# AIMultimediaLab at ImageCLEFmedical GANs 2023: Determining "Fingerprints" of Training Data in Generated Synthetic Images

Notebook for the ImageCLEFmedical GANs Lab at CLEF 2023

Alexandra-Georgiana Andrei[1,*], Bogdan Ionescu[1]

[1]AI Multimedia Lab, Politehnica University of Bucharest, Romania

**Abstract**

This paper presents the participation of the AI Multimedia Lab to the 2023 ImageCLEF medical GANs task. The 2023 ImageCLEFmedical GANs task challenges participants to examine the hypothesis that generative models (Generative Adversarial Networks — GAN) are generating medical images that contain "fingerprints" of the real images used for the training of the network. We present our team's approach to tackle this task, consisting of a method that implies generating synthetic images from the development dataset. Subsequently, features were extracted from both sets of generated images (the one provided in the development dataset and the one we generated). A binary Support Vector Machine (SVM) classifier was trained using these features and the labels were predicted for the real images from the test dataset. Experimentation on the testing dataset show promising results.

**Keywords**

synthetic medical images, data augmentation, Generative Adversarial Networks, ImageCLEFmedical GANs, CT images

## 1. Introduction

Generative models are are becoming a powerful tool with immense potential for various applications, including in the medical field. These models can generate synthetic data that resembles real medical images, opening new doors for research, diagnosis, and treatment.

The integration of generative models offers unprecedented opportunities, enabling researchers to overcome dataset limitations and develop more robust algorithms improving anomaly detection and disease prediction. Image synthesis methods using Generative Adversarial Networks (GAN) have been studied and developed to obtain different types of medical images. Nie et al. [1] use context-aware GANs to generate computed tomography (CT) images from magnetic resonance images (MRIs), Yang et al. [2] describe a method to generate MR images using cGANs, Salehinejad et al. [3] use DCGANs to generate fake chest x-ray images, Frid-Adar et al. [4] generate synthetic CT images for liver lesion classification and Madani et al. [5] prove that GAN-based data augmentation achieved higher accuracy than traditional augmentation in chest X-rays. Ho et al. [6] show that diffusion probabilistic models can generate high-fidelity images

*CLEF 2023: Conference and Labs of the Evaluation Forum, September 18–21, 2023, Thessaloniki, Greece*

*Corresponding author.

alexandra.andrei@upb.ro (A. Andrei); bogdan.ionescu@upb.ro (B. Ionescu)

**Table 1**
Description of development and test datasets made available for the task

| | *Number of images* | |
|---|---|---|
| ***Dataset*** | **Generated** | **Real** |
| **Development** | 500 | 80 (used)<br>80 (not used) |
| **Test** | 10.000 | 200 (used and not used) |

comparable to those generated by GANs. For example, Puria Azadi et al. [7] use diffusion probabilistic models to synthesize high quality histopathology images of brain cancer.

However, the use of generative models in healthcare raises important considerations regarding privacy, confidentiality, and ethical implications. Medical data is highly sensitive, and protecting patient privacy is of highly importance. Given these aspects, the ImageCLEFmedical GAN 2023 task which is part of the 2023 ImageCLEF evaluation campaign [8], proposes a task related to the need of addressing concerns about the protection of personal information and patient confidentiality and to study if generative models can uphold ethical principles and ensure the privacy of individuals. Participants are given a set of generated images and a set of real images and are tasked to determine which real images were used for obtaining the generated set.

This paper describes the methods employed by AIMultimedia Lab for this task. We propose an approach that starts from the development dataset from which we generate a set of synthetic images, extract features, train a classifier and use it on the real images. The rest of the paper is structured as follows. Section 2 presents the task and the datasets. The proposed methods are presented in Section 3 and the results are presented in Section 4. Finally, the paper closes with Section 5, where we present the conclusions.

## 2. The 2023 ImageCLEFmedical GANs Task

ImageCLEFmedical GANs [9] task is a new challenge in the ImageCLEFmedical track. The objective of this task is to investigate the hypothesis that GANs produce medical images that retain identifiable characteristics ("fingerprints") from the original images utilized during training. . If this hypothesis proves to be true, artificial biomedical images might be subject to similar restrictions and limitations in terms of sharing and usage as actual sensitive medical data. Conversely, if the hypothesis is incorrect, GANs could potentially be utilized to generate extensive datasets of biomedical images that do not require adherence to ethical and privacy regulations.

