Enhancing Lungs Tumor Detection with Generative Models

Pietro Picchione[1], Robin Ghyselinck[1], Bruno Dumas[1], and Benoît Frénay[1]

University of Namur, Namur 5000, Belgium
https://www.unamur.be

Abstract. Early diagnosis of lung cancer largely relies on the interpretation of bronchoscopic images, a complex task that strongly depends on clinical expertise. Deep learning has shown potential in computer-aided detection of lung cancer, but its performance remains limited by the scarcity of annotated data in endoscopic imaging. To address this constraint, we explore the use of synthetic data generated with SinGAN-Seg, a model capable of producing realistic images from a single example.
In this study, 257 bronchoscopic images from 64 patients were used to train 257 individual SinGAN-Seg models. The generated images were filtered using quantitative similarity metrics and integrated in increasing proportions (5% to 100%) into the training sets of three convolutional classifiers (ResNet-18, ResNet-50, EfficientNet-B7). Performance was evaluated through five-fold cross-validation on an independent validation set.
The results show that a moderate addition of synthetic data (up to 33% for ResNet-18 and 66% for EfficientNet-B7) led to small yet consistent improvements in accuracy and sensitivity, with gains ranging from +1% to +3% depending on the architecture. However, an excess of artificial data degraded precision and overall performance. These findings confirm the potential of generation-based augmentation to enhance generalization in data-scarce medical contexts, while emphasizing the importance of balanced integration and rigorous evaluation.

Keywords: Deep learning · Generative models · SinGAN-Seg · Bronchoscopy · Data augmentation · Lungs cancer

1 Introduction

Lung cancer remains the major cause of cancer-related mortality worldwide [3, 20]. It accounts for approximately 2.5 million new cases and 1.8 million deaths each year, representing about 12% of all cancers and nearly 19% of all cancer-related deaths [3]. Despite advances in screening and targeted therapies, survival outcomes remain unsatisfactory, largely due to late diagnosis and the complexity of the disease [20]. When detected at an early stage, however, the five-year survival rate can exceed 60%, compared to less than 15% for advanced stages [25].

Bronchoscopy is a medical procedure where physicians examine the inside of the airways with a flexible or rigid endoscope. It allows the clinician to directly visualize the lungs and identify abnormalities such as inflammation, obstructions, or tumoral lesions, and perform biopsies [5, 15]. In the context of lung cancer diagnosis, bronchoscopy is performed after suspicious nodules have been detected on a chest CT scan, in order to confirm or rule out their pathological nature.

The majority of lung cancers develop in the peripheral regions of the lungs, which are difficult to access with a conventional endoscope. Navigating to these areas requires crossing numerous increasingly narrow bifurcations of the bronchial tree. This anatomical complexity, combined with the physical size of the instruments, makes the procedure prone to navigation errors, which can result in blind biopsies with limited success rates [5]. Although ultra-thin bronchoscopy improves access to small bronchi, its use remains technically complex and not widely standardized.

Despite the usefulness of navigational bronchoscopies in diagnosing lung cancer, their interpretation remains a complex task that is highly dependent on the clinician's expertise. This limits the effectiveness of the procedure.

In this context, deep learning has emerged as a promising tool for computer-aided detection systems. Its performance, however, relies on the availability of large annotated datasets, which is a challenge in endoscopic image analysis, where data collection and annotation remain costly and limited. Moreover, the low inter-patient variability and visual redundancy of video sequences further exacerbate this issue, reducing the generalization capacity of models [18, 23].

To overcome these limitations, one option is to use synthetic image generation in order to increase the size and diversity of training datasets. To do so, there exist many deep learning models, such as *Generative Adversarial Networks* (GANs), *Variational Autoencoders* (VAEs), or diffusion-based approaches [6, 13, 9]. They typically require important computational resources and large volumes of training data. The SinGAN-Seg [22] architecture does not exhibit those constraints and has the ability to generate diverse images from single examples.

The objective of this work is to assess the contribution of synthetic bronchoscopic data generated by SinGAN-Seg in a supervised classification setting. We investigate the effect of different proportions of artificial images on the performance of several convolutional network architectures, in order to determine whether this strategy can help improve the robustness and sensitivity of models in the context of early detection of lung tumors.

