Detecting Retinal Fundus Image Synthesis by Means of Generative Adversarial Network

Francesco Mercaldo[1,3], Luca Brunese[1], Mario Cesarelli[2], Fabio Martinelli[3] and Antonella Santone[1]
[1]*Department of Medicine and Health Sciences "Vincenzo Tiberio", University of Molise, Campobasso, Italy*
[2]*Department of Engineering, University of Sannio, Benevento, Italy*
[3]*Institute for Informatics and Telematics, National Research Council of Italy (CNR), Pisa, Italy*

Keywords: Retina, GAN, Bioimage, Deep Learning, Classification.

Abstract: The recent introduction of Generative Adversarial Networks has showcased impressive capabilities in producing images that closely resemble genuine ones. As a consequence, concerns have arisen within both the academic and industrial communities regarding the difficulty of distinguishing between counterfeit and authentic images. This matter carries significant importance since images play a crucial role in various fields, such as biomedical image recognition and bioimaging classification. In this paper, we propose a method to discriminate retinal fundus images generated by a Generative Adversarial Network. Following the generation of the bioimages, we employ machine learning to understand whether it is possible to differentiate between real and synthetic retinal fundus images. We consider a Deep Convolutional Generative Adversarial Network, a specific type of Generative Adversarial Network, for retinal fundus image generation. The experimental analysis reveals that even though the generated images are visually indistinguishable from genuine ones, an F-Measure equal to 0.97 is obtained in the discrimination between real and synthetic images. Anyway, this is symptomatic that there are several retinal fundus images that are not classified as such and are thus considered authentic retinal fundus images.

1 INTRODUCTION AND RELATED WORK

Generative Adversarial Networks (GANs) are a type of neural network (Cimitile et al., 2017; Bacci et al., 2018; Mercaldo and Santone, 2020; Mercaldo et al., 2016) employed in unsupervised machine learning. They consist of two opposing components: a generator network responsible for creating synthetic data and a discriminator network designed to differentiate between authentic and fabricated instances. These components engage in a competitive process, with the discriminator trying to spot synthetic data while the generator endeavors to deceive the discriminator by producing realistic examples. Through this adversarial interaction, the GAN model learns to generate data that closely resembles the training dataset. This capability has various applications, including predicting future data or generating images, once the network has been trained on a specific dataset(Goodfellow et al., 2020).

One key advantage of GANs is their ability to generate high-quality synthetic data. The collaborative nature of the generator and discriminator enables the generator to learn from the feedback provided by the discriminator, resulting in the production of synthetic data that closely resembles real data. Furthermore, GANs often exhibit speed and efficiency benefits compared to traditional methods. By leveraging parallelization techniques, GANs use parallel neural networks for computational tasks, enabling faster processing. GANs also excel in generating diverse types of data, including images, videos, audio, and text, thanks to their inherent adaptability, as they are built upon neural networks that can be easily customized to handle different data types. In contrast, traditional methods often require specific techniques tailored to each data type, making GANs a more flexible solution.

Numerous research papers have delved into the use of GANs in biomedical contexts (Huang et al., 2022; Zhou et al., 2021; Huang et al., 2023; Huang et al., 2021), serving various purposes. For example, Orlando et al. (Orlando et al., 2018) proposed

Mercaldo, F., Brunese, L., Cesarelli, M., Martinelli, F. and Santone, A.
Detecting Retinal Fundus Image Synthesis by Means of Generative Adversarial Network.
DOI: 10.5220/0012403100003657

In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2024) - Volume 1*, pages 471-478
ISBN: 978-989-758-688-0; ISSN: 2184-4305

the generation of retinal fundus images with lesions, while Fu and colleagues (Fu et al., 2018) introduced retinal fundus image augmentation. Differently from these papers, the objective of this paper is to investigate whether GANs can be employed to generate retinal fundus images that are indistinguishable from authentic ones, constituting the primary contribution of this paper.

