

Deep Convolutional Generative Adversarial Networks for Imbalance Medical Image Classification.

Amelia Ritahani Ismail¹, Aisyah Saidah Mohd Khalili², Nur Farah Adilah Rahim³, Syed Qamrun Nisa⁴

Department of Computer Science, Kulliyah of ICT,
International Islamic University Malaysia

*Corresponding author: amelia@iiu.edu.my

(Received: 3rd April 2023; Accepted: 18th May 2023; Published on-line: 28th July 2023)

Abstract— Medical image classification is an essential task in clinical practice and research. It enables medical professionals to be assisted in diagnosing medical conditions accurately and efficiently, leading to improved patient outcomes and survival rates. However, traditional manual interpretation methods for diagnosing medical images have some drawbacks. Firstly, imbalanced medical images often exhibit a significant disparity in the number of samples across different classes, posing challenges in training accurate and robust models that can effectively learn from limited data in the minority class while avoiding biases towards the majority class. Secondly, the limited availability of labelled data will put a further load on the healthcare system, as labelling medical images is a time-consuming and resource-intensive task, often requiring expert knowledge. This paper proposed a generative adversarial network (GAN) with the purpose of improving the limitations associated with the imbalanced distribution of medical images. Based on the experiments conducted, it shows that the proposed model exhibits a high level of accuracy for two-class labelled dataset, with a low performance for the skin cancer dataset due to number of the labelled dataset is more than two.

Keywords— Generative Adversarial Network, Deep Convolutional Neural Network, Medical imaging

I. INTRODUCTION

The aim of this paper is to propose a generative adversarial network (GAN) with the purpose of mitigating the limitations associated with the imbalanced distribution of medical images which will be experimented with Deep Convolutional Neural Network (DCNN) and Generative Adversarial Networks (GANs). GANs are embedded in the DCNN architecture for the purpose of synthesizing medical images to address the challenge posed by the scarcity of annotated data leading to the imbalanced medical images for training the model. Based on the experiment conducted, the performance of generative adversarial network (GANs) designed for the purpose of medical image classification will be evaluated. The evaluation of the model will be conducted using suitable evaluation metrics, such as accuracy, loss, precision, and recall. Based on the experiments conducted, it shows that the proposed model exhibits a high level of accuracy for two-class labelled dataset, with a low performance for the skin cancer dataset due to number of the labelled dataset is more than two.

II. REVIEW OF PREVIOUS WORKS

With the emergence of Convolutional Neural Network in the area artificial intelligence (AI), has improved the research in medical imaging. It enables new possibilities in medical imaging by utilizing deep learning techniques together with computer vision in giving healthcare

practitioners reliable resources as well as assistance to make more informed decisions in their tasks and give patients better outcomes in understanding the medical issues that they are facing. Although there is great potential for AI integration in medical imaging, there are several issues that need to be resolved, including the issues of imbalance dataset.

A. Convolutional Neural Networks

Neural networks are a subset of machine learning, which consists of node layers containing an input layer, one or more hidden layers, and an output layer. Each node connects to another layers and has a specific weight and threshold. If the output on a node is above the threshold value, that node is activated and sends the data to the next layer of the network. Otherwise, there will be no data passing to the next layer of the network. Convolutional neural networks (CNN) are one type of neural network, which is currently a state-of-the-art method for image classification. CNN, as shown on Figure 1 relies on a large training dataset to train the model to achieve high classification performance [1]. In medical, a lot of medical image datasets suffer from the imbalance problem, which hampers the detection of outliers in the datasets as most of the classification methods especially neural network assume an equal occurrence of classes [2].