Our contributions are as follows:

- Introduction of a new medical use case for the SinGAN-Seg architecture.
- A study of the evolution of carefully selected performance metrics with increasing proportion of synthetic elements in a dataset.

2 Related Works

A first study by Amante et al. [2] used a ResNet-50 [14] trained on frames extracted from 41 patients and tested on frames from 20 patients based on a total

of over 60,000 images. In their approach, they first made per-frame predictions (cancer probability) with a ResNet-50, then applied post-processing techniques (ARIMA) to smooth the probabilities. Finally, they choose window sizes (between 30 to 60 frames) to make a unique prediction of the tumoral aspect per patient. If there was one window size where all the smoothed probabilities were over 50%, then the patient was marked as having cancer. Their model achieved 65.6% accuracy and 61.5% F1-score on the task of predicting lung cancer per patient, performing better than junior physicians but still below expert-level accuracy. These architectures were chosen as representative convolutional backbones of increasing depth and complexity, allowing a fair comparison of model capacity versus dataset scale (see section 4.1 for metric definitions).

To mitigate the lack of annotated data, generative models are widely investigated for data augmentation in medical imaging tasks [4]. VAEs and their variants have been used for representation learning and synthesis under constrained data regimes, despite sometimes yielding over-smooth samples [11, 16]. GANs have been used for realistic texture synthesis and domain transfer; in endoscopy, reported applications include capsule endoscopy augmentation (WCE-DCGAN) and navigation or feature enhancement (EndoL2H), as well as texture-focused adversarial schemes [24, 1, 7]. More recently, diffusion-based generative models have demonstrated high fidelity and controllability in image generation, albeit at higher computational costs [9]. Several studies caution that synthetic data must be validated carefully: while it can improve sensitivity and address class imbalance, excessive reliance on artificial samples may degrade precision and clinical fidelity [10, 12].

Beyond augmentation, recent research has also explored synthetic data for pre-training models before fine-tuning on limited real datasets. For instance, the Task2Sim framework [17] investigated the transferability of models pre-trained on fully synthetic data generated via controllable simulators, demonstrating that such pre-training could yield competitive downstream performance compared to ImageNet-based initialization. These approaches highlight that synthetic pre-training could serve as a viable alternative when real large-scale datasets are unavailable.

SinGAN-Seg is an extension of the SinGAN model, initially proposed by Shaham et al. for generating realistic images from a single example [19]. While SinGAN relies on a hierarchical architecture of Generative Adversarial Networks trained at multiple resolution scales, SinGAN-Seg adapts this principle to the medical domain by simultaneously incorporating the original image and its segmentation mask [22]. This approach makes it possible to generate not only diverse synthetic images but also their corresponding masks, which is an important advantage for supervised tasks such as classification or segmentation.

The training process follows a multi-scale strategy: at each level, the generator learns to reproduce the structure and local textures of the input image, while the discriminator evaluates fidelity against real samples. Adding the mask as an additional channel allows the model to preserve the anatomical consistency

between the image and the annotated regions, unlike the original SinGAN which only accounted for visual appearance.

The main strength of SinGAN-Seg lies in its ability to operate with a limited number of data samples. Whereas other generative models (GANs, VAEs, Diffusion Models) require thousands of annotated examples, SinGAN-Seg can be trained on a single image–mask pair to produce a large number of synthetic variations. This property makes it well suited for medical contexts, where data collection and annotation are particularly costly and restricted.

There remains a gap in understanding how far single-image generative augmentation can push classification robustness in bronchoscopy, and how much synthetic proportion is beneficial before precision degrades. This paper addresses this gap by systematically varying the synthetic ratio across three classifier families and quantifying the trade-offs on an independent validation set.

3 Methodology

3.1 Dataset

The dataset used in this study originates from anonymized bronchoscopic video recordings of 64 patients, yielding 62,072 frames (30 fps, 400×400 pixels). Each frame was annotated by an expert pulmonologist to indicate the presence or absence of a tumor, following the descriptors of tumoral and benign endoscopic patterns defined in previous work.