As a matter of fact, in this paper, we introduce an approach aimed at evaluating the potential impact of GANs on retinal fundus image classification tasks. Specifically, we employ a Deep Convolutional GAN (DCGAN) to generate a set of images using a dataset of retinal fundus images.

While GANs have been considered for various purposes in the biomedical field, such as retinal vessel segmentation and liver lesion classification, our focus is on generating synthetic images that are indistinguishable from real ones and evading dedicated classifiers. Our results demonstrate that as the number of training epochs increases, the synthetic images become progressively more realistic and are better at evading detection by the classifiers.

The paper is organized as follows: in Section 2 we provide a description of the method we designed and implemented to assess the DCGAN ability to generate indistinguishable retinal fundus images. The experimental results are presented in Section 3, followed by a discussion about future directions in the last section.

2 THE METHOD

In this section, we outline the proposed approach, which encompasses two primary objectives i.e., creating synthetic retinal fundus images and distinguishing these synthetic images from authentic ones obtained from real-world retinal fundus images.

The initial phase of our method involves the development and utilization of a DCGAN for the generation of synthetic retinal fundus images. This step is illustrated in Figure 1.

In every GAN architecture, there is at least one generator (referred to as "Generator" in Figure 1) and one discriminator (referred to as "Discriminator" in Figure 1). These two components engage in a competitive process where the generator strives to enhance its ability to produce images that closely match the distribution of the training data, taking cues from the feedback provided by the discriminator.

Hence, the training of a GAN is a critical process that involves two neural networks (Mercaldo and Santone, 2020; Canfora et al., 2015a; Canfora et al., 2015b; Canfora et al., 2013; Canfora et al., 2015c), a generator ("Generator" in Figure 1) and a discriminator ("Discriminator" in Figure 1), engaged in a competitive endeavor to enhance their performance. In the following, we provide an overview of the GAN training process:

1. Initialization: Initially, both the generator and discriminator networks are initialized with random weights.
2. Objective: The generator's goal is to create synthetic data that is virtually indistinguishable from real data, while the discriminator's objective is to accurately classify real data as real and generated data as synthetic.
3. Training Loop:
   (a) Generator Training ("Generator" in Figure 1):
   - The generator takes random noise as input and generates synthetic data.
   - This generated data is mixed with real data, if available, to create a training batch.
   - The generator's output is then passed through the discriminator, and the loss is calculated based on how effectively the discriminator was deceived (i.e., how well the generated data is classified as real).
   - The generator's weights are updated using gradient descent to minimize this loss, thereby improving its ability to generate more realistic data.

   (b) Discriminator Training ("Discriminator" in Figure 1):
   - The discriminator takes both real and generated data as input and classifies them as real or synthetic.
   - The loss for the discriminator is determined based on its accuracy in classifying real and generated data.
   - The discriminator's weights are updated to minimize this loss, making it better at distinguishing real from generated data.
4. Adversarial Training: The core concept in GANs is adversarial training, where the generator and discriminator iteratively enhance their performance by competing against each other. As the training progresses, the generator becomes more proficient at producing realistic data, and the discriminator becomes more adept at distinguishing real from synthetic data.
5. Convergence: Training continues for a predetermined number of epochs or until a convergence criterion is met. Convergence is achieved when the generator generates data so realistic that the

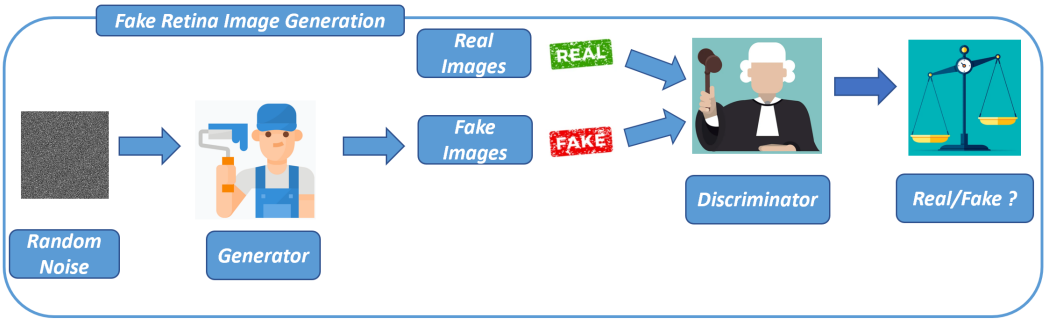

Figure 1: The step related to the synthetic retinal fundus images generation.

discriminator cannot reliably differentiate it from real data.

