Data Augmentation for Rare Disease Medical Imaging Using Transformer Autoencoders

Ayesha Maniyar[1], Michelle Hoolgeri[1], Gouri N. Banapurmath[1], Tanuja Ratan[1], Uday Kulkarni[1], and Sanjana Patil[1]

Department of Computer Science,
KLE Technological University, Hubballi, India

01fe23bci007@kletech.ac.in, 01fe23bci042@kletech.ac.in, 01fe23bci079@kletech.ac.in, 01fe23bci076@kletech.ac.in, uday_kulkarni@kletech.ac.in, 01fe22bcs069@kletech.ac.in

Abstract. Rare dermatological datasets are often small, unevenly distributed across classes, and only loosely annotated, making it difficult for deep learning models to learn features that generalize well. These limitations create a clear need for augmentation strategies that can strengthen the available training data, especially for the less-represented classes. In this work, we explored a Transformer-based Masked Autoencoder (MAE) to generate additional dermatology images that resemble real samples from the DermaMNIST dataset. The MAE was trained to reconstruct large missing regions in the images, which helped it learn broader lesion structures and class-specific visual cues. After producing the synthetic samples, we merged them with the original dataset to give rare classes more presence and reduce imbalance. We then trained a Data-efficient Image Transformer (DeiT-Tiny) on this expanded dataset and evaluated the model using accuracy and F1-based metrics. The performance improved in several areas: Macro-F1 went from 0.54 to 0.57, weighted-F1 increased from 0.76 to 0.77, and accuracy rose from 0.757 to 0.762. The generated images also appeared to retain important diagnostic patterns when inspected visually, which likely helped the model handle minority classes more reliably. Overall, the MAE-based augmentation method enhanced generalization for rare-disease skin-image classification and offers a practical way to work with limited medical-imaging data.

Keywords: Medical imaging · Data augmentation · Masked Autoencoder · Transformers · DermaMNIST · Rare disease detection

1 Introduction

Medical imaging is an important part of modern healthcare because it lets doctors find diseases early, make accurate diagnoses, and plan effective treatments. Magnetic Resonance Imaging (MRI) [14], Computed Tomography (CT) [14], Ultrasound [22], and Dermoscopy [24,10] are some of the most common ways to look at structural and pathological changes in the body. However, small, unbalanced, and poorly annotated datasets still limit the performance of Artificial

Intelligence (AI) [15] models. This shows how important it is to use stronger data augmentation for improving the model's ability to generalize.

Early Machine Learning (ML) [20] methods in medical imaging mostly relied on hand-crafted features and older classifiers such as Support Vector Machines (SVMs) [3] and Decision Trees [16]. These approaches did work to some extent, but they often missed the fine-grained details and subtle visual cues that medical images typically contain. With the rise of Deep Learning (DL) [13], especially Deep Neural Networks (DNNs) [9], this process changed quite a bit. Instead of manually designing features, models could now learn patterns on their own directly from image data. Within this broader group of deep models, Convolutional Neural Networks (CNNs) [12,21] became widely adopted because they worked particularly well for image-based tasks. Still, despite their success, CNNs tend to overfit when only a small amount of data is available, which reduces their ability to generalize, and this becomes a real concern when dealing with rare diseases.

To solve these problems, traditional methods of augmentation like rotation, flipping, cropping, and changing the brightness have been used to make the data more diverse [20]. More advanced approaches, including SMOTE [1], MixUp [31], CutMix [30], AutoAugment [4], and RandAugment [5,26], further help mitigate class imbalance and improve model regularization. At the same time, generative models like Autoencoders (AEs) [25], Variational Autoencoders (VAEs) [11], and Generative Adversarial Networks (GANs) [7] have shown strong potential for creating realistic synthetic medical data. GAN-based models have improved lesion detection in liver [6] and brain imaging [8], while hybrid architectures such as Disc-VAE [17] have achieved promising results with limited data. More recently, Transformer-based Masked Autoencoders (MAEs) [19,28,2] have emerged as effective self-supervised models that learn rich image representations suitable for reconstruction and augmentation.

In this paper, we propose a Transformer-based MAE fine-tuned on the DermaMNIST subset of the MedMNIST benchmark [29] to generate synthetic dermatological images and rebalance rare disease classes. The augmented dataset was used to retrain a baseline classifier, resulting in higher generalization and fairer performance across categories. Experimental results show 3% improvement in macro-F1 compared with the baseline, confirming the strength of transformer-based augmentation for medical imaging with limited data.