**Data**. The task organizers provide a development set and a training set consisting of axial slices of 3D CT images of about 8,000 lung tuberculosis patients. Fig. 3 depicts examples of images provided with the task.

- The development dataset for task includes artificial images, real images which were marked as used or not used for training generative neural networks.
- The training set was created in similar way, the only difference is that the two subsets of

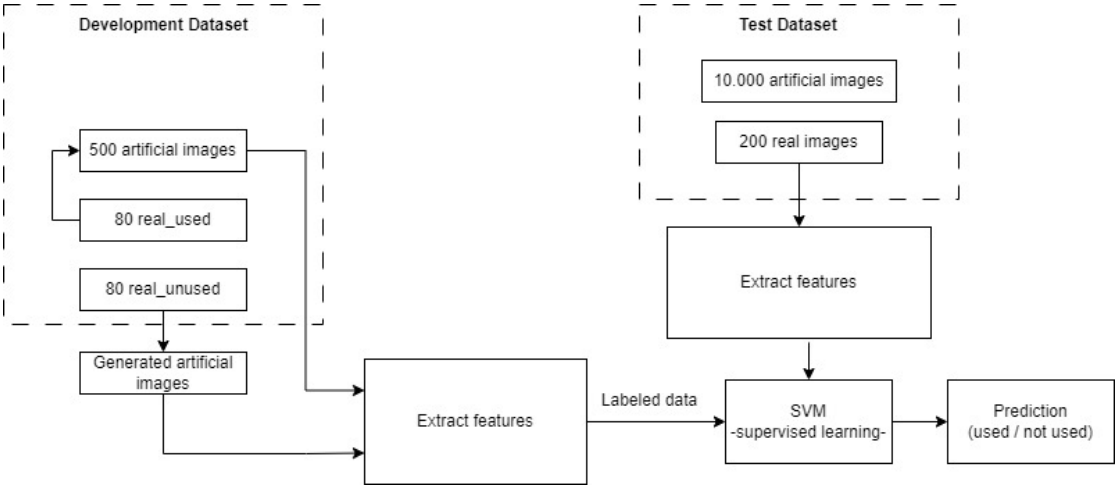


**Figure 1:** Overview of proposed approach for solving the task.

real images are mixed and no proportion of non-used and used ones has been disclosed. More information about the provided sets for the task are described in Table 1.

## 3. Proposed Methods

Two approaches are developed for addressing the task. They are depicted in Figure 1. Both approaches start by generating synthetic images from the real unused images provided in the development dataset. Subsequently, distinct descriptors/features are extracted and utilized to train a binary SVM classifier that we further used for identifying which of the 200 provided real images were used for generating the 10,000 artificial images from the test dataset.