6. Evaluation: After training, the generator can be employed to produce synthetic data, and the discriminator can be used to evaluate the authenticity of data samples.

The DCGAN architecture introduced the integration of Convolutional Neural Networks (CNNs) in both the discriminator and generator components.

DCGAN provides a set of architectural guidelines aimed at enhancing the stability of the training process(Radford et al., 2015):

1. Replacing pooling layers with strided convolutions in the discriminator and fractional-strided convolutions in the generator.
2. Incorporating batch normalization (batchnorm) in both the generator and discriminator.
3. Avoiding fully connected hidden layers in deeper architectures.
4. Applying ReLU activation for all generator layers, except the output layer, which uses Tanh activation.
5. Employing LeakyReLU activation in all discriminator layers.

Strided convolutions refer to convolutional layers with a stride of 2, used for downsampling in the discriminator. Conversely, fractional-strided convolutions (or Conv2DTranspose layers) utilize a stride of 2 for upsampling in the generator.

In the context of DCGAN, batch normalization (batchnorm) is employed in both the generator and discriminator to improve the stability of GAN training. Batchnorm normalizes the input layer by adjusting it to have a mean of zero and a variance of one. Typically, it is applied after the hidden layer and before the activation layer.

DCGAN incorporates four primary activation functions: sigmoid, tanh, ReLU, and LeakyReLU.

Sigmoid is used in the final layer of the DCGAN discriminator since it performs binary classification, providing an output of 0 (indicating synthetic) or 1 (indicating real).

Tanh is similar to sigmoid but scales the output to the range [-1, 1], making it suitable for the generator's final layer. Therefore, input data for training should be preprocessed to fit within the range of [-1, 1].

ReLU (Rectified Linear Activation) returns 0 for negative input values and the input value for non-negative inputs. In the DCGAN generator, ReLU is employed for all layers except the output layer, which uses tanh.

LeakyReLU is an extension of ReLU that introduces a small negative slope (controlled by a constant, typically set to 0.2) for negative input values. In DCGAN, LeakyReLU activation is used in all discriminator layers, except for the final layer.

The training process involves the concurrent training of both the generator and discriminator networks.

Initial data preparation for training the DCGAN is necessary. Since the generator is not intended for classification, there is no need to split the dataset into training, validation, and testing sets. The generator requires input images in the format (60000, 28, 28), indicating there are 60,000 grayscale training images with dimensions of 28×28. The loaded data already adheres to the shape (60000, 28, 28) as it is grayscale.

To ensure compatibility with the generator's final layer activation using tanh, input images are normalized to the range of [-1, 1].

The primary goal of the generator is to produce realistic images and deceive the discriminator into perceiving them as real.

The generator takes random noise as input and generates an image closely resembling the training images. Since the objective is to generate grayscale images with a size of 28×28, the model architecture needs to ensure the generator's output has a shape of $28 \times 28 \times 1$.

To achieve this, the generator performs the following operations:

- Conversion of the 1D random noise (latent vector) to a 3D format using the Reshape layer.
- Iterative upsampling of the noise through the Keras Conv2DTranspose layers to match the desired output image size. In this case, the goal is to generate grayscale images with a size of 28 $\times$ 28.

The generator consists of key layers serving as its building blocks:

1. Dense (fully connected) layer: primarily used for reshaping and flattening the noise vector.
2. Conv2DTranspose: employed for upsampling the image during the generation process.
3. BatchNormalization: applied to stabilize the training process. It is positioned after the convolutional layer and before the activation function.

