.

As with the results produced in GAN models so far, the image inevitably has vertical and horizontal streaks, which will make the image easier to detect as a fake image. The training process needs to be monitored and filtered to select the desired output epochs.

When using only images from the dataset, the classification accuracy was good in benign tumor images and very poor performance in malignant tumors. Furthermore, the number of malignant images classified as benign is much higher than the reverse – this carries the potential for great risk and presents an unreliable outcome. It can be explained that, in the original data set, there is a limited number of images and an imbalance between the two image classes, so the model has a bias when classifying. There are many other causes affecting the results to overcome, we can choose the direction of increasing the training set size.

It is possible that in the process of image generation and feature extraction, the model has preferred to take the benign tumor features (because of the larger number of images and the tumor is not as diverse as malignant tumors) leading to a situation where the resulting images have more benign tumor characteristics.

In summary, according to the qualitative method, the DCGAN model used has the ability to produce images that are close to the sample, but the image quality is not uniform, so it is necessary to choose the images

6. CONCLUSION AND FUTURE WORK

This paper focuses on generating ultrasound images of breast tumors with Deep Convolutional Generative Adversarial Network (DCGAN) to increase small dataset data and improve performance for classification problems using CNN. DCGAN is a fairly stable network for generating medical images, the generated images are quite similar to the images from the original dataset according to qualitative visual assessment. The results demonstrated have shown that the generated images significantly improve the performance of the classification model, especially reducing the misdiagnosis of malignant to benign imaging.

Finally, although enhancement techniques are useful in scaling datasets, the most important factor is still useful medical imaging datasets. The results of this study have some limitations. The study used only a dataset with limited quantity and diversity, the images were also transformed. The image processing step can lead to information loss and affect the performance of the generated models. Composite images are generated using data from a mixture of benign and malignant masses. The reliability of composite images should be improved by collecting data from more cases, generating graded ultrasound images for different pathological cases, the reader's subjective assessment of generated images and original images. Therefore, the bias cannot be completely eliminated.

Suggestions for similar studies in the future may be to combine with images generated from media imaging methods, to add to the data set to help make optimal use of existing images. These generated images have the potential to become open source for anyone to access and apply to problems using the CNN model.

7 ACKNOWLEDGMENT

We acknowledge Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for supporting this study.

8. REFERENCES

[1] N. Harbeck et al., "Breast cancer," Nature reviews Disease primers, vol. 5, no. 1, pp. 1-31, 2019.

[2] J. P. Kösters and P. C. Gøtzsche, "Regular self-examination or clinical examination for early detection of breast cancer," Cochrane Database of Systematic Reviews, no. 2, 2003.

[3] M. Ignatiadis and C. Sotiriou, "Luminal breast cancer: from biology to treatment," Nature reviews Clinical oncology, vol. 10, no. 9, pp. 494-506, 2013.

[4] Y. Jiang, R. M. Nishikawa, R. A. Schmidt, C. E. Metz, M. L. Giger, and K. Doi, "Improving breast cancer diagnosis with computer-aided diagnosis," Academic radiology, vol. 6, no. 1, pp. 22-33, 1999.

[5] J. Dahmen and D. Cook, "SynSys: A synthetic data generation system for healthcare applications," Sensors, vol. 19, no. 5, p. 1181, 2019.

[6] L. Torrey and J. Shavlik, "Transfer learning," in Handbook of research on machine learning applications and trends: algorithms, methods, and techniques: IGI global, 2010, pp. 242-264.

[7] W.-N. Hsu, Y. Zhang, and J. Glass, "Unsupervised domain adaptation for robust speech recognition via variational autoencoder-based data augmentation," in 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2017: IEEE, pp. 16-23.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention, 2015: Springer, pp. 234-241.

[9] Q. Zhang, Z. Cui, X. Niu, S. Geng, and Y. Qiao, "Image segmentation with pyramid dilated convolution based on ResNet and U-Net," in International conference on neural information processing, 2017: Springer, pp. 364-372.

[10] N. K. Singh and K. Raza, "Medical image generation using generative adversarial networks: a review," Health Informatics: A Computational Perspective in Healthcare, pp. 77-96, 2021.

[11] N. K. Singh and K. Raza, "Medical image generation using generative adversarial networks," arXiv preprint arXiv:2005.10687, 2020.

[12] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F.-Y. Wang, "Generative adversarial networks: introduction and outlook," IEEE/CAA Journal of Automatica Sinica, vol. 4, no. 4, pp. 588-598, 2017.

[13] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," arXiv preprint arXiv:1711.04340, 2017.

[14] T. Neff, C. Payer, D. Štern, and M. Urschler, "Generative adversarial networks to synthetically augment data for deep learning based image segmentation," in Proceedings of the OAGM workshop, 2018, pp. 22-29. Vision and Applications, vol. 30, no. 7, pp. 1135-1144, 2019.

[15] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," Data in brief, vol. 28, p. 104863, 2020.