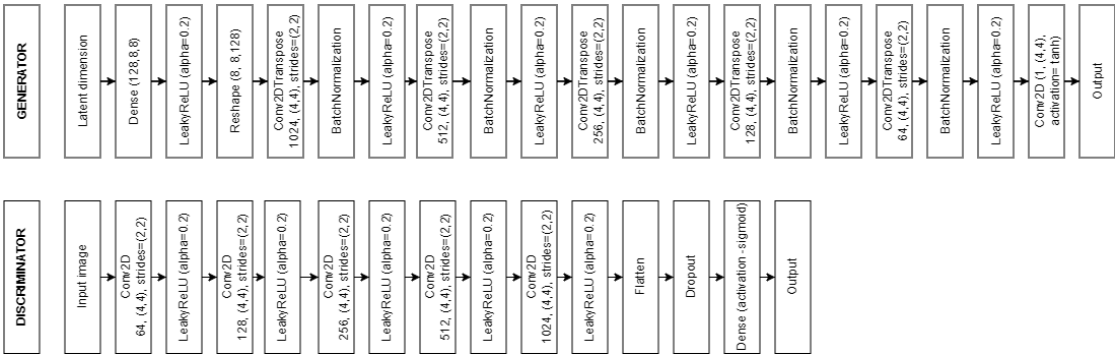


**Figure 2:** DCGAN architecture

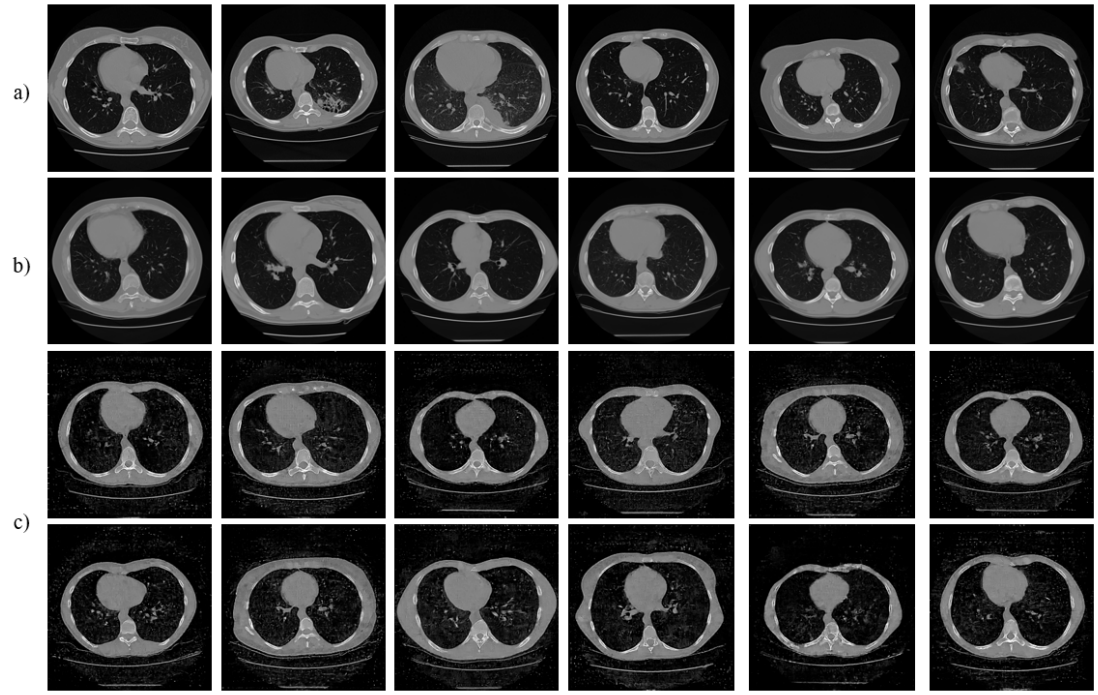


**Figure 3:** Example of images: a) real images from the dataset, b) synthetic images provided in the task' dataset, c) synthetic images generated by the method presented in this paper.

### 3.1. Generate Synthetic Images

Our research was based on the assumption that both the development and test sets of artificial images were produced utilizing the same Generative model so their characteristics may resemble. Consequently, we used the development dataset to generate synthetic images from the subset of 80 real unused images. We used a Deep Convolutional Generative Adversarial Network (DCGAN) to generate images as the ones presented in Figure 3 (c). The model is depicted in Figure 2. The generator consists of 5 deconvolutional layers with a kernel size (4,4), different number of filters for each layer and stride 2. Each deconvolutional layer is followed by a BatchNormalization layer and a LeakyRelu activation function with a slope of 0.2 and the last convolutional layer uses Tanh activation function. The discriminator contains 5 convolutional layer with kernel (4,4) and stride 2. Each convolutional layer is followed by a LeakyRelu activation layer. The last convolution layer of the discriminator is flattened and then fed into a sigmoid output. The model was trained using a batch size of 2 due to the small size of the data available for training. An Adam optimizer was used with learning rate 2e-4 ($\beta 1 = 0.5$). We used the 80 available images for training to generate 802 synthetic images that were further used as described in the diagram.