In the generator, ReLU activation is used for all layers except the output layer, which utilizes tanh activation.

For building the generator model, the Keras Sequential API is utilized. The process begins with the creation of a Dense layer to reshape the input into a 3D format, with the input shape specified in this layer.

Subsequently, BatchNormalization and ReLU layers are added to the generator model. Afterward, the previous layer is reshaped from 1D to 3D, and two upsampling operations are carried out using Conv2DTranspose layers with a stride of 2. These operations increase the size from 7 $\times$ 7 to 14 $\times$ 14 and ultimately to 28 $\times$ 28.

Following each Conv2DTranspose layer, a BatchNormalization layer is included, followed by a ReLU layer.

Finally, a Conv2D layer with a tanh activation function is included as the output layer.

Next, we delve into the implementation of the discriminator model.

The discriminator functions as a binary classifier that distinguishes whether an image is real or synthetic. Its primary objective is to accurately classify the provided images. However, there are a few notable distinctions between a discriminator and a conventional classifier: in the discriminator, the LeakyReLU activation function is employed and the discriminator confronts two categories of input images: real images from the training dataset, labeled as 1, and synthetic images generated by the generator, labeled as 0.

It is essential to note that the discriminator network is typically designed to be simpler or smaller compared to the generator. This is because the discriminator has a relatively easier task than the generator. In fact, if the discriminator becomes too powerful, it can impede the progress of the generator.

To construct the discriminator model, we have designed a function that takes input in the form of images, which can be either real images sourced from the training dataset or synthetic images generated by the generator. These images have dimensions of 28 $\times$ 28 $\times$ 1, and the function takes these dimensions (width, height, and depth) as arguments.

The construction of the discriminator model involves a sequence of layers, including Conv2D, BatchNormalization, LeakyReLU layers used twice for downsampling, a Flatten layer, and the application of dropout. In the final layer, we apply the sigmoid activation function to produce a single value for binary classification.

Loss computation plays a pivotal role in the training of both the generator and discriminator models in the DCGAN, as well as in any GAN architecture. Specifically, for the DCGAN under consideration, we employ a modified minimax loss and utilize the binary cross-entropy (BCE) loss function.

This involves calculating two separate losses: one for the discriminator and another for the generator. The Discriminator Loss is computed separately for the two sets of images (real and synthetic) that the discriminator evaluates, and the individual losses are then combined to yield the overall discriminator loss.

As for the generator loss, we aim to train the generator to maximize the probability of the discriminator incorrectly classifying the synthetic images as real. This approach employs the modified minimax loss.

For both the generator and discriminator models, we employ the Adam optimizer with a learning rate of 0.0002. Additionally, we utilize the Binary Cross-Entropy loss function for both the discriminator and generator.

The training process encompasses a total of 50 epochs.

After generating images using the DCGAN, the final step of our proposed method, illustrated in Figure 2, focuses on constructing models dedicated to discriminating between real and synthetic retina images.

In the proposed method, as illustrated in Figure 2, the second phase entails the creation of a model for distinguishing between generated and real images. This requires the availability of two distinct datasets. The first dataset comprises authentic retinal fundus images from the real world, while the second dataset consists of images produced by the DGCAN, also visualized in Figure 2. It is worth noting that the authentic real-world images used in the initial phase of the proposed method are the same images utilized here.