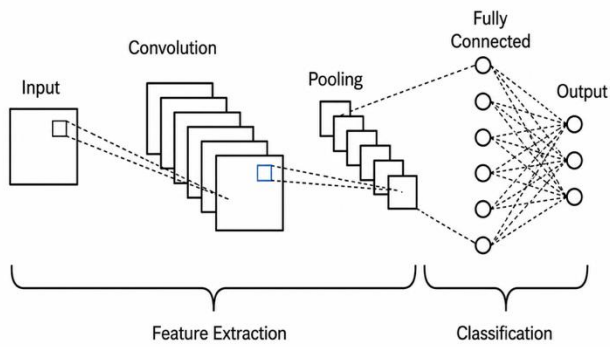


Fig. 1 CNN architecture diagram

B. Generative Adversarial Networks

GANs are a special type of neural network model where two networks are trained simultaneously, with one focused on image generator, focusing on creation of new or synthetic data and the other centred on discriminator that tries to classify examples as either real or fake [3,4,5,6] as depicted in Figure 2. [7] mentioned that synthetic sample data learned via a generative model offers greater diversity and enriches the dataset. The authors proved that the sensitivity and specificity of the synthetic data augmentation via a generative model are much better than using classic data augmentation techniques such as vertical flip, horizontal flip, and rotation [7]. Furthermore, [8] proposed a Deep Convolutional Generative Adversarial Networks (DCGAN) model that provides brain positron emission tomography (PET) images of Alzheimer’s disease in three stages, which are normal control (NC), mild cognitive impairment (MCI), and Alzheimer’s disease (AD) separately, which made the training process more complex. They suggest that the analysis of GAN architectures that generate multi-class samples can be the primary subject for future studies.

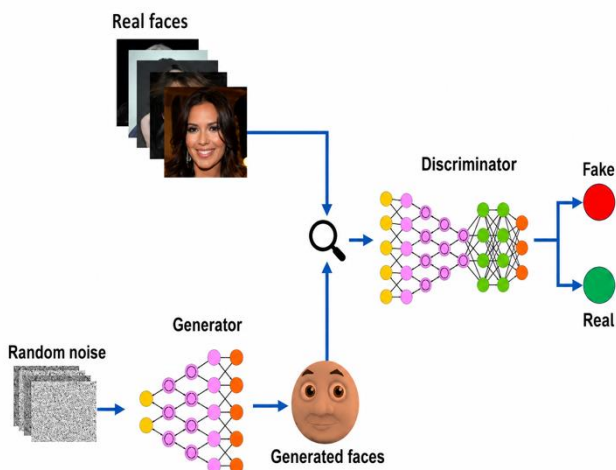


Fig. 2 GANs architecture

Researchers found out that DCGANs are capable of performing an effective analysis of dermoscopic pictures by drawing on the capabilities of both deep learning and generative adversarial networks. This enables dermatologists to detect early indicators of cancer with a surprising level of accuracy. The exceptional diagnostic capabilities of DCGANs have been collectively demonstrated by research conducted by [9],[10],[11],[12] and [13].

DCGAN also has shown significant benefits in skin cancer classification, where it is able to acquire important features and patterns indicative of a variety of skin lesion. This ability is made possible if it is train by a huge number of datasets. [9] and [10] demonstrated that DCGANs are capable of accurately classifying skin lesions, which enables them to be a decision-support systems for dermatologists. In addition, DCGANs also has the potential to assist in the reduction of diagnostic errors and the facilitation of early diagnosis to the patient suffering with skin cancer, which may ultimately lead to an increase in the percentage of patients who may survive and be cured.

Despite the encouraging findings, there are a number of issues that need to be resolved before the therapeutic application may be used on a large scale. There is still much work to be done in order to guarantee the generalizability and robustness of DCGANs over a wide variety of populations, as was demonstrated in research [11], [13]. Experts in machine learning and medical experts must work together to effectively understand the data given by DCGAN, fine-tune the algorithms, and maximize the performance of the model. In addition, ethical issues concerning patient privacy, data protection, and informed consent need to be resolved in order to keep patients and medical professionals alike trusting and accepting of this technology. This can only be accomplished by addressing these issues. DCGANs have the potential to revolutionize the detection and classification of medical images and profoundly impact the area of medical, provided that further research and development are carried out on them.

III. METHODOLOGY

A. Design

In this work, initially pre-processing and augmentation of training data is done. Following that, models are developed with privacy-enhancing techniques. Next, training model will be trained, predicted, and evaluated.