The remainder of this paper is organized as follows. Section 2 reviews augmentation techniques in medical imaging. Section 3 explains the proposed method and experimental setup. Section 4 presents the results and discussion, and Section 5 concludes with key insights and future directions.

2 Background and Related Work

Deep learning (DL) [13] has become a cornerstone of modern medical image analysis due to its ability to learn hierarchical representations directly from raw image data, eliminating the need for hand-crafted features. Despite these advantages, datasets involving rare dermatological conditions remain challenging

because they are typically small, highly imbalanced across classes, and often weakly annotated. These limitations significantly affect the generalization capability of deep models and motivate the use of data augmentation and generative learning strategies to enrich training data while preserving clinically meaningful patterns [10,20,18].

2.1 From Autoencoders to Generative Adversarial Models

Autoencoders (AEs) [25] introduced unsupervised representation learning by encoding input images into latent representations and reconstructing them using a reconstruction objective, as shown in Eq. (1).

$$\mathcal{L}_{\text{AE}} = \|x - \hat{x}\|_2^2 \tag{1}$$

Here, x denotes the input image and $\hat{x}$ represents the reconstructed output. This objective minimizes the pixel-wise reconstruction error and encourages the latent space to retain salient image features.

Denoising Autoencoders (DAEs) [25] extend this idea by reconstructing clean images from corrupted inputs, thereby improving robustness to noise. Variational Autoencoders (VAEs) [11,18] further introduce probabilistic latent modeling by combining reconstruction loss with Kullback–Leibler (KL) regularization, as defined in Eq. (2).

$$\mathcal{L}_{\text{VAE}} = \mathbb{E}_{q_\phi(z|x)}[\|x - \hat{x}\|_2^2] + D_{\text{KL}}(q_\phi(z|x)\|p(z)) \tag{2}$$

GANs [7,27] generate synthetic samples using an adversarial learning paradigm in which a generator and discriminator compete, as expressed in Eq. (3).

$$\min_G \max_D \mathbb{E}_{x\sim p_{\text{data}}}[\log D(x)] + \mathbb{E}_{z\sim p_z}[\log(1 - D(G(z)))] \tag{3}$$

GAN-based approaches have shown promise in medical image synthesis, including liver lesion generation [6] and brain metastasis detection [8]. Hybrid models such as Disc-VAE [17] combine VAE-based latent modeling with adversarial refinement to improve image realism under limited data conditions.

2.2 Masked Autoencoders for Limited-Data Regimes

Masked Autoencoders (MAEs) [19,28,2] reconstruct images from heavily masked inputs, making them well suited for scenarios involving small and imbalanced datasets. The MAE reconstruction loss is defined in Eq. (4).

$$\mathcal{L}_{\text{MAE}} = \|(1 - M) \odot (x - \hat{x})\|_2^2 \tag{4}$$

Here, M denotes a binary mask indicating visible patches, and $\odot$ represents element-wise multiplication. By penalizing reconstruction errors only on masked regions, MAEs are encouraged to model global contextual relationships rather than local textures.

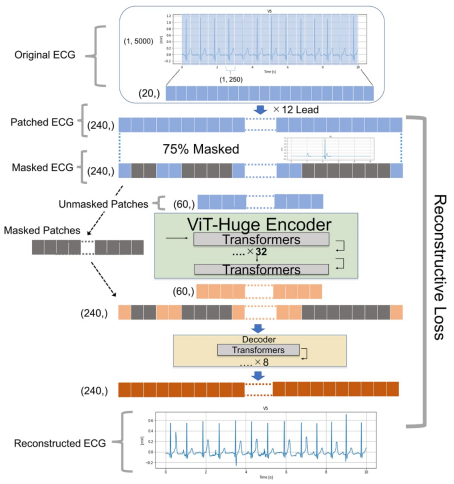

Fig. 1. Masked Autoencoder workflow with patch embedding, random masking, transformer encoding, and reconstruction [19]

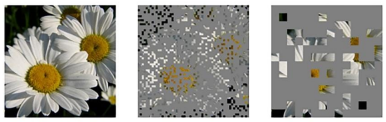

Fig. 2. Example MAE reconstruction showing original, masked, and reconstructed images [28]

The encoder processes only visible patches using the self-attention mechanism defined in Eq. (5).

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5}$$

This formulation enables the model to capture long-range dependencies between image regions. Fig. 1 illustrates the MAE training workflow, while Fig. 2 demonstrates representative reconstruction results.

