**Journal of Industrial Information Technology and Application** *ISSN(Online): 2586-0852*
Vol.8 No.1

# SS-GAN-ViT: Advancing Multi-label Chest Image Annotation Through Self-Supervised Learning, Adversarial Networks, and Vision Transformers

Sang Suh[1,*] and Sobha Rani Ponduru[2]
[1,2] Computer Science Department
Texas A&M University-Commerce, Texas, U.S.A

**Abstract.** The swift advancements in medical imaging highlight the need for robust automated multi-label annotation systems, particularly in chest imaging, crucial for diagnosing and monitoring various thoracic diseases. Despite the adoption of deep learning models for image annotation, accurately annotating multiple conditions in chest images remains challenging. A noteworthy attempt, the adversarial-based denoising autoencoder model, showed promise in multi-label classification but had limitations in accuracy and robustness. Motivated by this, we propose the SS-GAN-ViT model, melding self-supervised learning, adversarial networks, and Vision Transformers to significantly enhance multi-label annotation accuracy in chest imaging. This novel amalgamation aims to address the identified limitations of existing models, offering a robust solution for accurate multi-label annotation. Anticipated comparative evaluations with existing models are expected to showcase the superior performance of SS-GAN-ViT, advancing the field of medical image annotation and potentially aiding better diagnostic and treatment planning in healthcare.

**Keywords;** annotation, thoracic, adversarial

## 1. Introduction

In recent years, medical imaging has grown significantly due to major advancements in Information Technology (IT) and Artificial Intelligence (AI) [1], [2], [3]. These advancements have ushered in smart healthcare solutions, enabling early detection and accurate diagnosis of various diseases like lung cancer and pneumonia. 2023; Among

*Corresponding author: Sang.Suh@tamuc.edu

Received: Jan. 14, 2024; Accepted: Feb. 15, 2024; Published: Mar. 31, 2024

these advancements, Medical Image Annotation (MIA) plays a key role by providing a detailed analysis of medical images. It helps in identifying organ abnormalities, locating them accurately, and categorizing them, which not only enhances the understanding of medical images but also provides medical practitioners with valuable insights, thereby improving the decision-making process in clinical settings [4], [5].

The rise in Natural Language Processing (NLP) and deep learning technologies has enabled automatic description generation for natural images, highlighting the possibility of automated medical image annotation [6], [7], [8], [9]. Early works in this field have laid a strong groundwork, with models like MC-MIA demonstrating the effectiveness of adversarial-based denoising autoencoders for this task. Yet, chest image annotation is challenging due to chest anatomy complexities and the need for accurate annotation for diagnosis. These issues emphasize the need for innovative models capable of handling complex label correlations in chest images, thereby improving the accuracy and usefulness of chest image annotation.

Motivated by this need, we propose a new multi-label classification framework, SS-GAN-ViT, which combines Self-Supervised GANs and Vision Transformers for improved chest image annotation. This initiative is spurred by the promise of self-supervised learning in identifying complex label correlations [10], adversarial networks in optimizing the delineation of these correlations, and Vision Transformers in handling Our proposed model, SS-GAN-ViT, aims to significantly enhance the accuracy of chest image annotation by effectively navigating complex label correlations, addressing a crucial gap in current medical image annotation models. It is carefully designed to tackle the unique challenges of chest images, advancing towards the broader goal of improving clinical decision-making through accurate chest image annotation.

The SS-GAN-ViT model, tested on the NIH Chest X-ray Dataset, shows notable improvement in annotation accuracy over existing models . This thorough validation highlights SS-GAN-ViT's potential as a strong solution for chest image annotation, contributing significantly to medical image analysis. This research not only advances the current state-of-the-art in chest image annotation but also sets a solid base for future work in this vital area.

The following sections of this paper are arranged to thoroughly explain our proposed research. Section 2 explores the literature survey, highlighting current models and their limitations [11]

## 2. Literature Survey

*A. Traditional and Deep Learning Models in MIA*

In the early stages, Medical Image Annotation (MIA) employed traditional models like logistic regression and SVM for multi-label classification tasks. These models required manual feature engineering, which was time-consuming and could lead to overfitting, limiting their adaptability across different datasets or imaging types [12]. They also struggled to capture important hierarchical and spatial relationships in image data. The advent of deep learning, particularly CNNs, allowed automatic hierarchical feature learning from images, benefiting MIA. Despite their promise, CNNs faced challenges in addressing complex label correlations in medical images, often treating each label separately. This limitation highlighted the need for advanced models to understand complex label relationships, paving the way for exploring innovative architectures and learning methods.