From these two sets of images, denoted as ”Generated Images” and ”Real Images” in Figure 2, a col-

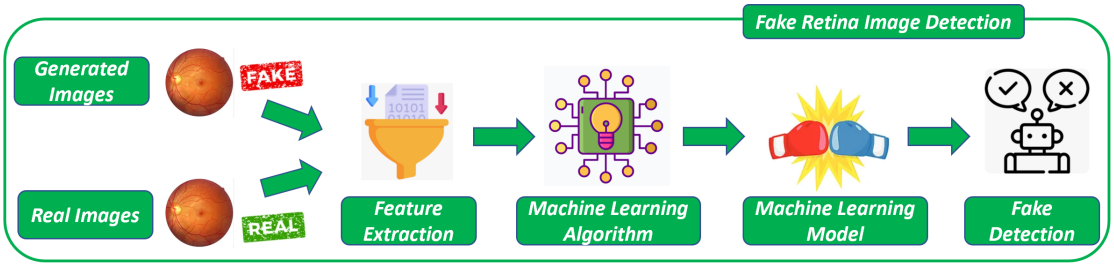

Figure 2: The step related to the synthetic retinal fundus images detection.

lection of numerical features is extracted, depicted as "Feature Extraction" in Figure 2. Specifically, the paper's experiments utilize the Simple Color Histogram Filter(Vijayan et al., 2023) for this purpose. This filter computes a histogram that represents the frequency of pixels in each image. Consequently, this filter generates 64 numerical features from each image.

After obtaining the feature set from both the generated and real images, these features are employed as inputs for a supervised machine learning algorithm, denoted as "Machine Learning Algorithm" in Figure 2. The objective is to construct a model capable of determining whether an image belongs to a synthetic or a real category.

Through the training of the machine learning algorithm with the extracted features, it acquires an understanding of the patterns and relationships between the features and the authenticity of the images. This enables the model to classify new images as either synthetic or real based on the acquired knowledge, as indicated by "Machine Learning Model" in Figure 2. The algorithm's training process involves providing it with labeled examples of images, along with their corresponding classification (synthetic or real), allowing the model to learn the decision boundaries between the two classes. Once the model is trained, it can be utilized to predict the authenticity of unseen images, as depicted by "Synthetic Detection" in Figure 2.

The effectiveness of the classifier is assessed by examining whether there is a significant distinction between the synthetic and original images. In contrast, if the machine learning models are unable to differentiate between the synthetic and original images, it suggests that the generated images closely resemble the originals.

To explore the evolution of image generation at various stages of GAN training, a model is constructed for each epoch (for a total of 50 models). This approach provides insights into whether the generated images become progressively more similar to the original images.

3 EXPERIMENTAL ANALYSIS

In the following, we present and elaborate on the outcomes of the conducted experimental analysis.

The primary objective of the experiment is to assess the potential impact of GANs on machine learning-based classification of retinal fundus images. To achieve this, we harnessed a DCGAN to generate a series of synthetic retinal fundus images. Subsequently, we trained a classifier with the aim of distinguishing between real-world retinal fundus images and their artificially generated counterparts.

The key focus of this investigation is to evaluate whether the exploited classifier could effectively differentiate between authentic and synthetic retinal fundus images. Given that the DCGAN generates a new dataset of retinal fundus images with each training epoch, we monitor the performance of the classifier throughout the training process. The objective was to determine whether the classifier's ability to distinguish between real-world and synthetic images would decline as the training advanced and the generated images presumably became more similar to real retinal fundus images.

By closely tracking the performance of the classifier, we could gauge their success in correctly identifying real images and distinguishing them from synthetic ones. A decrease in performance as training epochs increased would suggest that the classifiers encountered difficulties in distinguishing between authentic and synthetic retinal fundus images.

In summary, the experiment sought to assess the potential threat that GANs might pose to machine learning-based retinal fundus image classification by examining the classifiers' capacity to differentiate between real-world and artificially generated images as the GAN training unfolded.

For experimental purposes, we resort to the RetinaMNIST dataset, a publicly available dataset for research purposes[1], which is based on the DeepDRiD challenge. This dataset comprises 1,600 retinal fun-

[1]https://medmnist.com/

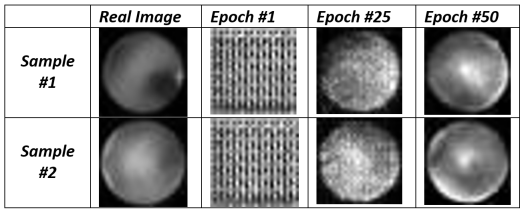

Figure 3: Different examples of synthetic retinal fundus images generated at different epochs, compared with the real images.

dus images related to the 5-level grading of diabetic retinopathy severity. The source images were centrally cropped and resized to dimensions of 3 $\times$ 28 $\times$ 28(Yang et al., 2021; Yang et al., 2023).

