Research Article GU J Sci, Part A, 11(3): 497-506 (2024) 10.54287/gujsa.1501098

Gazi University
Journal of Science
PART A: ENGINEERING AND INNOVATION
http://dergipark.org.tr/gujsa

TB-SMGAN: A GAN Based Hybrid Data Augmentation Framework on Chest X-ray Images and Reports

Hasan Berat ÖZFİDAN[1] Mehmet Ulvi ŞİMŞEK[1*]

[1] ASELSAN, Ankara, Türkiye

Keywords	
Generative Augmentation	
Latent Space Manipulation	
Generative Adversarial Networks	
Chest x-ray Data Augmentation	

Abstract

Data augmentation is a common practice in image classification, employing methods such as reflection, random cropping, re-scaling, and transformations to enhance training data. These techniques are prevalent when working with extended real-world datasets, focusing on improving classification accuracy through increased diversity. The use of Generative Adversarial Networks (GANs), known for their high representational power, enables learning the distribution of real data and generating samples with previously unseen discriminative features. However, intra-class imbalances in augmentations are problematic for conventional GAN augmentations. Hence, we propose a framework named Text-Based Style-Manipulated GAN augmentation framework (TB-SMGAN) aims to leverage the generative capabilities of StyleGAN2-ADA. In this framework, we utilize StyleCLIP to control disentangled feature manipulations and intra-class imbalances. We enhance the efficiency of StyleCLIP by fine-tuning CLIP with x-ray images and information extractions from corresponding medical reports. Our proposed framework demonstrates an improvement in terms of mean PR-AUC score when employing the text-based manipulated GAN augmentation technique compared to conventional GAN augmentation.

Cite

Özfidan, H. B., & Şimşek, M. U. (2024). TB-SMGAN: A GAN Based Hybrid Data Augmentation Framework on Chest X-ray Images and Reports. *GU J Sci, Part A*, *11*(3), 497-506. doi:10.54287/gujsa.1501098

Author ID (ORCID Number)	
0009-0006-0950-2017	Hasan Berat ÖZFİDAN
0000-0001-8017-4753	Mehmet Ulvi ŞİMŞEK

Article Process	
Submission Date	13.06.2024
Revision Date	06.08.2024
Accepted Date	12.09.2024
Published Date	28.09.2024

1. INTRODUCTION

Accurately classifying medical images is crucial for early diagnosis and effective treatment, but limited training data remains a critical obstacle (Liu et al., 2023). Traditional data augmentation techniques like geometric transformations and color jittering are used to expand datasets in literature. Their ability to introduce clinically relevant features is essential for applying medical tasks properly. Nevertheless, these methods are constrained to precise classification within medical images (Shorten & Khoshgoftaar, 2019).

Several approaches are explored to address data scarcity in medical image classification, each with its own set of limitations. One such approach is Conventional Data Augmentations, such as geometric transformations and color jittering, which are employed to expand datasets. However, these methods may not introduce clinically relevant features necessary for accurate diagnosis. Furthermore, they have the drawback of creating unrealistic augmentations that can potentially lead to biased models (Shorten & Khoshgoftaar, 2019; Jablonski et al., 2022; Wang & Qi, 2022). Another strategy is Self-Supervised Learning, which holds promise by learning representations from unlabeled data. Despite its potential, aligning learned features with specific diagnosis tasks proves to be challenging, and the performance in medical domains often falls short of optimal levels (Benčević et al., 2022; Ke et al., 2022; Hochberg et al., 2022). Additionally, Transfer Learning, where leveraging pre-trained models on large-scale datasets can enhance performance on smaller medical datasets. However, the presence of domain mismatch, characterized by differences between the training and target

*Corresponding Author, e-mail: mehmetulvi@gmail.com

datasets, poses a substantial challenge and can significantly impact the model's generalizability (Dao et al., 2022; 2023; Kariuki et al., 2023). Another notable approach is the use of the Deep Synthetic Minority Oversampling Technique (SMOTE), designed to address class imbalance by oversampling minority classes. Despite its effectiveness in tackling imbalanced datasets, this technique has drawbacks, including the risk of overfitting and the potential generation of unrealistic samples (Tarawneh et al., 2022; Altwaijry, 2023). Lastly, Physics-Informed Neural Networks offer an avenue to improve model generalizability by integrating physical constraints into the learning process. However, a significant challenge lies in the incorporation of accurate and relevant physical models specifically tailored for medical images (Wang et al., 2021; Islam & Mondal, 2019).