2.3 Reconstruction Objective and Role in Augmentation

The complete MAE optimization objective is expressed in Eq. (6).

$$\min_{\theta_E, \theta_D} \mathbb{E}_{x \sim \mathcal{D}} \left[\| (1 - M) \odot (x - D(E(M \odot x))) \|_2^2 \right] \tag{6}$$

This objective enables MAEs to reconstruct large missing regions using limited visible context, resulting in synthetic samples that preserve global structure and diagnostically relevant features. Such reconstructions are well suited for augmenting minority classes and improving generalization in rare-disease medical image classification tasks [10,18].

2.4 Motivation

Conventional augmentation techniques introduce only limited pixel-level variations and do not create new semantic content, while GAN-based models often suffer from training instability in extremely small-data regimes. MAEs overcome these limitations by learning contextual relationships across image patches and generating medically coherent synthetic samples from sparse inputs. This capability makes MAEs particularly attractive for addressing class imbalance in rare dermatological datasets and motivates their use as the augmentation backbone in this work.

3 Methodology

This section describes the proposed framework for rare-disease data augmentation using a Transformer-based Masked Autoencoder (MAE). The methodology consists of four main stages: MAE architecture design and training, synthetic image generation and class balancing, Vision Transformer-based classification, and evaluation using standard performance metrics.

3.1 Masked Autoencoder Architecture

The proposed MAE follows an encoder–decoder architecture designed to reconstruct missing image regions from partially observed inputs. Each input image is partitioned into non-overlapping patches, of which 75% are randomly masked. Only the visible patches are processed by the transformer encoder, while the decoder reconstructs the masked patches using the learned latent representation.

The MAE is trained by minimizing the Mean Squared Error (MSE) between the reconstructed and original patches, as defined in Eq. (7).

$$\mathcal{L}_{\text{rec}} = \frac{1}{N} \sum_{i=1}^{N} \|x_i - \hat{x}_i\|_2^2 \tag{7}$$

Here, x_i and $\hat{x}_i$ denote the original and reconstructed patches, respectively, and N represents the number of masked patches. This objective encourages the MAE to capture fine-grained structural and textural patterns that are critical for medical image reconstruction. The overall architecture is illustrated in Fig. 3.

Due to the high masking ratio, the encoder is forced to model long-range dependencies across visible patches rather than relying on local continuity. This enables the MAE to learn global lesion structures, color variations, and subtle

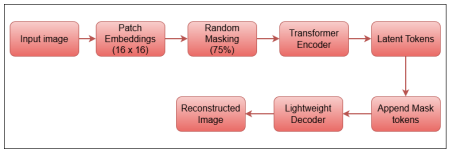

Fig. 3. Masked Autoencoder architecture with patch embedding, random masking, transformer-based encoding of visible patches, mask token insertion, and decoder-based reconstruction

Algorithm 1 Phase 1: Training the Masked Autoencoder

for *epoch* $= 1$ to E_{MAE} **do**
  Shuffle D_{train}
  for each minibatch X of size B_{MAE} **do**
    $X_p \leftarrow P(X)$
    Sample mask $M \sim \text{Bernoulli}(r_{\text{mask}})$
    $X_{\text{vis}} \leftarrow X_p \odot (1 - M)$
    $Z \leftarrow E_\theta(X_{\text{vis}})$
    $\hat{X} \leftarrow D_\phi(Z, M)$
    Compute $\mathcal{L}_{\text{rec}}$ using Eq. (7)
    Update (θ, ϕ) using Adam
  end for
end for

texture patterns commonly observed in dermatological images. The lightweight decoder expands the latent representation to reconstruct the missing regions, resulting in stable and efficient training, particularly suitable for datasets such as DermaMNIST.

Phase 1: MAE Pretraining The first phase of the pipeline (Fig. 4) involves MAE training on the original dataset. The training procedure is summarized in Algorithm 1.

3.2 Synthetic Data Augmentation Pipeline

Once trained, the MAE is employed to generate synthetic dermatological images by reconstructing heavily masked inputs from minority classes. The overall augmentation workflow is illustrated in Fig. 4, which includes preprocessing, MAE training, synthetic image generation, dataset balancing, and classifier fine-tuning.

The number of synthetic samples required for each class c is computed as:

$$N_{\text{aug}}^{(c)} = N_{\max} - N_{\text{orig}}^{(c)} \tag{8}$$

where $N_{\text{orig}}^{(c)}$ denotes the original number of samples in class c, and $N_{\max}$ is the maximum class size. Representative reconstruction results are shown in Fig. 5.