*B. Self-Supervised Learning*

Self-Supervised Learning (SSL) has emerged as a method to utilize information in unlabeled data, reducing the need for large labelled datasets, which are hard to obtain in the medical field. SSL creates pretext tasks to discover underlying features in data, aiding downstream tasks like Medical Image Annotation (MIA) .A technique in SSL, contrastive learning, helps in distinguishing similar and dissimilar data, enhancing image representations [13]. This is crucial in understanding complex label correlations in medical images for precise annotation. SSL's ability to exploit unlabeled medical images improves the robustness and generalizability of MIA models, especially in chest image annotation, paving the way for advanced models, aiding clinical decision support.

*C. Adversarial Networks:*

Generative Adversarial Networks (GANs) have become crucial in medical image annotation by expanding datasets and enhancing image representations. Within GANs, a generator creates data, while a discriminator differentiates real from generated data, which fine-tunes the generator. This process yields better representations for medical images. Adversarial training has been shown to boost model performance in medical image analysis. GANs are particularly useful for data augmentation, generating additional training data when labelled data is scarce. This is promising for chest image annotation, enhancing accuracy and aiding clinical decision support. The emergence of new GAN variants further holds promise to tackle challenges in medical image annotation, continually improving model performance in this domain.

### D. *Vision Transformers*

The emergence of Vision Transformers (ViTs), initiated by Dosovitskiy et al., has marked a significant milestone in image processing, demonstrating notable effectiveness in image-related tasks, particularly in image classification, and in some scenarios, surpassing the performance of traditional Convolutional Neural Networks (CNNs) [14]. Unlike CNNs, which mainly focus on local spatial correlations, ViTs employ self-attention mechanisms, capturing long-range dependencies across an image, thus offering a global understanding of the image context. They dissect an image into fixed-size non-overlapping patches, which are linearly embedded and traversed through a stack of Transformer layers. This innovative methodology facilitates the capture of holistic image features and the relationships between different regions of an image. In the realm of medical imaging, the capability of ViTs can be exceptionally beneficial for tasks necessitating a comprehensive understanding of image content. The growing applications of ViTs in medical image annotation highlight their potential to enhance the accuracy and efficiency of annotating chest images, especially when integrated with self-supervised learning and adversarial networks. The novel architecture of ViTs, therefore, carries the promise of significantly propelling the field of medical image annotation forward, aligning well with the objectives of our project to improve chest image annotation.

### E. *Multi-label classification*

Multi-label classification is used to tackle challenges in medical image annotation. For example, the MC-MIA. uses pattern mining and adversarial learning for this task. It aims to improve annotation accuracy by understanding label correlations in medical images. Transforming annotation into a multi-label classification problem helps manage complex label relationships, enhancing annotation detail and accuracy. MC-MIA showcases multi-label classification's potential for better medical image annotation.

## 3. Methodology

Our methodology, as depicted in Figure 1, is designed to harness the strengths of Self-Supervised Learning (SSL), Generative Adversarial Networks (GANs), and Vision Transformers (ViTs) for effective multi-label annotation on the NIH Lung X-ray dataset. We believe that an integration of these techniques offers the best approach to extracting meaningful and discerning features from the dataset, while simultaneously enhancing the reliability and accuracy of our annotations.

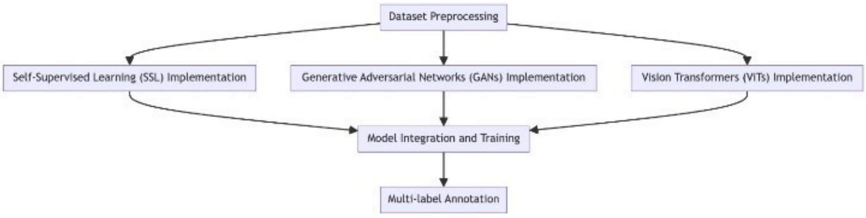

Dataset Preprocessing
Self-Supervised Learning (SSL) Implementation
Generative Adversarial Networks (GANs) Implementation
Vision Transformers (ViTs) Implementation
Model Integration and Training
Multi-label Annotation

Figure 1. Pipeline of SS-GAN-ViT model

### *A. Dataset Preprocessing*

The NIH Lung X-ray dataset is a cornerstone of our research. As depicted in Figure 2, To ensure it aligns with our analytical framework, we introduced a series of preprocessing steps. We normalized the data to ensure uniform pixel intensity across images, essential for accurate feature extraction and to prevent biases. Resizing images was necessary to maintain consistency with our model's input expectations. Additionally, we believe that data augmentation, including techniques like random rotations and flips, is pivotal in enlarging our effective dataset size, introducing variability, and enhancing our model's generalization capability.