Considering these limitations, a robust data augmentation technique is required to solve this problem. This paper introduces a novel Text-Based Style-Manipulated GAN augmentation framework (TB-SMGAN) that overcomes these limitations by leveraging the combined capabilities of StyleGAN2-ADA (Karras et al., 2020) and StyleCLIP (Patashnik et al., 2021). Utilizing StyleCLIP's ability to manipulate disentangled features based on textual descriptions, TB-SMGAN extracts key clinical findings from medical reports to fine-tune the generation of disease-specific variations within each class. This allows us to control precisely over intra-class imbalances and targeted generation of discriminative features relevant to specific medical diagnoses. To further refine TB-SMGAN's ability to generate clinically relevant augmentations, we fine-tune Contrastive Language-Image Pre-Training (CLIP) (Radford et al., 2021) with x-ray images and extracted information from corresponding medical reports. This domain-specific adaptation improves the framework's effectiveness in medical tasks. Our framework demonstrates significantly improved classification performance, as measured by PR-AUC score, compared to standard GAN augmentation. This improvement showcases the potential of TB-SMGAN for overcoming data scarcity and enhancing the accuracy and generalizability of medical image classification models, ultimately contributing to improved patient care.

Our proposed framework presents several contributions to the field of medical Generative Adversarial Network (GAN) augmentation:

1) Rule-Based Information Extraction from X-ray Reports: We develop a novel rule-based algorithm for accurately extracting relevant information from X-ray reports. This extracted information provides additional context for medical image analysis and enhances the effectiveness of downstream tasks.
2) Fine-Tuning CLIP for Medical Domain Adaptation: We fine-tune CLIP using various text extraction methods specifically tailored for the medical domain. This domain adaptation ensures the learned representations are relevant and informative for medical applications, leading to improved performance in subsequent tasks.
3) Text-Based Latent Space Manipulations for Medical Data Augmentation: This work introduces a novel approach that utilizes text-based information to manipulate the latent space of GANs. This approach enables the generation of synthetic medical data that is not only realistic but also semantically aligned with the extracted textual information, further enriching the training dataset and improving the generalizability of deep learning models trained on augmented data.

These contributions collectively address essential challenges within the domain of medical image analysis, offering substantial advancements that contribute to the refinement of deep learning models. The results shows that the proposed framework enhances accuracy and efficacy, establishing a more robust foundation for diverse clinical applications.

The rest of the paper is organized as follows: Section II overviews and discusses the related works in the literature. Then the details of TB-SMGAN framework are given in section III. We demonstrate the comparative results in performance evaluations in section IV. Finally, Section V summarizes key findings and the outputs of the paper.

2. RELATED WORKS

GAN is a type of deep learning model capable of generating synthetic data that is statistically similar to the real data (Alqahtani et al., 2021). This is achieved by training two neural networks in an adversarial manner: a generator network that attempts to create realistic synthetic data, and a discriminator network that attempts to distinguish between real and synthetic data. This adversarial process allows the generator to progressively learn to generate increasingly realistic and diverse data (Li et al., 2022).

In recent years, GAN-based data augmentation has emerged as a powerful alternative to classical data augmentation techniques (Lacan et al., 2023) One such approach is proposed to populate training datasets with synthetic data (Bowles et al., 2018) This approach utilizes the Progressive Growing of GANs (PGGAN) architecture for effective modeling of the input data distribution. The authors compare classical and GAN augmentation across various datasets to assess the generalizability of their synthetic data augmentation strategy. Additionally, they investigated the effects of varying augmentation strengths on model performance. Experiments demonstrate that GAN augmentation improves classification accuracy compared to classical data augmentation techniques.

Medical GAN augmentation emerges as a beneficial tool for enhancing the performance of medical image classification tasks. One study collected data from Sheba Medical Center for three specific diseases and experimented with DCGAN, ACGAN, and ACGAN discriminator for data augmentation (Frid-Adar et al., 2018). The best performance is observed when the classifier is trained with additional generated images alongside real data. Similarly, another study compares both classical and GAN data augmentation for pneumonia recognition and finds that the DCGAN-augmented classifier achieved the highest accuracy, recall, and F1 score (Kora Venu & Ravula, 2020). Low classification accuracies in medical datasets are attributed to class imbalance (Deepshikha & Naman, 2020). Researchers address this by balancing the dataset with samples generated by a DCGAN, leading to improved classification performance. Moreover, another study reported superior classification performance for underrepresented classes augmented with GANs compared to classical augmentation techniques (Sundaram & Hulkund, 2021). However, this work only focused on ROC-AUC score, which is insensitive to data imbalance. Our study addresses this limitation by reporting PR-AUC scores, providing a more comprehensive and robust evaluation of the proposed approach. We leverage the conditional StyleGAN2-ADA (Karras et al., 2020) model for synthetic data generation, further exploring its potential for enhancing medical image classification.