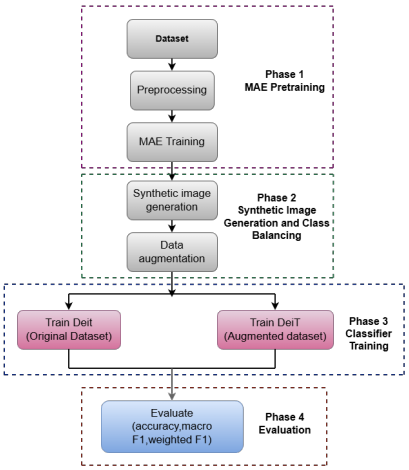

Fig. 4. Overview of the proposed pipeline, including MAE training, synthetic image generation, class balancing, and classifier evaluation

Algorithm 2 Phase 2: Synthetic Image Generation and Class Balancing

```
Compute N_orig(c) and N_max for all classes
for each class c do
    N_aug(c) ← N_max − N_orig(c)
    for k = 1 to N_aug(c) do
        Sample (x, c) from class c
        Generate synthetic sample x_syn using the MAE
        Add (x_syn, c) to D_aug
    end for
end for
```

Phase 2: Synthetic Image Generation and Class Balancing In this phase, synthetic images are generated using the trained MAE and iteratively added to the dataset until class balance is achieved, as described in Algorithm 2.

3.3 Classifier Training and Fine-Tuning

A Vision Transformer-based classifier is trained on the MAE-augmented dataset. Specifically, the DeiT-Tiny architecture [23] is selected due to its efficiency and strong performance in data-limited settings. The classifier is trained for 20 epochs using the Adam optimizer with a batch size of 32 and a learning rate of 1×10^{-4}.

The categorical cross-entropy loss used for classification is defined in Eq. (9).

$$\mathcal{L}_{\text{cls}} = -\frac{1}{M} \sum_{j=1}^{M} \sum_{k=1}^{C} y_{j,k} \log(\hat{y}_{j,k}) \tag{9}$$

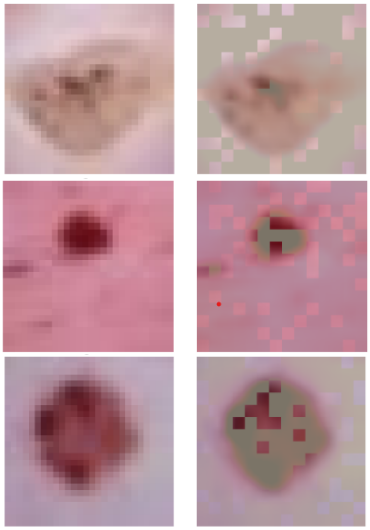

Fig. 5. Original and reconstructed dermatological images following 75% masking, demonstrating the MAE's ability to recover clinically relevant structures

Algorithm 3 Phase 3: Training the DeiT-Tiny Classifier

```
for epoch = 1 to E_cls do
    Shuffle D_aug
    for each minibatch (X, y) do
        ŷ ← f_ψ(X)
        Compute L_cls using Eq. (9)
        Update ψ using Adam
    end for
end for
```

Phase 3: Classifier Training

3.4 Evaluation Metrics

The F1-score for class i is defined as:

$$F1_i = \frac{2P_iR_i}{P_i + R_i} \tag{10}$$

The Macro-F1 score is computed as:

$$F1_{\text{macro}} = \frac{1}{C}\sum_{i=1}^{C} F1_i \tag{11}$$

Phase 4: Model Evaluation

Algorithm 4 Phase 4: Evaluation of Classification Performance

1: **for** each class c **do**
2: Compute precision P_c, recall R_c, and $F1_c$
3: **end for**
4: Compute Macro-F1, Weighted-F1, and Accuracy

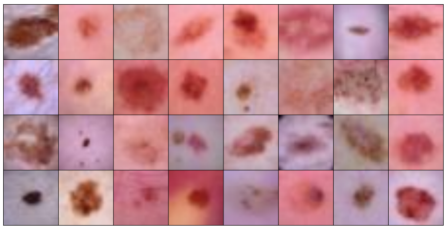

Fig. 6. Representative samples from the DermaMNIST dataset

4 Results

This section presents the experimental results of the proposed MAE-based data augmentation framework. We first describe the dataset and experimental setup, followed by an analysis of MAE reconstruction quality and its effect on dataset balance. Finally, the classification performance before and after augmentation is quantitatively evaluated.