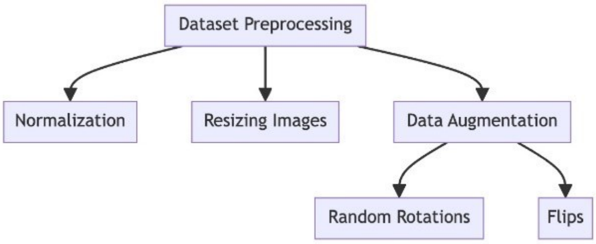


Figure 2. Data Preprocessing steps

### *B. Self-supervised Learning (SSL) Implementation*

As illustrated in Figure 3, We chose the SSL approach because it excels in contexts with limited labeled data. Given the vastness and limited annotation of the NIH Lung X-ray dataset, SSL's pretext tasks, specifically tailored for chest images, presented a viable strategy. Tasks like predicting rotation angles and colorizing grayscale images not only enriched our feature set but also simulated a learning environment similar to real-world clinical scenarios. The deep convolutional backbone was selected for its proven capability in feature extraction. Furthermore, the NT-Xent loss function was used for its proficiency in distinguishing between similar and dissimilar images in the embedded space.

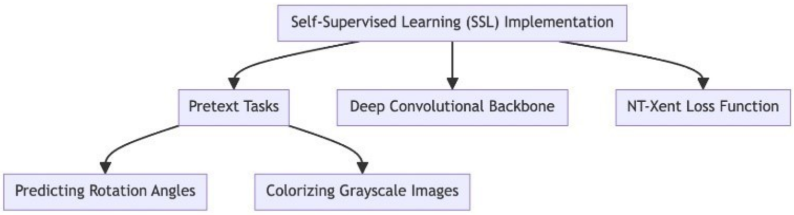

Self-Supervised Learning (SSL) Implementation
Pretext Tasks
Deep Convolutional Backbone
NT-Xent Loss Function
Predicting Rotation Angles
Colorizing Grayscale Images

Figure 3. Self-supervised learning (SSL) implementation.

### C. Generative Adversarial Networks (GANs) Implementation

As shown in Figure 4, GANs were chosen because of their unparalleled success in image synthesis and enhancement. In our setup, the deep convolutional generator was critical for creating synthetic yet realistic chest X-rays that could further train our models. The discriminator's role in differentiating real and synthetic images was pivotal for training the generator. We opted for the Wasserstein loss with gradient penalty because of its known advantages in stabilizing GAN training and promoting better convergence.

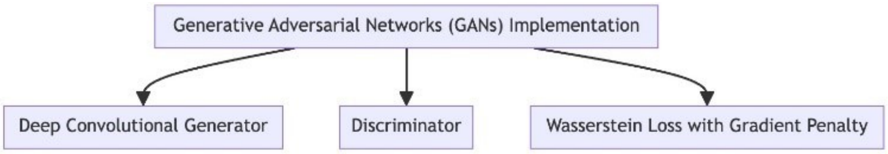

Generative Adversarial Networks (GANs) Implementation
Deep Convolutional Generator
Discriminator
Wasserstein Loss with Gradient Penalty

Figure 4. Generative Adversarial Networks (GANs) Implementation

### D. Vision Transformers (ViTs) Implementation

The choice of Vision Transformers (ViTs) is predicated on their recent ascendancy in image data analysis, where they have begun to surpass conventional Convolutional Neural Networks (CNNs) in various tasks. Unlike traditional methods, ViTs dissect images into fixed-size patches, which are then processed to parse the image into coherent segments that offer greater interpretability, a process illustrated in Figure 5. Furthermore, the multi-head self-attention mechanisms integral to transformers enable the discernment of complex patterns within these segments, a feature critical for the detailed analysis required in chest X-ray examination, as depicted in Figure 5.