The manipulation of latent spaces within GANs has emerged as a rapidly evolving field with significant potential for medical applications. However, the current research landscape exploring this technique in medical contexts remains relatively sparse, with limited investigations into its full capabilities.

One notable contribution employed StyleGAN architecture for manipulating latent spaces, achieving the generation of synthetic CT images corresponding to T2-weighted MR images and vice versa (Fetty et al., 2020). This work demonstrated the feasibility of cross-modality image synthesis through latent space walks between modalities. However, deeper investigations into the disentanglement properties of the style transfer process were not undertaken, leaving this area open for future research.

3. TB-SMGAN: A GAN BASED HYBRID DATA AUGMENTATION FRAMEWORK

In this section, we present the TB-SMGAN framework, which is shown Figure 1, the components of the framework are explained in detail.

3.1. Proposed TB-SMGAN System Architecture

We use a GAN framework, StyleGAN2-ADA, to generate realistic samples with a limited representational power dataset. The proposed adaptive discriminator augmentation technique stabilizes GAN training with tiny datasets and allows for diverse and realistic output generation without augmentation leakage or distortion (Karras et al., 2020). Our framework utilizes GAN augmentation to generate synthetic data samples and augment the dataset. Compared to traditional data augmentation techniques, GAN augmentation provides creative and realistic sample generation, preserving distinctive features with enough variation and fidelity. Therefore, we utilize StyleGAN2-ADA for both GAN augmentation and text-based style-manipulated augmentation experimentations.

Recent developments from OpenAI enable us to utilize the representational power of the CLIP pre-trained model for latent space manipulation (Radford et al., 2021). StyleCLIP enables semantic meaningful latent space manipulation with text guidance. Although several approaches enable text-based style manipulations in StyleCLIP, we focus solely on the input-agnostic global direction method, which does not require further optimization. For a given pair of text that defines neutral and target attributes, the global direction method exploits the colinearity between CLIP's image embedding space and StyleGAN2-ADA's style space for the difference of the given text pair. Since the global direction method offers style manipulation on real images, we train the encoder4editing framework for manipulating real x-ray images (Tov et al., 2021). After training, we combine the pre-trained StyleGAN2-ADA generator, encoder4encoder and StyleCLIP in order to create text-based style-manipulated augmentations.