Patients were divided into a training set (41 patients, including 25 cancer cases) and a validation set (20 patients, including 14 cancer cases). Overall, 37 out of 64 patients ($\approx 61\%$) had a final diagnosis of lung cancer, while 24 ($\approx 39\%$) presented benign lesions such as inflammatory nodules, aspergillomas, hamartomas, or cryptogenic organizing pneumonia. This distribution reflects a realistic clinical imbalance, with non-tumoral cases being more frequent at the frame level.

In the validation set, approximately 75% of frames corresponded to non-cancerous regions and 25% to tumoral ones, mirroring the real-world prevalence of visible malignant tissue during bronchoscopy. To mitigate this imbalance, both a class-weighted loss and oversampling were applied during training. Oversampling ensured that each mini-batch contained an equal proportion of positive (tumoral) and negative (non-tumoral) samples. Evaluation metrics (defined in Section 4.1) were prioritized to prevent bias toward the majority (non-tumoral) class. The validation set contained both malignant and benign lesions with heterogeneous endoscopic appearances, ensuring that generalization capacity was properly assessed.

3.2 Synthetic Data Generation

A total of 257,000 synthetic images were generated, representing a thousand-fold increase compared to the 257 manually selected originals, cold references

images here after. These reference images were chosen from the 64 available patients based on visual quality, sharpness, and the presence of relevant bronchial structures. Each was used to train an independent SinGAN-Seg model under controlled hyperparameters (five scales, scale factor 0.75, noise amplitude 0.1, 2000 iterations per scale, learning rate 5×10^{-4}), with an average training time of 50 minutes per image on an NVIDIA A100 GPU (40 GB RAM). To select the most relevant synthetic images, the following process was used. All generated samples were post-processed and filtered using quantitative similarity metrics (full definition given in appendix) —Structural Similarity Index Measure (SSIM), Peak Signal-to-Noise Ratio (PSNR), and Mean Squared Error (MSE)—as well as distance-based thresholds (cosine similarity 0.80–0.95, Euclidean distance 49k–70k). Each synthetic image was compared to its corresponding reference frame to discard both almost equivalent (SSIM > 0.95) and unrealistic outliers (SSIM < 0.80 or PSNR < 25). This filtering step automatically removed visually redundant or implausible samples, while retaining diverse yet clinically coherent variations. The selected synthetic images were integrated exclusively into the training set, with consistent labels, while the validation set remained untouched to prevent data leakage.

These enhanced datasets, combining real and filtered synthetic images in varying proportions (0%, 5%, 33%, 50%, 66%, and 100%), were subsequently used to train convolutional neural networks for binary lesion classification.

3.3 Generated Images

Figure 1 illustrated examples of synthetic bronchoscopic frames generated by SinGAN-Seg. For each case, the original image was compared with its synthetic variants, and the effect of style transfer post-processing was highlighted. Differences introduced by the generator were visualized in red in the last column.

Qualitatively, the model preserved the global bronchial anatomy (bifurcations, carina, mucosal texture) while introducing controlled variability in illumination, local texture, and edge definition. This variability mimicked the natural heterogeneity observed in clinical video sequences and helped to reduce redundancy within the training set. The style transfer step further enhanced realism, aligning synthetic samples more closely with the appearance of real bronchoscopic videos.

Such synthetic diversity did not create new pathological patterns from scratch, but perturbed the visual domain in a clinically coherent way. This made the generated images particularly suitable for data augmentation, as they expanded the variability available to the classifier without drifting outside the medical distribution.

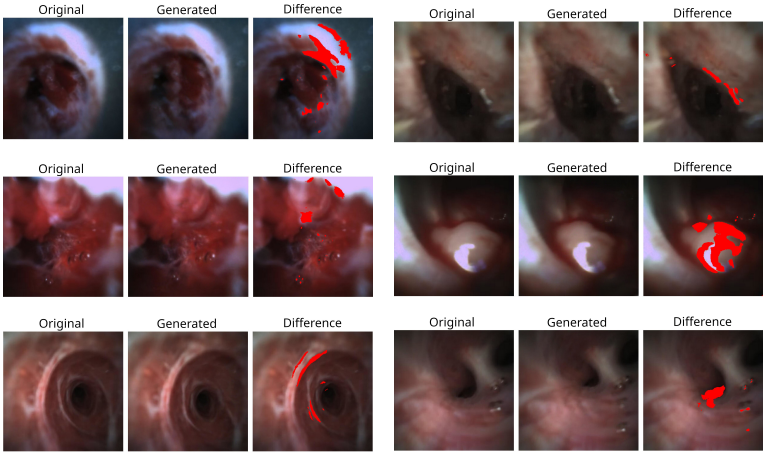

Fig. 1. Examples of synthetic bronchoscopic images generated by SinGAN-Seg. Each row shows the original frame, its synthetic counterparts, and the style transfer output. Differences introduced by the generator are visualized in red in the last column.