4.1 Dataset Description

Experiments were conducted on the DermaMNIST subset of the MedMNIST benchmark [29], which consists of dermatoscopic skin lesion images belonging to seven diagnostic categories. The dataset exhibits significant class imbalance, with certain classes containing thousands of samples while others contain fewer than two hundred. All images are provided at a standardized resolution of 28×28, preserving key lesion characteristics while maintaining computational efficiency. Representative samples from the dataset are shown in Fig. 6.

4.2 Experimental Setup

All experiments were conducted on the Lightning AI platform using an NVIDIA T4 GPU. The MAE was trained with a learning rate of 1×10^{-4}, a batch size of 64, and a masking ratio of 75%. The DeiT-Tiny classifier was trained for 20 epochs using the same optimizer configuration. To evaluate the impact of data augmentation, the classifier was trained once on the original dataset and once on the MAE-augmented balanced dataset.

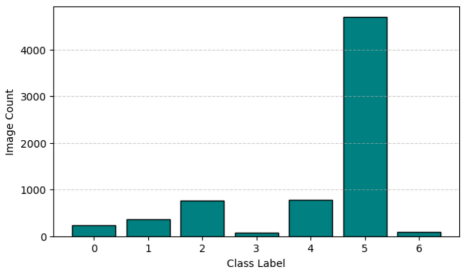

Fig. 7. Class distribution in the original DermaMNIST dataset before augmentation

Table 1. Class distribution before and after MAE augmentation

Class	Original	Augmented	Final Total
C0	807	5898	6705
C1	996	5709	6705
C2	1002	5703	6705
C3	115	6590	6705
C4	1053	5652	6705
C5	6705	0	6705
C6	142	6563	6705

4.3 MAE Reconstruction Quality

The reconstruction capability of the MAE was evaluated using images not seen during training. As illustrated earlier in Fig. 5, the MAE successfully reconstructs lesion structures even under heavy masking. This indicates that the model learns clinically meaningful texture and boundary information, enabling the generation of realistic synthetic dermatological images.

4.4 Effect of MAE Augmentation on Dataset Balance

The original DermaMNIST dataset is highly imbalanced, as shown in Fig. 7, where Class 5 dominates the dataset while several other classes are severely underrepresented.

To mitigate this imbalance, synthetic samples generated by the MAE were added to minority classes until all classes matched the size of the majority class. This resulted in a uniform class distribution across all seven categories, as summarized in Table 1.

Synthetic samples generated by the MAE are shown in Fig. 8. The images appear visually consistent with real dermatological lesions and do not exhibit noticeable artifacts, confirming their suitability for class-balancing augmentation.

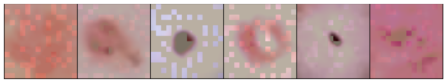

Fig. 8. Synthetic dermatological samples generated by the MAE (one per class)

Table 2. Classifier performance before and after MAE augmentation

Model	Accuracy	Macro-F1	Weighted-F1
Baseline (No Augmentation)	0.757	0.54	0.76
MAE-Augmented	0.762	0.57	0.77

4.5 Classification Performance Before and After Augmentation

Table 2 summarizes the classification performance before and after MAE-based augmentation.

Macro-F1 improves from 0.54 to 0.57, corresponding to an increase of approximately 3%, indicating enhanced balanced performance across all classes. This improvement is particularly significant for underrepresented categories, as Macro-F1 assigns equal importance to each class.

Weighted-F1 increases from 0.76 to 0.77, reflecting a smaller improvement of approximately 1%. This behavior is expected because Weighted-F1 is influenced more strongly by the majority class, which already achieves high performance in the baseline model.

Overall classification accuracy improves from 0.757 to 0.762. Although the absolute gain is modest, the combined improvement in Macro-F1 and stable accuracy indicates that MAE-based augmentation improves minority-class recognition without degrading overall predictive performance.

5 Conclusion and Future Scope

A transformer-based MAE framework for data augmentation in rare disease medical imaging was presented in this work. The suggested method successfully produced lifelike synthetic samples to balance the DermaMNIST dataset by reconstructing heavily masked dermatological images. In medical image classification tasks, transformer-based self-supervised models can improve generalization and lessen class imbalance, as demonstrated by experimental results showing that augmenting with MAE-generated images improved both Macro-F1 and Weighted-F1 scores. The MAE is well suited for clinical applications where annotated samples are hard to come by because of its capacity to learn contextual and structural relationships from sparse data.