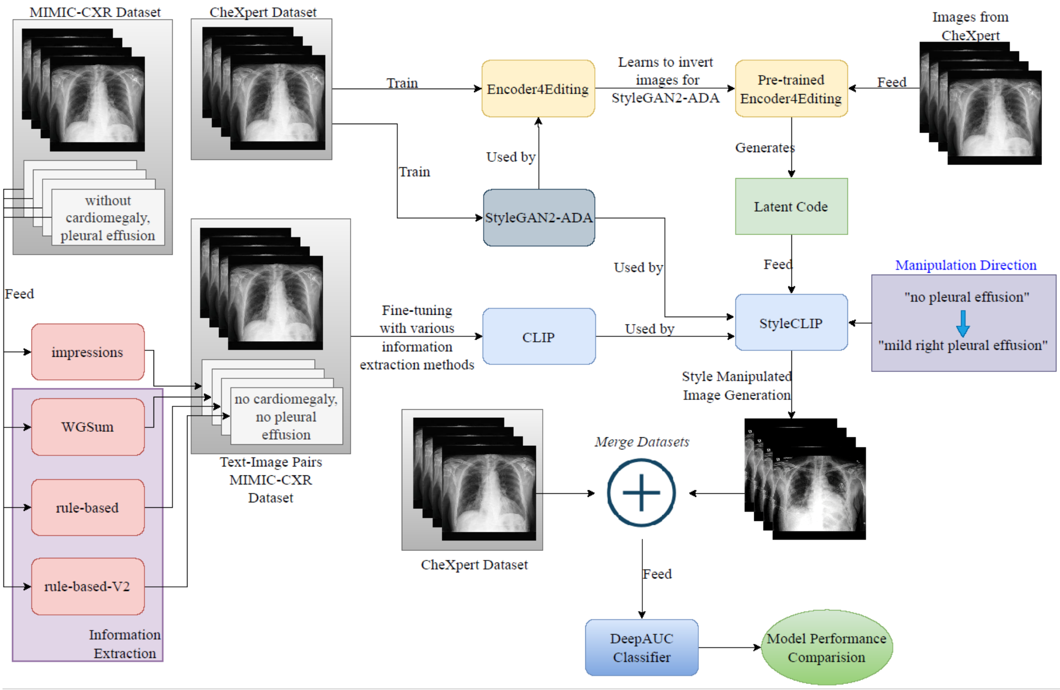

Figure 1. *Style manipulated data augmentation design overview*

During our initial experiments, we observed that the global optimization method does not yield the desired image manipulations in the context of chest X-rays since StyleCLIP using default CLIP weights. We infer this is due to the lack of medical representation in CLIP's training set. To address this issue, we fine-tune CLIP with medical image-text pairs originating from the MIMIC-CXR dataset (Johnson et al., 2019). We improve the fine-tuning process by incorporating NLP techniques for medical reports.

To create more informative reports, we extract entities using scispaCy, which contains spaCy models for processing biomedical-specific textual data (Neumann et al., 2019; Honnibal et al., 2020). In our experiments, we use the entities extracted by the encore sci scibert modell, which is a complete spaCy pipeline built on top of a transformer and pre-trained on a large biomedical corpus consisting of nearly 785,000 words. We use only the entities extracted for each report, instead of the full report, and also use dependency relations between words. We modify the entities to create less complex semantics for each report as shown in Algorithm 1 . For instance, considering an X-ray report stating "There is no focal consolidation, pleural effusion, or pneumothorax." Our rule-based approach combines entities linked to the word "No" directly or indirectly, resulting in "no focal consolidation", "no pleural effusion" and "no pneumothorax" as modified entities.

However, the results of our experiments show that the rule-based approach performs poorly for some classes. To address this issue, we analyze our rule-based approach to refine the disentanglement of disease-specific features, particularly for Edema and Consolidation classes. These classes present challenges due to their overlapping textual semantics. To overcome this, we introduce new indicator words and name the fine-tuning strategy "rule-based-V2".

GU J Sci, Part A 11(3) 497-506 (2024) 10.54287/gujsa.1501098

Algorithm 1 Rule-Based Information Extraction

Require: Medical report text T, scispaCy model M, list of words affecting semantic complexity L
Ensure: Modified list of entities E'

```
Extract entities E from T using M
Construct dependency relation graph G for T
E' ← ∅
for all entity e ∈ E do
    Find words W in L directly or indirectly related to e in G
    for all word w ∈ W do
        if relation between w and e then
            Add w prefix to e
            E' ← E' ∪ {e}
        else
            Add e to E'
        end if
    end for
end for
return E'
```

To quantify the effectiveness of style manipulations, we utilize DeepAUC, the top-performing solution in the CheXpert competition (Yuan et al., 2021) for evaluating the manipulated images.

To connect the dots formally, we incorporated all of the aforementioned approaches into a framework named TB-SMGAN. Our proposed framework consists of several steps and uses various models and methods for information extraction, image generation, and classification as in Figure 1. TB-SMGAN proceeds with the following steps:

1) Train StyleGAN2-ADA: The algorithm begins by training a StyleGAN2-ADA model, denoted as (G), on the set of real chest X-ray images $X_{CheXpert}$
2) Train Encoder4Encoder for GAN Inversion: Next, an Encoder4Encoder model, denoted as (E), is trained using the previously trained StyleGAN2-ADA model ($G_{CheXpert}$) and the set of real chest X-ray images ($X_{CheXpert}$).
3) Extract Information and Fine-tune CLIP: For each information extraction method (i) in the set (I) consists of WGSum, RuleBasedV1, RuleBasedV2 and Impressions, and for each pair of X-ray image and report (x,r) in ($X_{MIMIC-CXR}$, $R_{MIMIC-CXR}$) dataset, the algorithm fine-tunes a CLIP model, denoted as (T_i), using the information extracted from the report (F_i(r) and the X-ray image (x).
4) Determine Style Manipulation Direction: For each information extraction method (i) and each disease (d) in the set of diseases (mathcal{D}), the algorithm computes a manipulation direction (Δ_{z+d}) using the StyleCLIP model (S_i) which includes fine-tuned CLIP (T_i), the positive and negative text prompts for the disease (P_{+d}) and (P_{-d}) respectively.
5) Generate Style-Manipulated Images: The algorithm generates manipulated images for each information extraction method (i), each disease (\), and each X-ray image (x) in the set of X-ray images ($X^d_{cheXpert}$) with only disease (d) by applying the manipulation direction (Δ_{z+d}) to (S_i(x,$X_{CheXpert}$) which generates a synthetic image ($\hat{x}_{+d}$)
6) Data Selection with Classifier: Select the top performing style manipulated synthetic images for each disease (d) in the set of diseases (D) by comparing PR-AUC scores using DeepAUC classifier trained on X-ray images and labels ($X_{CheXpert}$, $L_{CheXpert}$).
7) Merge Datasets: The algorithm then merges the original dataset ($X_{CheXpert}$, $L_{CheXpert}$) with the set of manipulated images and corresponding labels ($\hat{x}_{+d}$, $\hat{l}_{+d}$) to create a new dataset ($\hat{X}_{\square}$, $\hat{L}_{\square}$).
8) Evaluate Representation Power of Augmented Dataset: Finally, the algorithm trains a DeepAUC model on the augmented dataset ($\hat{X}_{\square}$, $\hat{L}_{\square}$) and reports the results.