3.4 Architectures and Parameters

Three convolutional network architectures were selected for binary classification:

- **ResNet-18** and **ResNet-50**, introduced by He et al. [8], represented shallow and deep residual networks respectively.
- **EfficientNet-B7**, proposed by Tan and Le [21], leveraged compound scaling for optimal accuracy–efficiency trade-offs.

Table 1. Comparison of CNN architectures used in this study.

Model	Layers	Params	Key Features	Advantages
ResNet-18	18	11.7M	Simple residual blocks	Lightweight, fast training
ResNet-50	50	25.6M	Bottleneck residual blocks	Deeper, higher expressiveness
EfficientNet-B7[1]	∼66	66.3M	Compound scaling, MBConv blocks	High accuracy, efficient design

Those models were chosen because they had demonstrated good performances in medical imaging and analysis (see Section 2). It was expected that they would be almost equivalent in terms of performance despite their different

[1] Approximate layer count; EfficientNet-B7 uses composite MBConv blocks with multiple internal operations.

structural approaches. Mathematical definitions can be found in the following references [8, 21].

Each network was trained both on the original dataset and on the augmented datasets enriched with synthetic data. Training hyperparameters were adjusted consistently to ensure fair comparison across configurations: learning rate 5×10^{-5}, batch size 32, binary cross-entropy loss, and 15 epochs. The number of iterations and batches was kept identical across experiments to maintain training parity. Model selection was based on the highest validation AUC-ROC score (defined in Section 4.1).

Initially, the loss was a binary cross-entropy weighted by class imbalance. Moreover, oversampling of the minority class (cancer) was applied to the training set, such that 50% of the data seen by the models came from the positive class and 50% from the negative class. Therefore, the binary cross-entropy loss did not play a significant role in the training process. For the validation process, no binary cross-entropy function was used.

4 Experiments and Results

4.1 Performance Evaluation

Model performance was assessed on an independent validation set, completely separated from the data used for training and generation. Several standard metrics were employed to capture complementary aspects of classification performance:

- **Accuracy**, measuring the overall proportion of correct predictions;
- **AUC-ROC**, evaluating the discrimination capability between classes;
- **F1-score**, balancing precision and recall;
- **Precision** and **Recall**, examining false positives and false negatives separately.

For binary classification, let TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively. The metrics are defined as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}},$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \qquad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

The **AUC-ROC** (Area Under the Receiver Operating Characteristic curve) quantifies the trade-off between true positive rate (TPR) and false positive rate (FPR) across different thresholds, providing a threshold-independent measure of discriminative performance.

Each model was trained for 15 epochs with a batch size of 32 and a learning rate of 5×10^{-5}. The best checkpoint was selected based on the highest validation AUC-ROC, ensuring a consistent criterion across all configurations. No test-time augmentation or external pretraining was applied.

4.2 Data Integration Strategy

To evaluate the contribution of the generated images, three dataset configurations were considered: using exclusively real data (0% synthetic), a mixed dataset with 33% synthetic samples, and a fully synthetic configuration (100% synthetic). For each ratio, the required number of generated samples was randomly drawn from the pool of filtered images, ensuring unbiased integration while keeping the validation set identical across all experiments. This setup allowed for a fair comparison of model robustness and sensitivity under controlled synthetic data proportions.

4.3 Results

Table 2 reports the mean Accuracy, AUC-ROC, F1-score, Precision, and Recall for each architecture on the validation set. The values correspond to the best-performing checkpoint for each configuration. 100% means full synthetic dataset (257000) plus the original dataset (57290).

Table 2. Model performance according to the proportion of synthetic data (validation set).