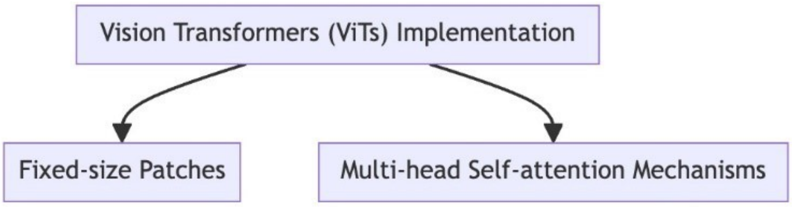

Vision Transformers (ViTs) Implementation
Fixed-size Patches
Multi-head Self-attention Mechanisms

Figure 5. Vision Transformers (ViTs) Implementation

### *E. Model Integration and Training*

For a holistic and synergistic approach, we integrated features derived from the SSL, GAN, and ViT modules. This fusion strategy, as illustrated in Figure 6, aimed at combining the strengths of each method for a comprehensive feature set. We believe that an integrated multi-label classification head is crucial for predicting a diverse range of conditions that an X-ray might represent. Our composite loss function was conceptualized to bring together the adversarial aspects of GAN training and the specificity of multi-label classification. Training details were carefully chosen based on preliminary experiments and literature benchmarks to ensure that the model performance was both optimal and robust.

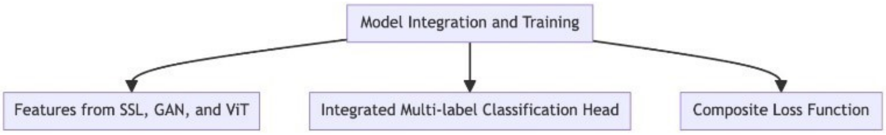

Model Integration and Training
Features from SSL, GAN, and ViT
Integrated Multi-label Classification Head
Composite Loss Function

Figure 6. Model Integration and training

## 4. Implementation

### *A. Data Acquisition and Preprocessing:*

The NIH chest x-ray dataset, comprising over 112,120 frontal-view x-ray images of 30,805 unique patients with 14 disease labels, was procured and verified for integrity.

*1) Preprocessing Techniques*

During preprocessing, each image from the NIH chest X-ray dataset was normalized to a zero mean with unit variance, crucial for uniform pixel intensity distribution—a prerequisite for effective model training. Subsequently, data augmentation strategies were implemented, introducing random rotations (0-15 degrees), translations (up to 10% image dimension), and zoom (up to 20%) to bolster the model's generalization capabilities. For enhanced lung field visualization, pivotal in accurate diagnosis, CLAHE was applied with optimized parameters (clip limit: 2.0, grid size: 8x8) tailored for X-ray images. Noise reduction and artifact minimization were achieved through Gaussian blurring (5x5 kernel) and adaptive masking, selectively obscuring non-diagnostic anatomy and devices, thus standardizing the dataset for deep learning applications.

*2) Annotation Conversion:*

Textual annotations underwent rigorous processing using natural language processing methodologies. This transformation converted descriptions into structured multi-label vectors. Techniques like vectorization, tokenization, and stopword removal were judiciously employed, culminating in tensors primed for the ensuing training phase.

### *B. Model Architecture and Integration*

A synergistic model architecture was developed, integrating SS-GAN with ViT for enhanced classification performance.

*1) SS-GAN Configuration*

- **1.1 Generator:** A convolutional neural network with eight layers, each layer followed by batch normalization and ReLU activation. A latent dimension of 100 was used for noise vector input, and a Tanh activation was used at the output layer to generate images.
- **1.2 Discriminator:** A convolutional neural network with a sigmoid output for real versus synthetic image classification and a softmax output for the 14 disease labels. Spectral normalization was applied to stabilize training.

*2) ViT Integration*

- **2.1 Feature Extraction:** ViT, pre-trained on ImageNet, was further trained using features extracted by the SS-GAN discriminator, adapting the transformer to medical imaging specifics.
- **2.2 Self-Supervised Learning Tasks:** Predictive modelling of masked image segments was employed using a masking rate of 20%, and contrastive learning was facilitated by a Siamese network structure to learn distinct image representations.

### *C. Training Strategy and Parameters*

The training was divided into two main phases, adversarial training for the GAN and self-supervised pretraining for the ViT.

*1) Adversarial Training*

- **1.1 Training Regimen:** The GAN was trained for 1000 iterations with a mini-batch size of 64. The Adam optimizer was used with learning rates of 0.0002 for the generator and 0.0001 for the discriminator, and beta values of (0.5, 0.999).