**Table 2**
Preliminary results

| Classifier | Features extraction method | F1-score | Accuracy | Precision | Specificity | Recall |
|---|---|---|---|---|---|---|
| SVM, radial kernel | hand-crafted | 0.54 | 0.45 | 0.46 | 0.25 | 0.65 |
| SVM, linear kernel | hand-crafted | 0.66 | 0.5 | 0.5 | 0 | 1 |
| SVM, radial kernel | deep-learning | 0.78 | 0.58 | 0.54 | 0.17 | 1 |
| SVM, linear kernel | deep-learning | 0.66 | 0.5 | 0.5 | 0 | 1 |

### 3.2. Features Extraction

Here is the step where our two approaches differ. Fist, we employed a a handcrafted feature extraction technique called Local Binary Pattern (LBP) to capture the local spatial patterns and the gray scale contrast of the images. Aditionally, we used a deep-learning approach utilizing a pre-trained VGG-16 convolutional network[10], which is available in Tensorflow library. We extracted characteristics from both sets of generated images, our set and the one provided in the development dataset, and from the 200 real images released for testing.

### 3.3. Prediction

The training process of the binary SVM classifier is outlined in Figure 1. It was trained on the features extracted from the two sets of generated images and tested on the features extracted from the 200 real images. To test the linear and non-linear relationship in the data we used both linear and kernel basis functions. The linear kernel was shown to be efficient just for the preliminary tests. While testing our method on the test set, using the linear kernel we only predicted that 9 images were used for training while the other 191 were classified as not used, so this proven that there is a non-linear relationship in the data.

## 4. Results and Discussion

**Dataset**.Both development and test sets were used. Development set consists of 500 artificial images, 80 real images used for training and 80 real images not used for training while the test set consists of 10,000 artificial images and 200 real images. Different evaluation metrics were used: f1-score, accuracy, precision, specificity, recall. The official evaluation metric for the task is F1-score.

### 4.1. Preliminary Tests

To validate the proposed method, we performed two types of preliminary runs on the development dataset before testing it on the test dataset. The first thing that was inspected was to verify that the features extracted from the two generated datasets are unique per class and can be used for classification. To verify this, both generated data sets were split into training and test sets, different data partitioning ratios were tested to train and test a binary SVM classifier

**Table 3**
Results on the test dataset.

| Method | f1_score | accuracy | precision | specificity | recall |
|---|---|---|---|---|---|
| run1: handcrafted feature extraction (LBP) | **0.626** | 0.54 | 0.527 | 0.31 | 0.77 |
| run 2: deep-learning features extraction | **0.585** | 0.47 | 0.4807 | 0.19 | 0.75 |

and in all cases accuracy values of 100% were obtained for both feature extraction methods. This confirmed that the proposed feature extraction methods can be used for a classification problem.
The method presented in this paper was first implemented and tested on the development dataset. In this case we are talking about a supervised classification method: the features extracted from the two sets of generated images were used to train the classifier, and the testing set consisted of the features extracted from the two types of real images provided in the development dataset. The preliminary results obtained are depicted in Table 2. The best preliminary result of F1-score of 0.78 of F1-score was obtained using a deep-learning feature extraction method and a SVM with radial kernel for classification.

## 4.2. Official Results

We have submitted two runs, one for each method presented previous section, as presented in the following:

- run1: uses the handcrafted feature extraction method (LBP).
- run2: uses a deep-learning feature extraction method

Our results are presented in Table 3 and the confusion matrices for the two runs are depicted in Figure 4. Best results were obtained using the handcrafted feature extraction method, the LBP. This can be explained as the fact that texture characteristics from the training images are maintained in the generated synthetic images.

