



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

DERMAFLUX: Synthetic Skin Lesion Generation with Rectified Flows for Enhanced Image Classification

Stathis Galanakis¹, Alexandros Koliouisis^{1,2}, and Stefanos Zafeiriou¹

¹ Imperial College London, UK
{s.galanakis21,s.zafeiriou}@imperial.ac.uk
² Northeastern University London, UK
alexandros.koliouisis@nulondon.ac.uk

Abstract. Despite recent advances in deep generative modeling, skin lesion classification systems remain constrained by the limited availability of large, diverse, and well-annotated clinical datasets, resulting in class imbalance between benign and malignant lesions and consequently reduced generalization performance. We introduce DERMAFLUX, a rectified flow-based text-to-image generative framework that synthesizes clinically grounded skin lesion images from natural language descriptions of dermatological attributes. Built upon FLUX.1, DERMAFLUX is fine-tuned using parameter-efficient Low-Rank Adaptation (LoRA) on a large curated collection of publicly available clinical image datasets. We construct image-text pairs using synthetic textual captions generated by Llama 3.2, following established dermatological criteria including lesion asymmetry, border irregularity, and color variation. Extensive experiments demonstrate that DERMAFLUX generates diverse and clinically meaningful dermatology images that improve binary classification performance by up to 6% when augmenting small real-world datasets, and by up to 9% when classifiers are trained on DERMAFLUX-generated synthetic images rather than diffusion-based synthetic images. Our ImageNet-pretrained ViT fine-tuned with only 2,500 real images and 4,375 DERMAFLUX-generated samples achieves 78.04% binary classification accuracy and an AUC of 0.859, surpassing the next best dermatology model by 8%. Code, data and pretrained models are available at dermaflux.github.io.

Keywords: Skin lesions · Rectified Flows · Dataset Augmentation

1 Introduction

Skin cancer is among the most common cancers worldwide, posing a significant global health burden. Globally, approximately 331,722 new melanoma cases were reported in 2022 [28], while in the United Kingdom alone, approximately 17,500 new melanoma cases were reported annually, making melanoma the fifth most common cancer in the country [4]. Accurate binary classification not only ensures that malignant lesions are identified early, when treatment is most effective, but also reduces unnecessary referrals for benign lesions.

Recent advances in deep learning have demonstrated strong potential for skin lesion classification, offering scalable support for clinical decision-making. However, deep learning models require large, well-annotated datasets that are difficult to obtain and share. The collection of dermatological images is subject to strict ethical, legal, and privacy regulations, often requiring informed consent and institutional approval. Consequently, publicly accessible datasets remain limited in size and may fail to capture sufficient variability in skin tone, lesion morphology, and rare malignant sub-types. This scarcity leads to class imbalance and reduced generalization capability in deep learning-based lesion classifiers.

Synthetic data generation is a promising solution to mitigate data scarcity. In this context, Bissoto *et al.* [2] used GAN-based methods to generate photorealistic skin lesion images given expected lesion morphology. The advent of diffusion models, particularly Stable Diffusion (SD) [25], enabled higher-fidelity and more diverse text-conditioned synthesis of skin lesion images [1,8]. However, two key challenges remain: (i) reliable control over clinically meaningful lesion attributes during image synthesis (e.g., asymmetry, border irregularity, or color heterogeneity) and (ii) strong semantic alignment between textual descriptions and synthesized visual features. Kim *et al.* [17] incorporated visual conditioning embeddings into the diffusion model to improve conditional synthesis, while **LesionGen** [9] introduced concept-guided diffusion using structured dermatological captions to improve attribute-level prompting. Despite these enhancements, diffusion-based lesion generators inherently rely on stochastic denoising and CLIP-based text encoders with a limited context capacity of 77 tokens [25]. These architectural constraints limit the efficiency and determinism of the generation process.

We propose DERMAFLUX, a rectified flow-based framework for synthetic lesion image generation. Unlike diffusion-based generators that rely on stochastic denoising, rectified flows learn a deterministic transport mapping between noise and data distributions, improving sampling efficiency. Specifically, we fine-tune FLUX.1 [3] via Low-Rank Adaptation (LoRA) [15], enabling parameter-efficient adaptation to skin lesion synthesis while preserving the pretrained model’s generative capacity. FLUX.1 utilizes a CLIP [23] and a T5-XXL [24] text encoder, thus enabling richer contextual conditioning and improved text-image alignment.