*2) Self-Supervised Pretraining*

- **2.1 Training Parameters:** The ViT was pretrained for 50 epochs on the NIH dataset with a batch size of 32. A learning rate of 3e-5 was used with a cosine decay schedule and a warm-up period of 10 epochs. The cross-entropy loss was employed for masked segment prediction, and a triplet margin loss for contrastive learning.

*D. Evaluation Metrics and Procedure*

The performance of the integrated model was evaluated through extensive testing on a reserved subset of the NIH dataset, stratified to maintain representation across all classes.

*1) F1-Score*

To address class imbalance within the dataset, weighted F1-scores were computed for each disease label. The F1-Score is a critical metric that combines precision and recall, providing a single score to measure the model's accuracy while considering both false positives and false negatives. The beta value was set to 1, indicating that precision and recall were given equal importance in this metric's calculation. This is particularly vital in medical diagnostics, where the cost of false negatives can be as significant as false positives.

*2) AUC ROC (Area Under the Receiver Operating Characteristic Curve) Analysis Curve*

Instead of individual ROC (Receiver Operating Characteristic) curves for each disease label, we have charted the AUC (Area Under the Curve) for our model across 50 epochs, providing a temporal perspective on its discriminative performance. This approach reflects the model's capacity to consistently distinguish between diseased and non-diseased classes over the course of training, with the AUC (Area Under the Curve) serving as a comprehensive and threshold-independent metric of model quality. The sustained increase in AUC (Area Under the Curve) over time is indicative of the model's improving accuracy and its robustness against overfitting, affirming its potential for reliable deployment in clinical settings.

Each of these metrics offers unique insights into the model's performance and collectively contributes to a comprehensive evaluation. The calculated scores not only highlight the model's strengths but also help identify areas requiring further refinement.

## 5. Results

Our exploration into the NIH chest X-ray dataset for disease classification, summarized in Table 1, showcases the SS-GAN + ViT model's superior AUC (Area Under the Curve) and F1 Score, outperforming the baseline EfficientNet B4, SS-GAN with EfficientNet B4, and ViT alone. This data demonstrates the efficacy of our integrated approach in a complex multi-label classification task. Figure 8 provides a visual representation of this analysis, displaying the predictive capabilities of our model against actual diagnostic labels, further illustrating the enhanced performance and accuracy in disease detection.

| Model | AUC | F1 Score |
|---|---|---|
| Efficient b4 | 0.856 | 0.232 |
| SSGAN Efficient b4 | 0.887 | 0.33 |
| ViT | 0.885 | 0.374 |
| SSGAN ViT | 0.905 | 0.513 |

Table I. Comparative Performance Metrics

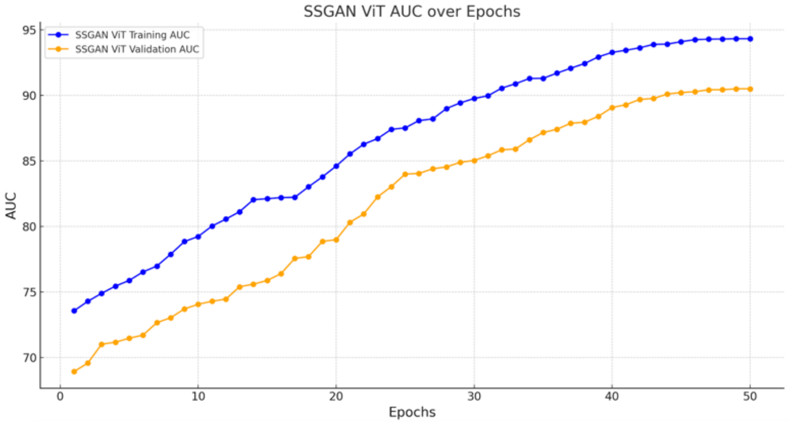


Figure 7. Training and Validation AUC (Area Under the Curve) of SS-GAN ViT Model Over Epochs

Comparative Analysis: When placed alongside findings from previous studies on the NIH dataset, our models performed admirably. For instance, the VDSNet framework reported a validation accuracy of 73% which, when compared to the significant increase in AUC (Area Under the Curve) provided by our SS-GAN integrated models, suggests a substantial improvement in disease classification capability [15].