In the experimental analysis, the DCGAN underwent training for a total of 50 epochs. Each epoch required approximately 25 seconds to complete, leveraging the computational capabilities of an NVIDIA T4 Tensor Core GPU. During each epoch, the DCGAN generated a batch of 1000 synthetic retinal fundus images.

In Figure 3, we provide a visual representation of a set of images generated by the DCGAN at different epochs, alongside the original input images utilized for the DCGAN in the generation of synthetic retinal fundus images.

Analyzing the images presented in Figure 3, we focus on two distinct original input images, denoted as "Real Image Sample #1" and "Real Image Sample #2." It is evident that at epoch #1, the DCGAN generated images that essentially resembled noise, which aligns with the expected behavior. However, as we progress to the 25th epoch, the synthetic images derived from both "Sample #1" and "Sample #2" start to exhibit a closer resemblance to the real images. By the time we reach the 50th epoch, the images noticeably resemble the authentic ones.

To evaluate the performance of the classifier, we took into consideration several key metrics, including Precision, Recall, and F-Measure.

As a classifier, the J48 algorithm (Bhargava et al., 2013; Canfora et al., 2014) is utilized to construct a model with the objective of distinguishing between counterfeit and authentic retinal fundus images.

To construct the model, a composite dataset was employed, comprising real-world retinal fundus images and synthetic ones generated by the DCGAN for a specific epoch. This signifies that, for each epoch, we considered a dataset that encompassed both genuine retinal fundus images sourced from real-world applications and synthetic retinal fundus images generated by the DCGAN.

We adopted a strategy of building a model exploiting synthetic images obtained for various epochs to assess the performance and efficacy of the classifier in distinguishing between authentic and synthetic retinal fundus images at different stages of the training process. This method allowed to observe any variations in classifier performance as the DCGAN generated images that were progressively becoming more akin to real retinal fundus images during the course of the training epochs.

Table 1: Experimental analysis results for the 1, 25, and 50 epochs obtained with the J48 algorithm.

Epoch	Precision	Recall	F-Measure
1	1	1	1
25	0.970	0.970	0.970
50	0.972	0.971	0.971

We opted for a cross-validation approach with a value of k=10. The experimental analysis results are presented in Table 1. To conserve space, the results pertaining to three specific epochs are displayed: the initial epoch (labeled as *1* in the "Epoch" column), the midway epoch (labeled as *25* in the "Epoch" column), and the concluding epoch (labeled as *50* in the "Epoch" column). This selection allows for an examination of the overall trends.

Looking at Table 1, it is evident that at epoch 1, the J48 model achieves an F-Measure of 1. By epoch 25, this value drops slightly to 0.970, and at epoch 50, it remains relatively stable at 0.971. This implies that the performance of the J48 model remains largely consistent from epoch 25 to epoch 50.

The observed diminishing trend in performance as the epoch number increases aligns with expectations. This decline can be attributed to the fact that the GAN progressively enhances its ability to generate improved synthetic retinal fundus images with each successive epoch. However, as indicated by the results in Table 1, the decrease in performance is minimal but still noticeable. Consequently, the series of retinal fundus images can not be reliably distinguished by the classifier.