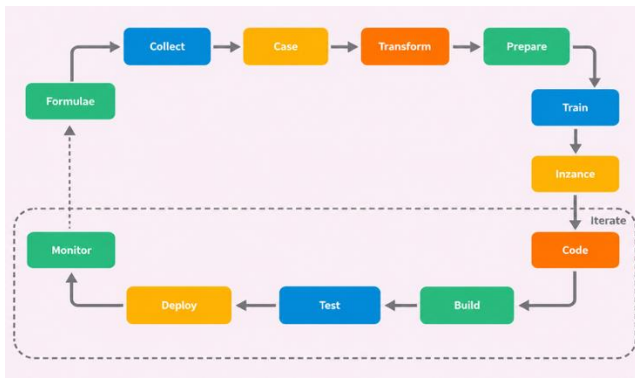


Fig. 3 ML Ops Lifecycle

Furthermore, to make sure model deployment is efficient and reliable, the Machine Learning Operations (ML Ops) lifecycle will be followed. Basically, the ML Ops lifecycle strings model and software development together in a unified machine learning life cycle. Figure 3 demonstrates the ML Ops life cycle. The first cycle is formulation. In this cycle, the problem statement, objectives, and success criteria for research will be defined. Next, the data collection and preparation will be performed. Dataset was obtained online from Kaggle. After that, the data was curated. The cleaning process of the data will happen in this cycle. After the data has been cleaned, it is transformed to make sure it suits the analytics and ML modeling. The next cycle is exploration, where the target value is defined for the model to predict. Then, the training cycle begins. Model will be trained and compare the model's performance and select a couple of the best-performing models. After the training phase is done, the model is evaluated. However, this is not a sequential cycle, which means if problem is encountered, it allows to return to the relevant cycle to fix it and continue the cycle again. Next, the development cycle progresses by initiating the planning of the outcome of research. Some requirements or problems will be obtained from users and collect the required data to use in development process. Then, code will be developed, build, and do the testing. After the model has passed the cycles, the release and deployment of the model can take place. Next, the operation part is performed, and lastly constantly monitor the health of services, errors, latencies, model predictions, etc. If a problem arises, it will depend on its severity and diagnosis. The monitoring phase is a continuous process that will monitor the performance and behaviour of deployed model.

Apart from that, there are several key components of ML Ops that help improve model performance and quality. Firstly, there is version control, which enables reproducibility and tracks changes throughout this project's development. Next, continuous deployment to automate the process of building, testing, and deploying the models.

Apart from that, monitoring and logging are also key components of ML Ops which play a crucial role in tracking the performance and behaviour of deployed models. Besides, model deployment and serving, which involve setting up a scalable and robust serving infrastructure, and lastly, model monitoring and retraining help detect performance degradation and ensure that they stay accurate and up to date.

B. Implementation

This research experimented on dataset that was obtained online from Kaggle. The dataset consists of skin cancer images, and it contains four classes. The number of images used in training, testing, and validation for dataset has been parallelized to the 8:1:1 ratio with the imbalanced number of images between the classes.

Next, models with different GANs have been implemented in Python using Google Colab and Jupyter Notebook. All these models use the same CNN architecture as an algorithm, such as the same input shape, optimizer, learning rate, and loss. In the end, these models will be evaluated and compared with appropriate evaluation metrics such as accuracy, loss, precision, and recall. This paper will also highlight how imbalanced images affect the model's performance. The overall implementation flowchart is shown in Figure 4.

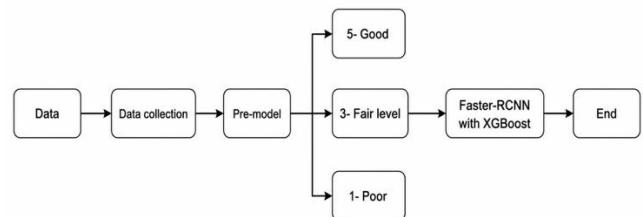


Fig. 4 Overall Implementation Flowchart

C. Testing

Since the images used as input data and a class of deep neural networks, TensorFlow with GPU is used. TensorFlow with GPU is a package of one of the top libraries for deep learning, known as TensorFlow. It requires a few steps of system and environment setup to be able to use this package. First and foremost, it is essential to ensure that the laptop is equipped with a GPU.

The next step will be the installation of the CUDA toolkit, CuDNN, and NVIDIA Driver. This includes the step of verifying the paths of CUDA and CuDNN in environment variables once the installation is complete. Finally, is to create a virtual environment in Anaconda complete with the installation of the TensorFlow package. For verification, TensorFlow is imported and see if the GPU works. Once everything is set, the development of model can proceed.