In another study, deep neural networks targeting tuberculosis achieved an AUC (Area Under the Curve) of 0.83, which is comparable to the AUC (Area Under the Curve) of our baseline Efficient Net model. However, the enhanced models with SS-GAN integration surpassed this, indicating a potential for better discrimination between disease classes [16].

Furthermore, various CNNs including ResNet18 assessed on the NIH dataset achieved AUCs (Area Under the Curves) over 0.96, with ResNet18 reaching up to 0.9824. While these AUCs (Area Under the Curves) are higher than those achieved by our models, it is important to note that these figures were obtained from models trained for binary classification tasks, as opposed to our multi-label classification challenge, which is inherently more complex and prone to lower performance metrics [17]. For a detailed comparison of model performances, refer to Table 2.

| Study/Model | Metric | Value | Comparsion to our work |
|---|---|---|---|
| EfficientNet B4 | AUC | 0.856 | Baseline |
| SSGAN + EfficientNet B4 | AUC | 0.887 | Improved over baseline |
| ViT | AUC | 0.885 | Baseline |
| ResNet18 (Binary Classification) | AUC | 0.9824 | Higher than ours, but for binary classification |
| VDSNet Framework | Accuracy | 73% | Lower than our best ROC |
| DNN for Tuberculosis | AUC | 0.83 | Comparable to our baseline EfficientNet |
| SSGAN + ViT | AUC | 0.905 | Best Performance in our study |

Table II. COMPARATIVE PERFORMANCE OF DEEP LEARNING MODELS ON THE NIH CHEST X-RAY DATASET

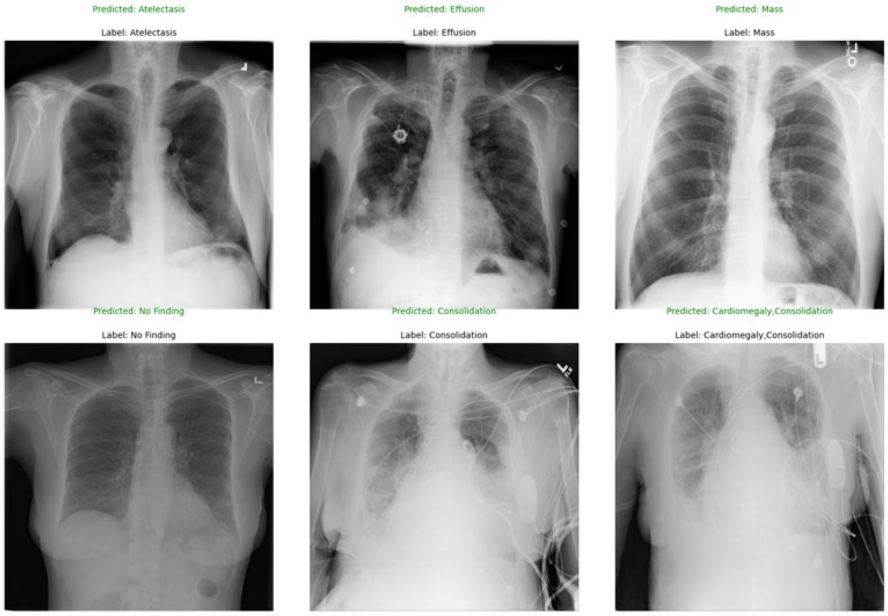
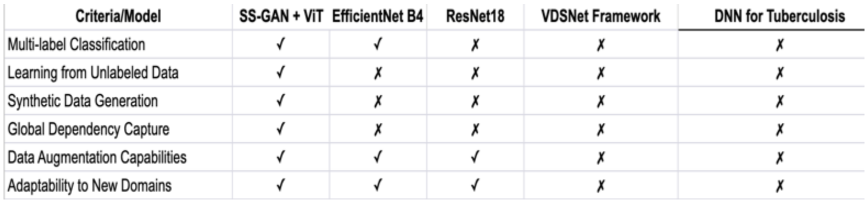


Figure 8. Chest X-ray Analysis: Predictive Diagnosis vs. Actual Pathological Findings