Our proposed method represents a sophisticated approach to data augmentation, leveraging advanced models and techniques to generate new, manipulated images that are capable of improving the representational power of the final augmented dataset.

GU J Sci, Part A 11(3) 497-506 (2024) 10.54287/gujsa.1501098

4. EXPERIMENTAL RESULTS

4.1. Datasets

In this paper, we employ the CheXpert dataset (Irvin et al., 2019), comprising about 225,000 chest X-ray images from Stanford University Medical Center. The dataset includes 14 classes, with uncertain labels categorized as positive, negative, or uncertain.

On our framework, we focus on five specific classes—Atelectasis, Cardiomegaly, Consolidation, Edema, and Pleural Effusion—central to the challenges in the CheXpert competition (Irvin et al., 2019). By excluding multi-class labeled samples, we refined the dataset to approximately 85,000 samples, optimizing our model's training for more focused and efficient learning in these specific classes.

4.2. TB-SMGAN Framework

This section presents the experimental results of our proposed method for augmenting chest X-ray datasets using text-based style manipulations. We evaluate the representation capacity of the augmentations by DeepAUC classifier. The key focus for classification performance evaluation is on PR-AUC scores instead of ROC-AUC scores. This metric provides a robust evaluation of our methodology, including its effectiveness in addressing class imbalance issues often encountered in medical image analysis tasks.

Table 1 shows the PR-AUC scores of the DeepAUC classifier for different fine-tuning strategies on only synthetic data generated by TB-SMGAN. The DeepAUC classifier used for evaluation is trained on the CheXpert dataset without any augmentations. Each row represents a distinct fine-tuning strategy along with plus or minus signs that indicate positive or negative manipulation direction. Columns denote the PR-AUC scores for each disease. The "Mean AUC" column presents the average PR-AUC across all diseases. Results demonstrate significant variations in performance across strategies. Notably, rule-based (+) strategy achieves the highest PR-AUC for Cardiomegaly emphasizing the effectiveness of incorporating rule-based information during fine-tuning. The impression (+) strategy leads in PR-AUC for Consolidation and Atelectasis, underlining the importance of specific fine-tuning strategies for different diseases.

Table 1. *Classification results on only synthetic data generated by TB-SMGAN*

Fine-Tuning Strategy	**Cardiomegaly (PR-AUC)**	**Edema (PR-AUC)**	**Consolidation (PR-AUC)**	**Atelectasis (PR-AUC)**	**Pleural-Effusion (PR-AUC)**	**Mean AUC**
rule-based (-)	0,147	0,198	0,196	0,194	0,193	0,186
impression (-)	0,191	0,203	0,159	0,169	0,175	0,179
rule-based-V2 (-)	0,234	0,228	0,126	0,169	0,227	0,197
original-CLIP (-)	0,151	0,41	0,225	0,197	0,411	0,279
WGSum-generated (-)	0,254	0,241	0,207	0,247	0,24	0,238
randomized (-)	0,19	0,269	0,182	0,261	0,312	0,243
inverted	0,241	0,294	0,219	0,221	0,277	0,251
rule-based (+)	**0,579**	0,235	0,17	0,364	0,432	0,356
impression (+)	0,27	0,232	**0,353**	**0,516**	0,455	**0,365**
rule-based-V2 (+)	0,208	0,382	0,222	0,281	**0,619**	0,342
original-CLIP (+)	0,339	0,38	0,336	0,26	0,205	0,304
WGSum-generated (+)	0,159	**0,579**	0,281	0,221	0,315	0,311
randomized (+)	0,318	0,181	0,288	0,217	0,187	0,238