Model	Synthetic	Accuracy	AUC-ROC	F1-score	Precision	Recall
ResNet-18	0%	0.6908	0.7050	0.4785	0.4051	0.5844
	33%	0.7044	0.6989	0.4714	0.4189	0.5389
	100%	0.6531	0.6573	0.4242	0.5263	0.3552
ResNet-50	0%	0.7113	0.7550	0.5262	0.4374	0.6605
	33%	0.6913	0.7448	0.5126	0.4174	0.6640
	100%	0.6363	0.7150	0.4700	0.3636	0.6644
EfficientNet-B7	0%	0.6400	0.6877	0.4613	0.3623	0.6350
	50%	0.6836	0.7343	0.4968	0.4074	0.6364
	100%	0.6632	0.6923	0.4512	0.3732	0.5704

ResNet-18. An improvement in overall accuracy was observed when adding up to 33% synthetic data, though AUC-ROC and recall slightly decreased. This can be explained by the fact that the training set was balanced while the validation set was unbalanced (75% negative and 25% positive). For the same reason precision did follow the same behaviour as accuracy (slight increase) At 100% synthetic data, performance dropped across all metrics.

ResNet-50. This deeper model achieved the best overall performance on real data alone. Adding synthetic data did not yield measurable gains and instead slightly reduced accuracy and AUC-ROC, while recall remained stable. This indicates that the base dataset already provided sufficient diversity for this architecture.

EfficientNet-B7. For this more complex model, moderate synthetic augmentation (50%) improved accuracy and AUC-ROC compared to the baseline, before decreasing at 100%. This suggests that higher-capacity models may benefit from larger synthetic contributions, though excessive artificial data still impairs generalization.

The evaluation metrics confirm that the most significant improvement concerns sensitivity (accuracy), i.e., the ability to correctly identify positive cases. This suggests that the inclusion of synthetic data generated by SinGAN-Seg helps to better capture intra-patient variability while reducing the risk of under-detection of lesions. However, this is not true for all the models suggesting that the inclusion of synthetic data should be model dependent.
Conversely, all the metrics tends to deteriorate when the proportion of artificial data becomes too high, reflecting an increase in false positives. Moreover, the positive effect is mainly observed for moderate proportions, while an over-representation of synthetic data compromises the balance between sensitivity and precision.

5 Discussion

The integration of images generated by SinGAN-Seg showed a differentiated effect depending on both the proportion of synthetic data and the classification architecture considered. In general, a moderate addition of synthetic data (around one-third of the total corpus) led to small but measurable improvements in robustness, particularly for lighter architectures such as ResNet-18. For deeper or more complex models such as EfficientNet-B7, the benefits were consistent with overall good performances. On the other hand, adding synthetic data with ResNet-50 did not show any performance improvement, regardless of the metric considered.

Although the original SinGAN-Seg architecture is designed to jointly generate images and segmentation masks, in this work it was used in RGB-only mode to ensure compatibility with the original dataset, which did not include mask annotations. This configuration maintained training stability while leveraging the model's ability to capture structural and textural variability from single bronchoscopic examples. Consequently, the generated images preserved the overall bronchial geometry and lighting patterns without explicit anatomical constraints, resulting in visually coherent yet mask-free samples suitable for classification tasks.

These findings confirm that generation-based augmentation can partially compensate for the limited diversity of real datasets [22], but its effect remains

bounded. While synthetic data occasionally improved sensitivity, precision often decreased when the synthetic proportion was too high, revealing a trade-off between false-positive control and lesion detectability. This aligns with prior studies showing that excessive artificial content can reduce clinical fidelity [10, 12].

Overall, generation-based augmentation remains a promising yet delicate strategy for improving deep learning models in data-scarce medical contexts. Future work could explore adaptive sampling strategies—such as dynamically varying synthetic subsets per epoch—to better exploit intra-model diversity and mitigate overfitting to generated features. More recent generative paradigms, including diffusion-based models [9], could also offer finer control over image realism and variability, supporting more clinically reliable augmentation in subsequent research.

6 Limitations

While SinGAN-Seg has the advantage of operating from a single example [19], it still faces certain limitations. The generated images, although diverse, may introduce noise or visual artifacts that could bias the training process. Moreover, the approach does not guarantee that all clinically relevant features are faithfully reproduced. More thorough validation by medical experts therefore remains essential [22].