To gain a more comprehensive understanding of the classifiers' performance across the numerous epochs, Figure 4 illustrates the graphical representation of the F-Measure trend over the course of 50 epochs. It's worth noting that the decline in the metrics (Precision, Recall, and F-Measure) becomes no-

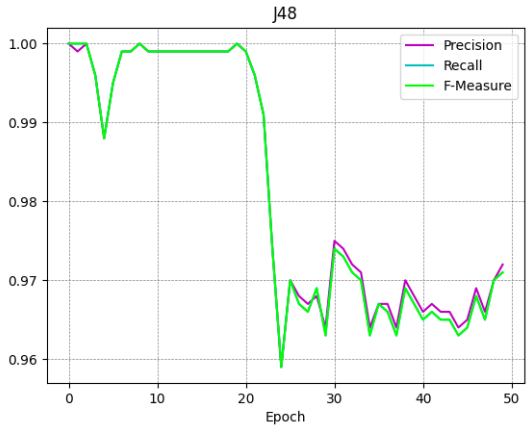

Figure 4: The Precision, Recall, and F-Measure trend for the 50 epochs.

ticeable around the 20-epoch mark.

Hence, we observe that a slight performance decline persists. While the classifier continues to deliver commendable results even with images generated after 50 epochs, it is still evident that a fraction of the synthetic images remains indistinguishable for the classifier.

In summary, the findings from the experimental analysis indicate that, at present, GANs do not present a substantial threat from a biomedical point of view, as the current classifiers are adept at effectively discerning between authentic and synthetic images. Nonetheless, it is important to acknowledge that a small fraction of images can still evade detection, which could potentially evolve into a threat in the future, especially in the realm of biomedical image classification.

4 CONCLUSION AND FUTURE WORK

Given the lifelike nature of images produced by GANs, it is imperative to evaluate their potential impact on image recognition systems, especially in the realm of biomedical image classification. This paper introduced a method aimed at assessing whether retinal fundus images generated by a DCGAN can be distinguished from genuine images. To accomplish this, we employed machine learning to construct a model capable of discerning between real and synthetic retinal fundus images. The experimental analysis uncovered that all the models achieved an F-Measure exceeding 0.95, demonstrating their proficiency in recognizing the majority of synthetic images. Nevertheless, it was observed that a subset of retinal fundus images managed to elude detection by the classifiers designed for synthetic image identification.

While GANs offer solutions to limited dataset challenges by generating synthetic data that closely resembles actual biomedical data, which can assist in training robust models and excel in producing high-quality medical images for medical image analysis, disease diagnosis, and treatment planning, they are also sensitive to input data quality and may inherit errors or biases. Adversarial attacks, involving deliberate manipulations to disrupt predictions, pose a significant threat to the practical application of machine learning. These attacks encompass evasion attacks, which manipulate only test data, and poisoning attacks, where the attacker introduces contaminated test and/or training data. A comprehensive understanding of adversarial attacks and the development of appropriate defenses are essential for upholding the reliability of machine learning applications.

Moreover, it is essential to acknowledge that while GANs offer significant advantages in biomedical image analysis, their utilization in critical medical applications necessitates thorough validation and careful consideration of ethical and regulatory concerns. The quality of generated images and their clinical relevance must undergo a rigorous assessment before implementing GAN-based solutions in real healthcare settings.

In future research endeavors, we intend to evaluate the effectiveness of the proposed approach using different types of GANs and various biomedical images acquired from diverse sources. Specifically, we aim to explore alternative types of biomedical images and assess the performance of various GAN architectures, such as conditional generative adversarial networks and cycle-consistent generative adversarial networks, in comparison to the DCGAN employed in this study.

ACKNOWLEDGEMENTS

This work has been partially supported by EU DUCA, EU CyberSecPro, SYNAPSE, PTR 22-24 P2.01 (Cybersecurity) and SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the EU - NextGenerationEU projects.

REFERENCES

Bacci, A., Bartoli, A., Martinelli, F., Medvet, E., and Mercaldo, F. (2018). Detection of obfuscation techniques in android applications. In *Proceedings of the 13th International Conference on Availability, Reliability and Security*, pages 1–9.

Bhargava, N., Sharma, G., Bhargava, R., and Mathuria, M. (2013). Decision tree analysis on j48 algorithm for data mining. *Proceedings of international journal of advanced research in computer science and software engineering*, 3(6).