Table 2 presents the results of style-manipulated GAN augmentations. The augmented dataset includes the full CheXpert dataset and the corresponding generated data, depending on the fine-tuning strategy. Experiments are conducted for various fine-tuning strategies. "Pure-dataset" and "randomized" strategies are included for benchmarking and sanity-check purposes respectively. On the one hand, "Pure-dataset" demonstrates the results of the classifier on pure CheXpert dataset without any augmentations. On the other hand, "randomized" indicates the fine-tuning strategy which uses reports with random word order. This strategy demonstrates the significance of the word order in the context of fine-tuning. Additionally, "Inverted" fine-tuning strategy demonstrates the results for the inverted images originating from Encoder4Encoder. Moreover, "StyleGAN2-Augmentation" is included to compare GAN augmentation with our proposed framework. This strategy includes class conditional image generation of StyleGAN2-ADA. In addition to the singular fine-tuning strategies, we experiment with the combinations of fine-tuning strategies. The ensemble method unifies the generated samples from various training strategies with maximum PR-AUC scores based on Table 1. For example, we include synthetic Cardiomegaly samples generated from the rule-based (+) method and synthetic Edema samples generated from the WGSum-Generated (+) method. To account for the unavailability of an impressions section for other datasets, we also employ an ensemble-without-impression method, in which we ignore the data created by fine-tuning strategies using impressions and select the data with the maximum PR-AUC scores. For instance, we include data created by the original-CLIP (+) method for Consolidation, and we select data generated by the rule-based-V2(+) method instead of the impression (+) method. Based on the results in Table 2, the TB-SMGAN(ensemble) framework performs slightly better than GAN augmentation and performs the best among all manipulation strategies in terms of PR-AUC score.

Table 2. *Classification performance results of fine-tuning strategies on text-based style manipulated GAN augmented dataset*

Fine-Tuning Strategy	**Cardiomegaly (PR-AUC)**	**Edema (PR-AUC)**	**Consolidation (PR-AUC)**	**Atelectasis (PR-AUC)**	**Pleural-Effusion (PR-AUC)**	**Mean AUC**
pure-dataset	0,719	0,745	0,615	0,633	0,842	0,711
rule-based (+), inverted and rule-based-V2 (+)	0,707	0,756	0,592	0,644	0,847	0,709
impression (+)	0,688	0,766	0,562	0,68	0,855	0,71
original-clip (+)	**0,738**	0,779	0,566	0,667	0,843	0,719
rule-based-V2 (+)	0,658	0,8	0,643	0,66	0,841	0,72
rule-based (-)	0,641	0,782	0,728	0,624	0,837	0,722
inverted	0,7	0,788	0,66	0,651	0,841	0,728
rule-based (+)	0,684	0,796	0,659	0,693	0,821	0,731
WGSum-generated (-)	0,659	0,791	0,671	0,698	0,834	0,731
randomized (+)	0,678	**0,818**	0,614	0,712	0,835	0,731
rule-based-V2 (-)	0,649	0,787	0,703	0,707	0,838	0,737
randomized (-)	0,682	0,791	0,688	0,701	0,827	0,738
WGSum-generated (+)	0,687	0,79	0,669	0,691	0,852	0,738
ensemble-without-impression	0,694	0,796	0,69	0,701	0,818	0,74
impression (-)	0,664	0,796	0,758	0,682	0,823	0,744
StyleGAN2-Augmentation	0,71	0,772	0,676	0,702	**0,863**	0,745
original-clip (-)	0,67	0,81	0,706	0,727	0,827	0,748
rule-based (+) and inverted	0,723	0,796	0,658	**0,743**	0,847	0,753
TB-SMGAN (ensemble)	0,672	0,802	**0,783**	0,667	0,845	**0,754**

5. CONCLUSION

GANs have demonstrated significant capabilities in representing complex data distributions. By effectively capturing the underlying variance of real-world data, GANs enable the generation of synthetic images with highly discriminative features. This work leverages the generative power of StyleGAN2-ADA to perform data augmentation for medical image datasets. Moreover, we introduce a text-based style manipulated GAN augmentation technique named TB-SMGAN for the medical domain. We utilize DeepAUC, the top solution of the CheXpert competition, to demonstrate the effectiveness of our GAN augmentation technique. Our methodology reveal that the classification performance of TB-SMGAN outperforms the classical GAN augmentation technique due to the increased representation of augmentations created by TB-SMGAN.

While our findings are promising, there are several directions for future research. First, it would be beneficial to explore the application of our proposed framework to other domains beyond medical imaging, to assess its generalizability. Second, developing a consistent measurement methodology for the quality of the dataset could yield valuable insights. Lastly, as text-based manipulated GAN augmentation has shown promise, future work could focus on refining this technique and exploring its potential in other contexts. This could involve the incorporation of more complex textual information or the development of more sophisticated text-to-image translation methods.