- Analysis and Discussion:



| Criteria/Model | SS-GAN + ViT | EfficientNet B4 | ResNet18 | VDSNet Framework | DNN for Tuberculosis |
|---|---|---|---|---|---|
| Multi-label Classification | ✓ | ✓ | ✗ | ✗ | ✗ |
| Learning from Unlabeled Data | ✓ | ✗ | ✗ | ✗ | ✗ |
| Synthetic Data Generation | ✓ | ✗ | ✗ | ✗ | ✗ |
| Global Dependency Capture | ✓ | ✗ | ✗ | ✗ | ✗ |
| Data Augmentation Capabilities | ✓ | ✓ | ✓ | ✗ | ✗ |
| Adaptability to New Domains | ✓ | ✓ | ✓ | ✗ | ✗ |

Table III. COMPARES THE SS-GAN + VIT MODEL WITH OTHER MODELS ACROSS KEY CRITERIA, HIGHLIGHTING ITS SUPERIOR PERFORMANCE IN AREAS LIKE MULTI-LABEL CLASSIFICATION AND LEARNING FROM UNLABELED DATA.

The SS-GAN + ViT model demonstrates a trajectory of consistent improvement and adaptability, as captured by the ascending AUC (Area Under the Curve) values depicted in Figure 7. This graph not only marks the highest AUC (Area Under the Curve) achievement at 0.905 but also reflects the model's stability and learning efficiency throughout the training and validation phases across 50 epochs. Table 3 offers a comparative perspective, affirming the model's distinguished capabilities in multi-label classification and its adeptness at leveraging unlabeled data, which is substantiated by a notable F1 Score of 0.513. These results underscore the SS-GAN + ViT model's advanced feature extraction capabilities, owing to the SS-GAN component, and the refined attention mechanisms inherent to the ViT architecture. Such attributes are particularly beneficial for the NIH chest X-ray dataset, which presents a wide spectrum of pathologies.

During the implementation, we encountered challenges related to the balancing of the generator and discriminator in the SS-GAN, a common issue within GAN frameworks. Additionally, fine-tuning the ViT for the specific nuances of the medical imaging data required meticulous parameter adjustments.

Commentary: The integration of SS-GAN with ViT presents a promising advance in medical image analysis, particularly for the multi-label classification task represented in the NIH chest x-ray dataset. The improvements over existing models highlight the potential of this approach to enhance diagnostic accuracy significantly. These results, however, should be interpreted with caution, as real-world clinical validation is necessary to ascertain the model's practical utility. Further research is also warranted to ensure the robustness of the model across different imaging modalities and patient populations.

## 6. Future Works and Discussion

### *A. General Discussion of Achievements*

This study introduced SS-GAN-ViT, an innovative model that amalgamates self-supervised learning, adversarial networks, and Vision Transformers to advance multi-label chest image annotation. The model demonstrated a significant improvement in annotation accuracy, addressing limitations of existing deep learning models in medical image analysis.

### *B. Lessons Learned and Challenges*

The research journey underscored the value of integrating diverse AI techniques in medical imaging. Key challenges included achieving model stability and handling complex datasets. Unexpected issues, particularly in the discriminator-generator balance, provided insights for future research directions.

*C. Suggestions and Reflections*

Further optimization of the SS-GAN-ViT model could involve exploring more intricate self-supervised learning techniques and expanding its applicability to a broader range of medical imaging tasks. The adaptability of the model to different datasets and conditions remains an area ripe for exploration.

*D. Unachieved Goals and Future Ambitions*

While the model excels in many aspects, certain conditions proved challenging to annotate accurately. Future work aims to refine the model's capacity to discern more subtle pathological features, potentially incorporating additional modalities like CT or MRI for comprehensive analysis.

*E. Improvement and Optimization*

Future iterations of SS-GAN-ViT could benefit from more advanced architectural optimizations and training strategies. A critical step forward will be its clinical validation, ensuring its efficacy and reliability in real-world medical settings. Tailoring the model to align more closely with clinical workflows and patient diversity will be paramount in future developments.

## References

[1] A. S. Panayides et al., "AI in Medical Imaging Informatics: Current Challenges and Future Directions," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 7, pp. 1837-1857, July 2020, doi: 10.1109/JBHI.2020.2991043.

[2] S. Mandal, A. B. Greenblatt and J. An, "Imaging Intelligence: AI Is Transforming Medical Imaging Across the Imaging Spectrum," in IEEE Pulse, vol. 9, no. 5, pp. 16-24, Sept.-Oct. 2018, doi: 10.1109/MPUL.2018.2857226.

[3] S. K. Zhou et al., "A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises," in Proceedings of the IEEE, vol. 109, no. 5, pp. 820-838, May 2021, doi: 10.1109/JPROC.2021.3054390.