To further assess the scalability of transformer-based augmentation, future research will investigate the integration of multi-modal medical data (such as MRI, CT, and histopathology). Diffusion models and conditional generative transformers can be used to increase control over disease-specific synthesis and realism. Furthermore, using explainable AI (XAI) techniques and expanding this

strategy to 3D medical imaging could improve interpretability and trust in clinical settings. There is a lot of promise for developing trustworthy, data-efficient, and equitable AI systems in healthcare in this direction.

The authors would like to acknowledge KLE Technological University for providing the necessary resources and infrastructure.

References

1. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research **16**, 321–357 (2002)
2. Chen, X., et al.: Adaptive masked autoencoder transformer for image classification. arXiv preprint arXiv:2401.00000 (2024)
3. Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning **20**, 273–297 (1995)
4. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Autoaugment: Learning augmentation policies from data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
5. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: Practical automated data augmentation with a reduced search space. In: Advances in Neural Information Processing Systems (NeurIPS) (2020)
6. Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: Synthetic data augmentation using gan for improved liver lesion classification. arXiv preprint arXiv:1803.01229 (2018)
7. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 27, pp. 2672–2680 (2014)
8. Han, C., Murao, K., Noguchi, T., Kawata, Y., Uchiyama, F., Rundo, L., Nakayama, H., Satoh, S.: Learning more with less: Conditional pggan-based data augmentation for brain metastases detection using highly-rough annotation on mr images. arXiv preprint arXiv:1904.00609 (2019)
9. Hinton, G.E., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. Neural Computation **18**(7), 1527–1554 (2006)
10. Islam, T., et al.: A systematic review of deep-learning data augmentation in medical imaging. PLOS ONE (2024)
11. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
12. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Proceedings of the 26th International Conference on Neural Information Processing Systems (NeurIPS). pp. 1097–1105 (2012)
13. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
14. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical Image Analysis **42**, 60–88 (2017)
15. Pinto-Coelho, L.: How artificial intelligence is shaping medical imaging technology: A survey of innovations and applications. Bioengineering **10**(12), 1370 (2023)
16. Quinlan, J.R.: Induction of decision trees. Machine Learning **1**, 81–106 (1986)

17. Rais, K., Haouam, Y., Amroune, M.Y.: Medical image generation techniques for data augmentation: Disc-vae versus gan (2023)
18. Rais, K., et al.: Deep learning approaches for data augmentation in medical imaging: VAEs, GANs, diffusion models. Journal of Imaging **9**(4), 81 (2023)
19. Sawano, S., et al.: Applying masked autoencoder-based self-supervised learning to improve performance with limited data. PLOS ONE (2024)
20. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. Journal of Big Data **6**(1), 60 (2019)
21. Simard, P.Y., Steinkraus, D., Platt, J.C.: Best practices for convolutional neural networks applied to visual document analysis. In: Proc. International Conference on Document Analysis and Recognition (ICDAR) (2003)
22. Smistad, E., Falch, T.L., Bozorgi, M., Elster, A.C., Lindseth, F.: Medical image segmentation on gpus: A comprehensive review. Medical Image Analysis **20**(1), 1–18 (2015)
23. Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., Jégou, H.: Training data-efficient image transformers & distillation through attention. In: Proceedings of the International Conference on Machine Learning (ICML). pp. 10347–10357 (2021)
24. Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data **5**, 180161 (2018)
25. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A.: Extracting and composing robust features with denoising autoencoders. In: Proc. 25th International Conference on Machine Learning (ICML). pp. 1096–1103 (2008)
26. Xu, M., et al.: A comprehensive survey of image augmentation. arXiv preprint arXiv:2312.00000 (2023)
27. Xu, Y., et al.: Application and analysis of generative adversarial networks in medical image analysis: A survey. Artificial Intelligence Review (2024)
28. Xu, Z., et al.: Swin mae: Masked autoencoders for small datasets. arXiv preprint arXiv:2301.00000 (2023)
29. Yang, J., Shi, R., Wei, D., et al.: Medmnist: A large-scale lightweight benchmark for 2d and 3d biomedical image classification. Medical Image Analysis **74**, 102304 (2021)
30. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 6023–6032 (2019)
31. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. In: International Conference on Learning Representations (ICLR) (2018)