Canfora, G., Medvet, E., Mercaldo, F., and Visaggio, C. A. (2014). Detection of malicious web pages using system calls sequences. In *Availability, Reliability, and Security in Information Systems*, pages 226–238. Springer.

Canfora, G., Medvet, E., Mercaldo, F., and Visaggio, C. A. (2015a). Detecting android malware using sequences of system calls. In *Proceedings of the 3rd International Workshop on Software Development Lifecycle for Mobile*, pages 13–20. ACM.

Canfora, G., Mercaldo, F., and Visaggio, C. A. (2013). A classifier of malicious android applications. In *Proceedings of the 2nd International Workshop on Security of Mobile Applications, in conjunction with the International Conference on Availability, Reliability and Security*.

Canfora, G., Mercaldo, F., and Visaggio, C. A. (2015b). Evaluating op-code frequency histograms in malware and third-party mobile applications. In *E-Business and Telecommunications*, pages 201–222. Springer.

Canfora, G., Mercaldo, F., and Visaggio, C. A. (2015c). Mobile malware detection using op-code frequency histograms. In *Proceedings of International Conference on Security and Cryptography (SECRYPT)*.

Cimitile, A., Martinelli, F., and Mercaldo, F. (2017). Machine learning meets ios malware: Identifying malicious applications on apple environment. In *ICISSP*, pages 487–492.

Fu, H., Cheng, J., Xu, Y., Wong, D. W. K., Liu, J., and Cao, X. (2018). Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE transactions on medical imaging*, 37(7):1597–1605.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11):139–144.

Huang, P., He, P., Tian, S., Ma, M., Feng, P., Xiao, H., Mercaldo, F., Santone, A., and Qin, J. (2022). A vit-amc network with adaptive model fusion and multiobjective optimization for interpretable laryngeal tumor grading from histopathological images. *IEEE Transactions on Medical Imaging*, 42(1):15–28.

Huang, P., Tan, X., Zhou, X., Liu, S., Mercaldo, F., and Santone, A. (2021). Fabnet: fusion attention block and transfer learning for laryngeal cancer tumor grading in p63 ihc histopathology images. *IEEE Journal of Biomedical and Health Informatics*, 26(4):1696–1707.

Huang, P., Zhou, X., He, P., Feng, P., Tian, S., Sun, Y., Mercaldo, F., Santone, A., Qin, J., and Xiao, H. (2023). Interpretable laryngeal tumor grading of histopathological images via depth domain adaptive network with integration gradient cam and priori experience-guided attention. *Computers in Biology and Medicine*, 154:106447.

Mercaldo, F., Nardone, V., Santone, A., and Visaggio, C. A. (2016). Hey malware, i can find you! In *Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2016 IEEE 25th International Conference on*, pages 261–262. IEEE.

Mercaldo, F. and Santone, A. (2020). Deep learning for image-based mobile malware detection. *Journal of Computer Virology and Hacking Techniques*, 16(2):157–171.

Orlando, J. I., Barbosa Breda, J., Van Keer, K., Blaschko, M. B., Blanco, P. J., and Bulant, C. A. (2018). Towards a glaucoma risk index based on simulated hemodynamics from fundus images. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11*, pages 65–73. Springer.

Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

Vijayan, T., Sangeetha, M., Kumaravel, A., and Karthik, B. (2023). Feature selection for simple color histogram filter based on retinal fundus images for diabetic retinopathy recognition. *IETE Journal of Research*, 69(2):987–994.

Yang, J., Shi, R., and Ni, B. (2021). Medmnist classification decathlon: A lightweight automl benchmark for medical image analysis. In *IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 191–195.

Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., and Ni, B. (2023). Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Scientific Data*, 10(1):41.

Zhou, X., Tang, C., Huang, P., Mercaldo, F., Santone, A., and Shao, Y. (2021). Lpcanet: classification of laryngeal cancer histopathological images using a cnn with position attention and channel attention mechanisms. *Interdisciplinary Sciences: Computational Life Sciences*, 13(4):666–682.