AUTHOR CONTRIBUTIONS

Conceptualization, methodology, manuscript-review, editing supervision, Mehmet Ulvi ŞİMŞEK; field and laboratory works, sources, software, data curation and visualization Hasan Berat Özfidan; research, validation, formal analysis, manuscript-original draft and funding, both authors. All authors have read and legally accepted the final version of the article published in the journal.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Alqahtani, H., Kavakli-Thorne, M., & Kumar, G. (2021). Applications of generative adversarial networks (gans): An updated review. *Archives of Computational Methods in Engineering*, *28*, 525-552. https://doi.org/10.1007/s11831-019-09388-y

Altwaijry, N. (2023). Probability-based synthetic minority oversampling technique. *IEEE Access*, *11*, 28831-28839. https://doi.org/10.1109/ACCESS.2023.3260723

Benčević, M., Habijan, M., Galić, I., & Pizurica, A. (2022, August 29 - September 02). *Self-supervised Learning as a Means to Reduce the Need for Labeled Data in Medical Image Analysis*. In: Proceedings of the 30th European Signal Processing Conference (EUSIPCO 2022) (pp. 1328-1332). Belgrade, Serbia. https://doi.org/10.23919/EUSIPCO55093.2022.9909542

Bowles, C., Chen, L., Guerrero, R., Bentley, P., Gunn, R., Hammers, A., Dickie, D. A., Hernández, M. V., Wardlaw, J., & Rueckert, D. (2018). Gan augmentation: Augmenting training data using generative adversarial networks. https://doi.org/10.48550/arXiv.1810.10863

Dao, H. N., Quang, T. N., & Paik, I. (2022, October 26-28). *Transfer Learning for Medical Image Classification on Multiple Datasets using PubMedCLIP*. In: Proceedings of the 2022 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia) (pp. 1-4). Yeosu, Korea, Republic of. https://doi.org/10.1109/ICCE-Asia57006.2022.9954669

Dao, H. N., Nguyen, T., Mugisha, C., & Paik, I. (2023). A Multimodal Transfer Learning Approach Using PubMedCLIP for Medical Image Classification. *IEEE Access*, *12*, 75496-75507. https://doi.org/10.1109/ACCESS.2024.3401777

Deepshikha, K., & Naman, A. (2020). Removing Class Imbalance using Polarity-GAN: An Uncertainty Sampling Approach. https://doi.org/10.48550/arXiv.2012.04937

Fetty, L., Bylund, M., Kuess, P., Heilemann, G., Nyholm, T., Georg, D., & Löfstedt, T. (2020). Latent space manipulation for high-resolution medical image synthesis via the StyleGAN. *Zeitschrift für Medizinische Physik*, *30*(4), 305-314. https://doi.org/10.1016/j.zemedi.2020.05.001

Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). 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

Hochberg, D. C., Greenspan, H., & Giryes, R. (2022). A self supervised StyleGAN for image annotation and classification with extremely limited labels. *IEEE Transactions on Medical Imaging*, *41*(12), 3509-3519. https://doi.org/10.1109/TMI.2022.3187170

Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). spaCy: Industrial-strength natural language processing in python.

Islam, S. M., & Mondal, H. S. (2019, July 06-08). *Image Enhancement Based Medical Image Analysis*. In: Proceedings of the 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). Kanpur, India. https://doi.org/10.1109/ICCCNT45670.2019.8944910

Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., Seekins, J., Mong, D. A., Halabi, S. S., Sandberg, J. K., Jones, R., Larson, D. B., Langlotz, C. P., Patel, B. N., Lungren, M. P., & Ng, A. Y. (2019, January 27 - February 1). *Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison.* In: Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, No. 01, pp. 590-597). Honolulu, Hawaii, USA. https://doi.org/10.1609/aaai.v33i01.3301590

Jablonski, J. A., Angadi, S. S., Sharma, S., & Brown, D. E. (2022, March 10-11). *Enabling Clinically Relevant and Interpretable Deep Learning Models for Cardiopulmonary Exercise Testing*. In: Proceedings of the 2022 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT) (pp. 50-53). Houston, TX, USA. https://doi.org/10.1109/HI-POCT54491.2022.9744068

Johnson, A. E. W., Pollard, T. J., Berkowitz, S. J., Greenbaum, N. R., Lungren, M. P., Deng, C., Mark, R. G., & Horng, S. (2019). MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific Data*, *6*(1), 317. https://doi.org/10.1038/s41597-019-0322-0

Kariuki, P. W., Gikunda, P. K., & Wandeto, J. M. (2023, April 14-16). *Deep Transfer Learning Optimization Techniques for Medical Image Classification: A Review*. In: Proceedings of the 2022 International Conference on Intelligent Computing and Machine Learning (2ICML) (pp. 7-15). Qingdao, China. https://doi.org/10.1109/2ICML58251.2022.00013

Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., & Aila, T. (2020, December 6-12). Training generative adversarial networks with limited data. In: H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.) Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS'20) (pp. 12104-12114). Vancouver BC Canada.