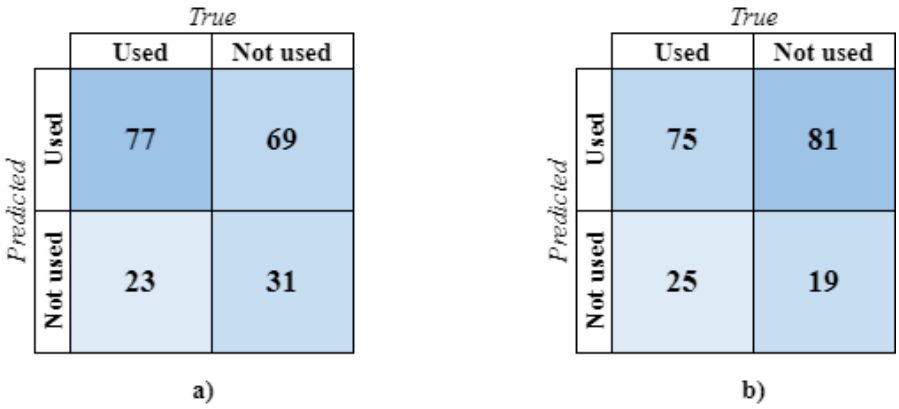


**Figure 4:** Confusion matrices: a) run1, b) run2

## 5. Conclusions

This paper describes the methods developed by AIMultimediaLab team for the ImageCLEFmedical GANs task. The purpose of the task was to determine the images from the real data set that were used to train the models that generated synthetic images. In our system, we have generated a new set of synthetic images from the development dataset, and extracted features that we used to train a classifier. The trained classifier was further used to predict the label (used/not used) for the 200 real images. The best result with an f1-score of 0.62 was obtained using the handcrafted feature extraction technique demonstrating that classical texture descriptors can be involved in analysing synthetic medical images.of our approach in practical applications involving the generation and analysis of artificial medical images.
In conclusion, 2023 ImageCLEFmedical GANs aim to bring attention to the possible privacy hazards associated with the utilization of synthetic medical data in practical settings and this paper presents a couple of methods for examining the existing hypothesis that GANs are generating medical images that contain the "fingerprints" of the real images used for generative network training.
To enhance the scope of this paper, additional exploration can be conducted by extracting and examining alternative hand-crafted features. This investigation aims to determine whether a combination of these features can provide more insights into the origin of the generated data. Furthermore, an opportunity for further development lies in the method proposed in this paper for generating synthetic images. By testing and implementing alternative methods, it is possible to generate images of superior quality.

## Acknowledgments

The contribution to this task is supported under project AI4Media, A European Excellence Centre for Media, Society and Democracy, H2020 ICT-48-2020, grant #951911.

## References

[1] D. Nie, R.Trullo, C. Petitjean, S. Ruan, D. Shen, Medical image synthesis with context-aware generative adversarial networks, in: International conference on medical image computing and computer-assisted intervention, Springer, 2017, pp. 417–425.
[2] Q. Y. et al., Mri cross-modality image-to-image translation, Scientific reports 10 (2020) 1–18.
[3] H. Salehinejad, S. Valaee, T. Dowdell, E. Colak, J. Barfett, Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks, in: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, 2018, pp. 990–994.
[4] M. F.-A. et al., Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification, Neurocomputing 321 (2018) 321–331.
[5] M. Moradi, A. Mandani, A. Karargyris, T. Syeda-Mahmood, Chest x-ray generation and

data augmentation for cardiovascular abnormality classification, in: Medical imaging 2018: Image processing, volume 10574, SPIE, 2018, pp. 415–420.

[6] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, in: Advances in Neural Information Processing Systems, volume 33, Curran Associates, Inc., 2020, pp. 6840–6851.

[7] P. A. M. et al., A morphology focused diffusion probabilistic model for synthesis of histopathology images, arXiv e-prints (2022) arXiv–2209.

[8] B. Ionescu, H. Müller, A. Drăgulinescu, W. Yim, A. Ben Abacha, N. Snider, G. Adams, M. Yetisgen, J. Rückert, A. Garcıa Seco de Herrera, C. M. Friedrich, L. Bloch, R. Brüngel, A. Idrissi-Yaghir, H. Schäfer, S. A. Hicks, M. A. Riegler, V. Thambawita, A. Storås, P. Halvorsen, N. Papachrysos, J. Schöler, D. Jha, A. Andrei, A. Radzhabov, I. Coman, V. Kovalev, A. Stan, G. Ioannidis, H. Manguinhas, L. Ştefan, M. G. Constantin, M. Dogariu, J. Deshayes, A. Popescu, Overview of ImageCLEF 2023: Multimedia retrieval in medical, socialmedia and recommender systems applications, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction, Proceedings of the 14th International Conference of the CLEF Association (CLEF 2023), Springer Lecture Notes in Computer Science LNCS, Thessaloniki, Greece, 2023.

[9] A. Andrei, A. Radzhabov, I. Coman, V. Kovalev, B. Ionescu, H. Müller, Overview of ImageCLEFmedical GANs 2023 task – Identifying Training Data "Fingerprints" in Synthetic Biomedical Images Generated by GANs for Medical Image Security, in: CLEF2023 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Thessaloniki, Greece, 2023.

[10] S. Karen, Z. Andrew, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).