Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240612

Deep Learning-Based Synthetic Skin Lesion Image Classification

Saadullah Farooq ABBASI[a], Muhammad BILAL[b], Teesta MUKHERJEE[a], James CHURM[a], Omid POURNIK[a] , Gregory EPIPHANIOU[c] and Theodoros N. ARVANITIS[a,1]

[a] *Department of Electronic, Electrical and Systems Engineering, University of Birmingham, Birmingham B15 2TT, United Kingdom*

[b] *Department of Artificial Intelligence, Rare Sense Inc, Covina, California, USA*

[c] *WMG, University of Warwick, Coventry CV4 7AL, United Kingdom*

ORCiD ID: Saadullah Farooq Abbasi https://orcid.org/0000-0001-9814-3023, Muhammad Bilal https://orcid.org/0009-0005-1715-4123,Teesta Mukherjee https://orcid.org/0009-0003-8023-8317, James Churm https://orcid.org/0000-0003-0654-5960, Omid Pournik https://orcid.org/0000-0001-7938-0269 Gregory Epiphaniou https://orcid.org/0000-0003-1054-6368, Theodoros N. Arvanitis https://orcid.org/0000-0001-5473-135X

Abstract. Advances in general-purpose computers have enabled the generation of high-quality synthetic medical images that human eyes cannot differ between real and AI-generated images. To analyse the efficacy of the generated medical images, this study proposed a modified VGG16-based algorithm to recognise AI-generated medical images. Initially, 10,000 synthetic medical skin lesion images were generated using a Generative Adversarial Network (GAN), providing a set of images for comparison to real images. Then, an enhanced VGG16-based algorithm has been developed to classify real images vs AI-generated images. Following hyperparameters tuning and training, the optimal approach can classify the images with 99.82% accuracy. Multiple other evaluations have been used to evaluate the efficacy of the proposed network. The complete dataset used in this study is available online to the research community for future research.

Keywords. Synthetic data, generative adversarial network, convolutional neural network, medical images, VGG16

1. Introduction

Image synthesis is a process of generating images using artificial intelligence from image, text, sketch, or audio [1]. This has been used in multiple real-world applications like the generation of art [2], editing [3], and painting [4]. For this reason, image synthesis has been extensively used since its origin. After the development of Generative Adversarial Networks (GANs) [5], the use of image synthesis increased exponentially. GAN mainly

[1] Corresponding Author: Professor Theodoros N. Arvanitis, Department of Electronic, Electrical and Systems Engineering, School of Engineering, College of Engineering and Physical Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom; E-mail: T.Arvanitis@bham.ac.uk.

consists of two networks: the generator and the discriminator. While the generator generates images that can fool the discriminator, the discriminator tries to differentiate between real and artificial images. Initially, the generator creates obvious fake images, and the discriminator easily classifies artificial images. However, as the training progresses, the generator learns features that can fool the discriminator, and finally, it is unable to classify real vs artificial images.

Over the last decade, GANs have been extensively used in the fields of image generation [6], text generation [7] and bio-physiological data generation [8]. Medical images contain critical information and features; therefore, it is important to generate images of high quality. For this reason, Maayon et al. [9] proposed a GAN-based synthetic medical image augmentation to generate liver lesion images. Initially, GAN was used to increase the dataset, and then a convolutional neural network (CNN) based classifier was developed for liver lesion classification. The classifier performance improved by 7% by adding the augmented data for training and testing. Furthermore, in 2018, Talha et al. [10] proposed a novel GAN-based algorithm for medical retinal image generation (MI-GAN). The proposed algorithm generated images and their masks with a dice coefficient of 0.837. To enhance the performance of breast cancer detection, conditional GANs (cGANs) were proposed to generate complex structured images [11]. In cGANs, additional information like class labels, feature correlation, or other auxiliary information has been added. The proposed study efficiently represented the complex patterns in the generated images. In addition, this study showed improvement in the classification accuracy of binary classification i.e., malignant, or benign. Zhiwei et al. [12] proposed a style-based GAN architecture to classify skin lesions. The results show that GAN-based architecture effectively generates images with high resolution. A significant breakthrough in research was achieved by Jyoti et al. [13] in 2020. They used GANs to generate high-quality brain PET images. This research was a breakthrough as the dataset of PET images is very limited.

Multiple other studies were proposed to generate images using GANs for applications like 3D Echocardiography images [14] and Cervical cancer [15]. Researchers have proposed multiple algorithms to evaluate the efficacy of the generated facial images [16]. However, the studies to test the efficacy of the generated images or the proposed GANs are limited.

This study proposed a modified VGG16-based algorithm to detect artificial medical images. Initially, 10000 fake skin lesion images were generated using spectrally normalised GAN (SNGAN). Then, a modified VGG16-based algorithm based on CNN architecture has been developed for binary classification. The proposed algorithm achieved an accuracy of 99.82% with sensitivity and specificity of 99.65% and 99.95%, respectively.