Ke, B., Lu, H., Huo, W., & Wang, Y. (2022, July 22-24). *Semi-supervised Medical Image Classification Combining Metric Pseudo-Label and Classification Pseudo-Label*. In: P. Lin, & Y. Yang (Eds.) Proceedings of the 2022 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI) (pp. 794-799). Shijiazhuang, China. https://doi.org/10.1109/ICCEAI55464.2022.00166

Kora Venu, S., & Ravula, S. (2020). Evaluation of Deep Convolutional Generative Adversarial Networks for Data Augmentation of Chest X-ray Images. *Future Internet*, *13*(1), 8. https://doi.org/10.3390/fi13010008

Lacan, A., Sebag, M., & Hanczar, B. (2023). GAN-based data augmentation for transcriptomics: survey and comparative assessment. *Bioinformatics*, *39*(S1), i111-i120. https://doi.org/10.1093/bioinformatics/btad239

Li, Z., Xia, P., Tao, R., Niu, H., & Li, B. (2022). A New Perspective on Stabilizing GANs Training: Direct Adversarial Training. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *7*(1), 178-189. https://doi.org/10.1109/TETCI.2022.3193373

Liu, L., Zhang, Y., & Sun, L. (2023). Medimatrix: innovative pre-training of grayscale images for rheumatoid arthritis diagnosis revolutionises medical image classification. *Health Information Science and Systems*, *11*(1), 44. https://doi.org/10.1007/s13755-023-00246-7

Neumann, M., King, D., Beltagy, I., & Ammar, W. (2019, August 1). *ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing*. In: Dina D.-F., Kevin B. C., Sophia A., & J. Tsujii (Eds.) Proceedings of the 18th BioNLP Workshop and Shared Task (pp. 319-327). Florence, Italy. https://doi.org/10.18653/v1/W19-5034

Patashnik, O., Wu, Z., Shechtman, E., Cohen-Or, D., & Lischinski, D. (2021, October 10-17). *StyleCLIP: Text-Driven Manipulation of StyleGAN Imagery*. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 2065-2074). Montreal, QC, Canada. https://doi.org/10.1109/ICCV48922.2021.00209

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021, July 18-24). *Learning transferable visual models from natural language supervision*. In: M. Meila, & T. Zhang (Eds.) Proceedings of the 38th International Conference on Machine Learning, PMLR (pp. 8748-8763). https://doi.org/10.48550/arXiv.2103.00020

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1), 60. https://doi.org/10.1186/s40537-019-0197-0

Sundaram, S., & Hulkund, N. (2021). GAN-based Data Augmentation for Chest X-ray Classification. https://doi.org/10.48550/arXiv.2107.02970

Tarawneh, A. S., Hassanat, A. B., Altarawneh, G. A., & Almuhaimeed, A. (2022). Stop oversampling for class imbalance learning: A review. *IEEE Access*, *10*, 47643-47660. https://doi.org/10.1109/ACCESS.2022.3169512

Tov, O., Alaluf, Y., Nitzan, Y., Patashnik, O., & Cohen-Or, D. (2021). Designing an encoder for StyleGAN image manipulation. *ACM Transactions on Graphics (TOG)*, *40*(4), 133. https://doi.org/10.1145/3450626.3459838

Wang, Y., Ge, X., Ma, H., Qi, S., Zhang, G., & Yao, Y. (2021). Deep Learning in Medical Ultrasound Image Analysis: A Review. *IEEE Access*, *9*, 54310-54324. https://doi.org/10.1109/ACCESS.2021.3071301

Wang, X., & Qi, G.-J. (2022). Contrastive Learning With Stronger Augmentations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(5), 5549-5560. https://doi.org/10.1109/TPAMI.2022.3203630

Yuan, Z., Yan, Y., Sonka, M., & Yang, T. (2021, October 10-17). *Large-scale Robust Deep AUC Maximization: A New Surrogate Loss and Empirical Studies on Medical Image Classification*. In: Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (pp. 3040-3049). Montreal, QC, Canada. https://doi.org/10.1109/ICCV48922.2021.00303