D. Tools and Environment Setup

The first step is to set up the tools and environment that are suitable for the model's development. NVIDIA driver is installed, the CUDA toolkit, and CuDNN. The paths of CUDA and CuDNN need to be recognized by the system, so these paths in the environment variable are verified. Next, a new environment is created with Python 3.10 and TensorFlow 2.8.4 is installed in the Anaconda prompt. TensorFlow with GPU is being installed successfully when a list of GPU devices is returned.

E. Data Preprocessing and Preparation

The images are split into training, validation, and testing phases with a ratio of 8:1:1. Next, the pixel values of the images are transformed to a common scale by dividing them by 255. Since pixels have values between 0 and 255, dividing them by 255 will scale the data to the range of 0 to 1 so that the network can learn the image patterns and features better without getting impacted by the wide pixel value ranges.

F. Data Modelling and Training

Deep Convolutional Generative Adversarial Networks (DCGAN) are used to create synthetic images for each class of the dataset. These imbalanced images of synthetic and original images are then evaluated by CNN.

IV. RESULTS

A. Synthetic Images via Deep Convolutional Generative Adversarial Networks

DCGAN is used to train the model to create synthetic images. It involves adversarial learning, where both the discriminator and generator have opposite goals but are trained simultaneously. The discriminator aims to accurately classify real and synthetic samples, while the generator tries to deceive it by producing samples that are indistinguishable from actual data. During the training, both networks engage in a feedback loop that enhances the generator's capacity to produce realistic images. As a result, the generated images' quality continues to improve. The Figure 5 shows how the synthetic images for dataset improve during the training process.

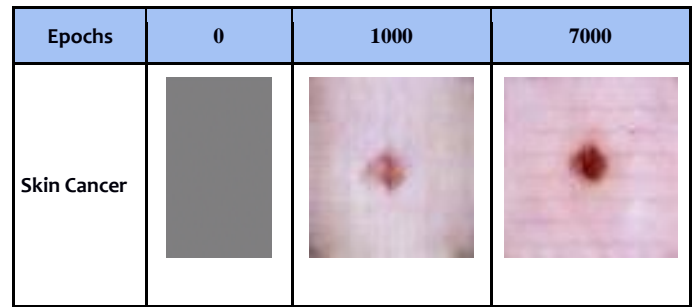


Fig. 5 Sample Generation of Synthetic data by GAN

B. Model Evaluation of Imbalanced Original and Synthetic Images of Pneumonia

The pneumonia dataset, which is obtained from Mendeley Data, consists of two classes, which are pneumonia and normal. The class normal has the majority data with a total of 235 images and pneumonia as the minority class with only 127 images. Below is the result of an imbalance between original and synthetic images using CNN.

TABLE I
RESULTS FOR PNEUMONIA DATASET

		Original Images	Synthetic Images
Train accuracy		0.955	1.00
Val accuracy		0.722	1.00
Precision	Normal	0.69	0.60
	Pneumonia	0.00	0.00
Recall	Normal	1.00	1.00
	Pneumonia	0.00	0.00
Number of images		362	2500

C. Model Evaluation of Imbalanced Original and Synthetic Images of Skin Cancer

The skin cancer dataset, which is obtained from Kaggle, consists of only four classes which are malignant, benign, vascular lesion, and dermatofibroma. Since the original images are already imbalanced between these classes, original images are used for the evaluation without any modification or exclusion. Benign will be the majority class with 1800 images, followed by malignant with 1497 images, vascular lesions with 142 images, and dermatofibroma with only 111 images. Table II shows the result of an imbalance between original and synthetic images of the skin cancer dataset using CNN.

TABLE II
 RESULTS FOR SKIN CANCER DATASET

		Original Images	Synthetic Images
Train accuracy		0.048	0.132
Val accuracy		0.051	0.732
Precision	Benign	0.51	0.39
	Malignant	0.00	0.00
	Vascular lesion	0.00	0.00
	Dermato fibroma	0.00	0.00
Recall	Benign	1.00	1.00
	Malignant	0.00	0.00
	Vascular lesion	0.00	0.00
	Dermato fibroma	0.00	0.00
Number of images		3550	3800

V. CONCLUSIONS

Deep Convolutional Generative Adversarial Networks (DCGAN) are developed to create synthetic images. It utilizes adversarial learning to train a generator network to create high-quality synthetic data that closely reflects the distribution of real data. Since it is a resource and time-consuming, the training is done only up until epoch 7000.