2. Methods

The method section is structured into two primary parts: creating SNGAN and refining a modified VGG16-based classification algorithm tailored for artificial image detection.

2.1. SNGAN-Based Image Generation

Fig.1 illustrates the block diagram of the proposed SNGAN for skin lesion image generation. The proposed method is mainly divided into two main components: the discriminator and the generator. The generator is responsible for generating images that look like real images, and then the discriminator is responsible for differentiating the real images from those generated by the generator. The process continues unless the discriminator cannot differentiate between the two groups.

Spectral normalisation serves as a weight regularisation method employed in the discriminator of a GAN to address the issue of exploding gradients. Its function is to stabilise the training process by rescaling the weights of convolution layers using a norm. This norm is calculated through the power iteration method and is applied just before the forward function call. To elaborate, spectral normalisation focuses on the Lipschitz constant as its sole hyperparameter. This constant pertains to a regularisation property of continuous functions that constrains their values. More precisely, in our case, where the activation function is a LeakyReLU, the Lipschitz constant is set to 1. The control of this parameter within the discriminator is facilitated by bounding it through the spectral norm. The Lipschitz norm is equivalent to the upper bound of the layer's gradient. The spectral normalisation of a weight normalises the weight of each layer and, consequently, the entire network, effectively mitigating issues related to exploding gradients. For the proposed study, we have generated 10000 images using SNGAN.

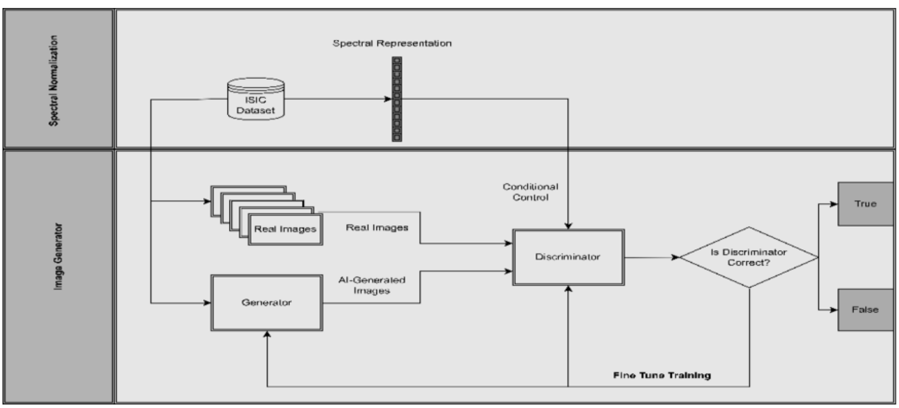

Figure 1. Block Diagram of SNGAN for Synthetic Image Generation

2.2. Modified VGG16 algorithm for AI-generated Image Detection

The proposed research used the pre-trained VGG16 model, also recognised as VGGNet, by integrating additional layers to classify artificial medical images from real ones. VGG16 comprises 16 CNN layers. Input to the network comprises images of dimensions (128, 128, 3). The initial two layers comprise 64 channels with a 33 filter size and identical padding. Subsequently, following a max pool layer with a stride of (2, 2), two layers entail convolutional layers with 128 filters and a filter size of (3, 3). Another max-pooling layer with the same stride as the preceding one succeeds this. Then, there are two convolution layers with a filter size of (3, 3) and 256 filters. Moreover, a max pool layer and two sets of three convolution layers are present. Every set has 512 filters with the same padding and a size of (3, 3). Next, two convolution layers are stacked on top of the image. After every convolution layer, a 1-pixel padding is inserted to maintain the image's spatial characteristics. A (7, 7, 512) feature map is the end result. The result of flattening this output is a feature vector (1, 25088). Subsequently, three fully connected layers of size (1, 4096), (1, 4096), and a softmax function of size 1000 are employed.

Following the existing classifier of the VGG16 model, we appended three more fully connected layers. These layers aim to further extract and process features from the flattened output of the convolutional layers:

- Two layers of fully connected neural network nodes, each consisting of 4096 units, utilising ReLU activation functions and batch normalisation to introduce non-linearities and improve feature processing.
- A single unit fully connected layer, responsible for generating the binary classification decision.

These additional layers enhance the model's capacity to capture complex patterns and improve its performance on binary classification tasks by introducing more non-linearity and feature processing capabilities. In addition, dropout layers have been used to decrease overfitting in the proposed architecture.

3. Results and Discussion

The proposed study conducted many experiments, and the results are reported in this section. The proposed network has been trained and tested on 12th Gen Intel Core i9 with NVIDIA GeForce RTX 3080 Ti Laptop GPU. In addition, the network was implemented using PyTorch in Python. The proposed network achieved an accuracy of 99.82% for real vs artificial image classification. Fig. 2 shows the confusion matrix of the proposed network.