7 Conclusion

This work examined the contribution of synthetic bronchoscopic data generated by SinGAN-Seg [22] in a supervised classification setting. By systematically varying the proportion of generated images across three architectures of increasing complexity (ResNet-18, ResNet-50, EfficientNet-B7), we assessed their impact on multiple validation metrics.

Our results show that a moderate integration of synthetic data can strengthen robustness and sensitivity. For example, Accuracy increased from 0.69 to 0.70 with ResNet-18 at 33% synthetic augmentation, while EfficientNet-B7 reached its peak with 50% synthetic data (Accuracy 0.68, AUC-ROC 0.73). Beyond these thresholds, however, an excess of artificial samples systematically reduced sensitivity and F1-score, underlining the risk of over-representation. Taken together, these findings suggest that SinGAN-Seg helps mitigate intra-patient redundancy and enrich training variability, but remains a complement rather than a substitute for real annotated data.

In conclusion, this study highlights SinGAN-Seg as a practical one-shot generative approach for medical endoscopy, especially in contexts where data scarcity constrains the deployment of deep learning models. Looking forward, promising directions include extending the evaluation to larger multi-centric datasets, involving clinicians in the validation of synthetic frames, and testing more recent generative paradigms such as diffusion-based models [9], which may provide finer control over variability and realism.

A Quantitative Evaluation Metrics

For completeness, this appendix summarizes the quantitative metrics used to assess the similarity between generated and real bronchoscopic images.

Mean Squared Error (MSE). The MSE measures the average squared difference between corresponding pixel intensities of two images:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2,$$

where x_i and y_i denote pixel values of the generated and reference images, respectively, and N is the total number of pixels. Lower MSE indicates higher similarity.

Peak Signal-to-Noise Ratio (PSNR). Derived from MSE, PSNR expresses the ratio between the maximum possible pixel value (MAX_I) and the noise power:

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{\text{MSE}}\right).$$

Higher PSNR values correspond to better image fidelity.

Structural Similarity Index (SSIM). The Structural Similarity Index (SSIM) measures perceptual similarity between two images by comparing their luminance, contrast, and structural information. It is defined as:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)\,(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)\,(\sigma_x^2 + \sigma_y^2 + C_2)},$$

where μ_x and μ_y are the local means, σ_x^2 and σ_y^2 are the local variances, and σ_{xy} is the covariance between x and y. Constants C_1 and C_2 are included to stabilize the division when the denominators are small. SSIM values range from 0 (no similarity) to 1 (perfect similarity), and the metric correlates more closely with human visual perception than pixel-based measures such as MSE or PSNR.

References

1. Almalioglu, Y., Ozyoruk, K.B., Gokce, A., Incetan, K., Gokceler, G.I., Simsek, M.A., Ararat, K., Chen, R.J., Durr, N.J., Mahmood, F., et al.: Endol2h: deep super-resolution for capsule endoscopy. IEEE Transactions on Medical Imaging **39**(12), 4297–4309 (2020)
2. Amante, E., Ghyselinck, R., Thiberville, L., Trisolini, R., Guisier, F., Delchevalerie, V., Dumas, B., Frénay, B., Duparc, I., Mazellier, N., Farhi, C., Jubert, C., Salaün, M., Lachkar, S.: Human and deep learning predictions of peripheral lung cancer using a 1.3 mm video endoscopic probe. Respirology **n/a**(n/a) (2025). https://doi.org/https://doi.org/10.1111/resp.70057, `https://onlinelibrary.wiley.com/doi/abs/10.1111/resp.70057`

3. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., Jemal, A.: Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians **74**(3), 229–263 (2024)
4. Cai, L., Gao, J., Zhao, D.: A review of the application of deep learning in medical image classification and segmentation. Annals of translational medicine **8**(11), 713 (2020)
5. Ghyselinck, R., Delchevalerie, V., Poitier, P., Frénay, B., Dumas, B.: Deep learning for in vivo bronchial carinae detection in flexible bronchoscopy. In: ECAI 2024, pp. 1527–1534. IOS Press (2024)
6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014)
7. Guo, L., Nahm, W.: Texture synthesis for generating realistic-looking bronchoscopic videos. International Journal of Computer Assisted Radiology and Surgery **18**(12), 2287–2293 (2023)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models (2020)
10. Hu, Q., Yuille, A., Zhou, Z.: Synthetic data as validation (2023), `https://arxiv.org/abs/2310.16052`
11. Kebaili, A., Lapuyade-Lahorgue, J., Ruan, S.: Deep learning approaches for data augmentation in medical imaging: A review. Journal of Imaging **9**(4) (2023). https://doi.org/10.3390/jimaging9040081
12. Khosravi, B., Li, F., Dapamede, T., Rouzrokh, P., Gamble, C.U., Trivedi, H.M., Wyles, C.C., Sellergren, A.B., Purkayastha, S., Erickson, B.J., Gichoya, J.W.: Synthetically enhanced: unveiling synthetic data's potential in medical imaging research. eBioMedicine **104**, 105174 (Jun 2024). https://doi.org/10.1016/j.ebiom.2024.105174, `http://dx.doi.org/10.1016/j.ebiom.2024.105174`
13. Kingma, D.P., Welling, M., et al.: An introduction to variational autoencoders. Foundations and Trends® in Machine Learning **12**(4), 307–392 (2019)
14. Koonce, B.: ResNet 50, pp. 63–72. Apress, Berkeley, CA (2021). https://doi.org/10.1007/978-1-4842-6168-2_6, `https://doi.org/10.1007/978-1-4842-6168-2_6`
15. Lachkar, S., Duparc, I., Piton, N., Dantoing, E., Thiberville, L., Guisier, F., Salaün, M.: Direct endoscopic visualization of small peripheral lung nodules using a miniaturized videoendoscopy probe **29**(10), 914–917
16. Larsen, A.B.L., Sønderby, S.K., Larochelle, H., Winther, O.: Autoencoding beyond pixels using a learned similarity metric (2016)
17. Mishra, S., Panda, R., Phoo, C.P., Chen, C.F.R., Karlinsky, L., Saenko, K., Saligrama, V., Feris, R.S.: Task2sim: Towards effective pre-training and transfer from synthetic data. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9194–9204 (2022)
18. Rohlfs, C.: Generalization in neural networks: A broad survey. Neurocomputing **611**, 128701 (2025)
19. Shaham, T.R., Dekel, T., Michaeli, T.: SinGAN: Learning a Generative Model from a Single Natural Image (Sep 2019), `http://arxiv.org/abs/1905.01164`, arXiv:1905.01164 [cs]

20. Smolarz, B., Łukasiewicz, H., Samulak, D., Piekarska, E., Kołaciński, R., Romanowicz, H.: Lung cancer—epidemiology, pathogenesis, treatment and molecular aspect (review of literature). International Journal of Molecular Sciences **26**(5), 2049 (2025)
21. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: Chaudhuri, K., Salakhutdinov, R. (eds.) Proceedings of the 36th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 97, pp. 6105–6114. PMLR (09–15 Jun 2019), `https://proceedings.mlr.press/v97/tan19a.html`
22. Thambawita, V., Salehi, P., Sheshkal, S.A., Hicks, S.A., Hammer, H.L., Parasa, S., Lange, T.d., Halvorsen, P., Riegler, M.A.: SinGAN-Seg: Synthetic training data generation for medical image segmentation. PLOS ONE **17**(5), e0267976 (May 2022). https://doi.org/10.1371/journal.pone.0267976, publisher: Public Library of Science
23. Wang, Z., Wang, K., Chen, X., Zheng, Y., Wu, X.: A deep learning approach for inter-patient classification of premature ventricular contraction from electrocardiogram. Biomedical Signal Processing and Control **94**, 106265 (2024)
24. Xiao, Z., Lu, J., Wang, X., Li, N., Wang, Y., Zhao, N.: Wce-dcgan: A data augmentation method based on wireless capsule endoscopy images for gastrointestinal disease detection. IET Image Processing **17**(4), 1170–1180 (2023)
25. Zigman Suchsland, M., Kowalski, L., Burkhardt, H.A., Prado, M.G., Kessler, L.G., Yetisgen, M., Au, M.A., Stephens, K.A., Farjah, F., Schleyer, A.M., et al.: How timely is diagnosis of lung cancer? cohort study of individuals with lung cancer presenting in ambulatory care in the united states. Cancers **14**(23), 5756 (2022)