For future improvement, the training can be done for more than 7000 epochs to get more high-quality images. The consequences of using imbalanced images in a model's performance are also recorded. The model's accuracy is high for medical images with two-labelled class except for the skin cancer dataset. High accuracy does not necessarily mean that the model is good. Accuracy alone is not enough to evaluate the model's performance, so it is crucial to consider other evaluation metrics as well, such as precision, recall, and confusion matrix, to properly understand the way a model performed across all classes. From the result it can be seen that the model is biased toward the majority class and cannot generalize well to the minority samples. In other words, the model is overfitting. As for the skin cancer dataset, its accuracy is the lowest, and it has very poor performance due to the number of classes. Increasing the classes in an imbalanced dataset can lead to worse model performance, as it can be challenging for the model to learn to differentiate more classes. As the number of classes grows, the model's sensitivity to minority classes decreases. Future work will include the improvement of the model to facilitate the imbalance issues in medical imaging for multi-labelled classes.

ACKNOWLEDGMENT

The authors would like to thank the Kulliyah of ICT IIUM

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

- [1] Basic CNN architecture: Explaining 5 layers of Convolutional Neural Network. upGrad blog. (2022, October 3). <https://www.upgrad.com/blog/basic-cnn-architecture/>
- [2] Gao, L., Zhang, L., Liu, C., & Wu, S. (2020, August). Handling imbalanced medical image data: A deep-learning-based one-class classification approach. Artificial intelligence in medicine. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7519174/>
- [3] Brauneck, A., Schmalhorst, L., Kazemi Majdabadi, M. M., Bakhtiari, M., Völker, U., Baumbach, J., Baumbach, L., & Buchholtz, G. (2023, March 30). Federated Machine Learning, Privacy-Enhancing Technologies, and Data Protection Laws in Medical Research: Scoping Review. *J Med Internet Res*, 25, 1. 10.2196/41588
- [4] Li, L., Xie, N., & Yuan, S. (2022, 11 16). A Federated Learning Framework for Breast Cancer Histopathological Image Classification. *Electronics*, 3767. <https://doi.org/10.3390/electronics11223767>
- [5] Senekane, M. (2019). Differentially Private Image Classification Using Support Vector Machine and Differential Privacy. *Machine Learning and Knowledge Extraction*, 1(1), 483–491. <https://doi.org/10.3390/make1010029>
- [6] Yi, X., Walia, E., & Babyn, P. (2019, 12). Generative adversarial network in medical imaging: A review. *Medical Image Analysis*, 58. <https://doi.org/10.1016/j.media.2019.101552>
- [7] Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018, 12 10). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion

- classification. *Neurocomputing*, 321, 321-331.
<https://doi.org/10.1016/j.neucom.2018.09.013>
- [8] Islam, J., & Zhang, Y. (2020, 3 30). GAN-based synthetic brain PET image generation. *Brain Informatics*, 7. <https://doi.org/10.1186/s40708-020-00104-2>
- [9] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). A classification of skin cancer comparable to that of a dermatologist using deep neural networks. *Nature*, 542(7639), 115-118. doi:10.1038/nature21056
- [10] Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., ... & Thomas, L. (2018). The diagnostic performance of a deep learning convolutional neural network was compared to that of 58 dermatologists for the purpose of dermoscopic melanoma identification. *Annals of Oncology*, Volume 29 Number 8 (Pages 1836-1842). doi:10.1093/annonc/mdy166
- [11] Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., ... & Zalaudek, I. (2020). Collaboration between humans and computers in the detection of skin cancer. *Nature Medicine*, 26(8), 1229-1234. doi:10.1038/s41591-020-0975-z
- [12] Menon, S., Shiny, N. N., Thirunavukarasu, R., Chandra Mouli, P. V., & Ramanathan, R. (2021). The use of deep convolutional generative adversarial networks for the diagnosis of skin cancer more effectively. The eighth edition of *Healthcare Technology Letters* includes the pages 101-107. doi:10.1049/htl2.12004
- [13] Matsunaga, K., Hamada, A., Minagawa, A., Koga, H., & Kiyohara, Y. (2022). Skin cancer categorization through the use of deep convolutional generative adversarial networks: A case study of face basal cell carcinoma. *The Journal of Dermatological Science*, Vol. 97, No. 1, Pages 18-24. doi:10.1016/j.jdermsci.2019.12.002