		Original	Synthetic
True	Original	1999	1
	Synthetic	6	1994
		Predicted	

Figure 2. Confusion matrix of the proposed study.

The performance metrics, used to evaluate the efficacy of the proposed network architecture, have been calculated using the confusion matrix given in Fig. 2: Accuracy (99.82%), sensitivity (99.65%), specificity (99.95%), F1-Score (99.80%) and Recall (99.65%).

From the above, it is evident that the proposed enhanced VGG16 network architecture can efficiently classify real vs AI-generated images of skin lesions. This is the pilot study, and it can be further explored using different datasets to check the efficacy of the generated medical images. The results of the proposed network architecture are commendable. However, there is no existing study with which this study can be compared; therefore, that can be considered as a limitation of the proposed study. The complete dataset used in this study is publicly available to the research community for future research [17].

4. Conclusions

This research demonstrates the potential of deep learning techniques to accurately differentiate between real and AI-generated medical images, which is crucial for validating the authenticity and reliability of synthetic data in medical diagnostics.

Despite the promising results, this pilot study acknowledges certain limitations, including the lack of direct comparisons with existing studies due to the novelty of the approach. Future research should explore the application of our methodology across diverse datasets and medical imaging domains to further validate and enhance its generalisability and robustness.

Our study sets a precedent for the continued development and application of AI in medical imaging, aiming to achieve higher accuracy and reliability in clinical practice. The integration of advanced GANs and deep learning models holds significant promise for augmenting medical image datasets and enhancing diagnostic capabilities.

Acknowledgement

This work is partially funded by both Advanced Security-for-safety Assurance for Medical Device IoT (MedSecurance) - Grant Agreement number: 101095448 and Innovative Applications of Assessment and Assurance of Data and Synthetic Data for Regulatory Decision Support (INSAFEDARE) - Grant Agreement number: 101095661.

References

[1] Baraheem SS, Le TN, Nguyen TV. Image synthesis: a review of methods, datasets, evaluation metrics, and future outlook. Artificial Intelligence Review. 2023 Oct;56(10):10813-65.
[2] Elgammal A, Liu B, Elhoseiny M, Mazzone M. Can: Creative adversarial networks, generating" art" by learning about styles and deviating from style norms. arXiv preprint arXiv:1706.07068. 2017 Jun 21.
[3] Chen J, Shen Y, Gao J, Liu J, Liu X. Language-based image editing with recurrent attentive models. InProceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 8721-8729).
[4] Nazeri K, Ng E, Joseph T, Qureshi FZ, Ebrahimi M. Edgeconnect: Generative image inpainting with adversarial edge learning. arXiv preprint arXiv:1901.00212. 2019 Jan 1.
[5] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. Advances in neural information processing systems. 2014;27.
[6] Siarohin A, Lathuilière S, Sangineto E, Sebe N. Appearance and pose-conditioned human image generation using deformable gans. IEEE transactions on pattern analysis and machine intelligence. 2019 Oct 16;43(4):1156-71.
[7] Li Y, Pan Q, Wang S, Yang T, Cambria E. A generative model for category text generation. Information Sciences. 2018 Jun 1;450:301-15.
[8] Fahimi F, Dosen S, Ang KK, Mrachacz-Kersting N, Guan C. Generative adversarial networks-based data augmentation for brain–computer interface. IEEE transactions on neural networks and learning systems. 2020 Aug 25;32(9):4039-51.
[9] Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing. 2018 Dec 10;321:321-31.
[10] Iqbal T, Ali H. Generative adversarial network for medical images (MI-GAN). Journal of medical systems. 2018 Nov;42(11):231.
[11] Abedi M, Hempel L, Sadeghi S, Kirsten T. GAN-based approaches for generating structured data in the medical domain. Applied Sciences. 2022 Jul 13;12(14):7075.
[12] Qin Z, Liu Z, Zhu P, Xue Y. A GAN-based image synthesis method for skin lesion classification. Computer Methods and Programs in Biomedicine. 2020 Oct 1;195:105568.
[13] Islam J, Zhang Y. GAN-based synthetic brain PET image generation. Brain informatics. 2020 Dec;7(1):3.
[14] Tiago C, Gilbert A, Beela AS, Aase SA, Snare SR, Šprem J, McLeod K. A data augmentation pipeline to generate synthetic labeled datasets of 3D echocardiography images using a GAN. IEEE Access. 2022 Sep 16;10:98803-15.
[15] Torfi A, Fox EA, Reddy CK. Differentially private synthetic medical data generation using convolutional GANs. Information Sciences. 2022 Mar 1;586:485-500.
[16] Baraheem SS, Nguyen TV. AI vs. AI: Can AI Detect AI-Generated Images?. Journal of Imaging. 2023 Sep 28;9(10):199.

[17] Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data. 2018 Aug 14;5(1):1-9.