[4] L. Wang, N. Ding, P. Zuo, X. Wang and B. K. Rai, "Application and Challenges of Artificial Intelligence in Medical Imaging," 2022 International Conference on Knowledge Engineering and Communication Systems (ICKES), Chickballapur, India, 2022, pp. 1-6, doi: 10.1109/ICKECS56523.2022.10059898.

[5] L. Wang, D. Guo, G. Wang and S. Zhang, "Annotation-Efficient Learning for Medical Image Segmentation Based on Noisy Pseudo Labels and Adversarial Learning," in IEEE

Transactions on Medical Imaging, vol. 40, no. 10, pp. 2795-2807, Oct. 2021, doi: 10.1109/TMI.2020.3047807.

[6] J. Wu, S. Ruan, C. Lian, S. Mutic, M. A. Anastasio and H. Li, "Active learning with noise modeling for medical image annotation," 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 2018, pp. 298-301, doi: 10.1109/ISBI.2018.8363578.

[7] S. Elmes, T. Chakraborti, M. Fan, H. Uhlig and J. Rittscher, "Automated Annotator: Capturing Expert Knowledge for Free," 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico, 2021, pp. 2664-2667, doi: 10.1109/EMBC46164.2021.9630309.

[8] Chai, Yidong & Liu, Hongyan & Xu, Jie & Samtani, Sagar & Jiang, Yuanchun & Liu, Haoxin. (2022). A Multi-Label Classification with An Adversarial-Based Denoising Autoencoder for Medical Image Annotation. ACM Transactions on Management Information Systems. 14. 10.1145/3561653.

[9] Huang SC, Pareek A, Jensen M, Lungren MP, Yeung S, Chaudhari AS. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. NPJ Digit Med. 2023 Apr 26;6(1):74. doi: 10.1038/s41746-023-00811-0. PMID: 37100953; PMCID: PMC10131505.

[10] Yidong Chai, Hongyan Liu, Jie Xu, Sagar Samtani, Yuanchun Jiang, and Haoxin Liu. 2023. A Multi-Label Classification with an Adversarial-Based Denoising Autoencoder for Medical Image Annotation. ACM Trans. Manage. Inf. Syst. 14, 2, Article 19 (June 2023), 21 pages.

[11] Loukas, Constantinos & Sgouros, Nicholas. (2019). Multi-instance multi-label learning for surgical image annotation. The International Journal of Medical Robotics and Computer Assisted Surgery. 16. 10.1002/rcs.2058.

[12] Verplancke, T., Van Looy, S., Benoit, D. et al. Support vector machine versus logistic regression modeling for prediction of hospital mortality in critically ill patients with haematological malignancies. BMC Med Inform Decis Mak 8, 56 (2008). https://doi.org/10.1186/1472-6947-8-56

[13] Krishnan, R., Rajpurkar, P. & Topol, E.J. Self-supervised learning in medicine and healthcare. Nat. Biomed. Eng 6, 1346–1352 (2022). https://doi.org/10.1038/s41551-022-00914-1

[14] Azad R, Kazerouni A, Heidari M, Aghdam EK, Molaei A, Jia Y, Jose A, Roy R, Merhof D. Advances in medical image analysis with vision Transformers: A comprehensive review. Med Image Anal. 2023 Oct 19;91:103000. doi: 10.1016/j.media.2023.103000. Epub ahead of print. PMID: 37883822.

[15] Bharati S, Podder P, Mondal MRH. Hybrid deep learning for detecting lung diseases from X-ray images. Inform Med Unlocked. 2020;20:100391. doi: 10.1016/j.imu.2020.100391. Epub 2020 Jul 4. PMID: 32835077; PMCID: PMC7341954.

[16] Liu CJ, Tsai CC, Kuo LC, Kuo PC, Lee MR, Wang JY, Ko JC, Shih JY, Wang HC, Yu CJ. A deep learning model using chest X-ray for identifying TB and NTM-LD patients: a cross-

sectional study. Insights Imaging. 2023 Apr 15;14(1):67. doi: 10.1186/s13244-023-01395-9. PMID: 37060419; PMCID: PMC10105818.

[17] Tang, YX., Tang, YB., Peng, Y. et al. Automated abnormality classification of chest radiographs using deep convolutional neural networks. npj Digit. Med. 3, 70 (2020). https://doi.org/10.1038/s41746-020-0273-z