Beyond architectural considerations, the design of open and well-annotated training datasets containing image-text pairs is equally critical for medical image synthesis. PanDerm [32] and Derm1M [30] introduced large-scale datasets containing approximately one million dermatology image-text pairs. However, a substantial proportion of PanDerm consists of in-house data that is not fully accessible to the research community, while Derm1M annotations were automatically extracted mainly from medical videos and online forums, potentially introducing noise. In contrast, we construct an open, curated image-text dataset to train DERMAFLUX, derived exclusively from publicly available dermatology datasets with human-verified annotations. Textual descriptions are generated using Llama 3.2 [12] and structured around clinically meaningful skin lesion attributes. Combined with the expressive dual text encoders of FLUX.1, this structured supervision enables finer control over skin lesion characteristics such

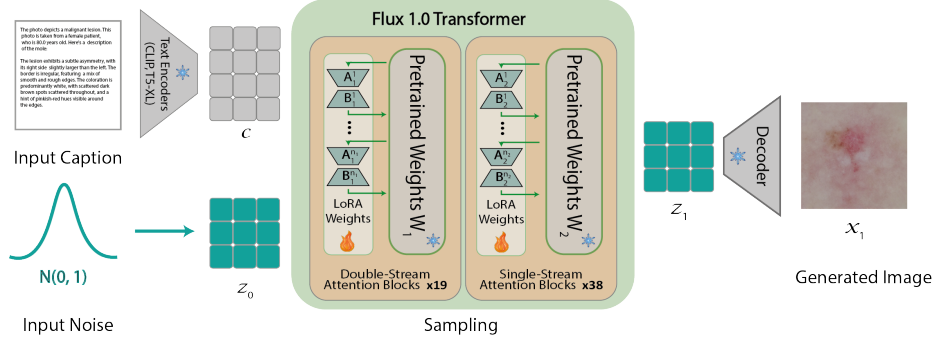


Fig. 1. DERMAFLUX synthesizes a skin lesion image x_1 by transporting Gaussian noise z_0 to a clean latent representation z_1 , conditioned on the input caption. The FLUX.1 [3] backbone is frozen (✳) and only the injected LoRA parameters are trained (🔥).

as asymmetry, border irregularity, and color variation. Motivated by the need to support early screening and reduce unnecessary referrals to dermatologists, the curated dataset focuses on the benign and malignant skin lesion classes.

Overall, our main contributions are:

1. We construct a large-scale dermatology dataset of approximately 500k image-text pairs with structured, attribute-level captions that explicitly encode skin lesion asymmetry, border irregularity, and color variation, enabling fine-grained alignment between clinically meaningful lesion descriptors and corresponding visual features.
2. We develop DERMAFLUX, a rectified flow-based generative framework for synthesizing diverse and semantically consistent skin lesion images, offering a deterministic alternative to diffusion-based skin lesion generation.
3. We experimentally demonstrate that classifiers trained with DERMAFLUX-generated synthetic lesions improve binary classification performance by up to 6% when augmenting limited real-world datasets, achieve up to 9% higher accuracy than diffusion-based synthetic augmentation under identical training settings, outperform existing state-of-the-art classifiers by 8%, and reduce unnecessary referrals.

2 Method

DERMAFLUX is a rectified flow-based framework for text-conditioned lesion synthesis (Fig. 1). We first present our dataset curation efforts (§2.1), followed by an outline of the backbone model and the adaptation strategy (§2.2).

2.1 Dataset Creation

Dataset collection. We curate a large-scale dermatology image corpus by aggregating publicly available datasets for automated skin lesion analysis, including

Table 1. Summary of the aggregated dermatology datasets used for training, reporting imaging modality (clinical , dermoscopic , or both), benign and malignant counts, and training resolution.

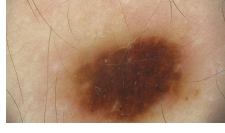
Dataset	Type	Benign	Malignant	Resolution
MedNode [10]		100	70	512×512
HIBA [14]	 	802	751	512×512
Derm12345 [33]		11,120	1,167	512×512
ISIC 2019 [27,5,13]		15,991	9,340	512×512
ISIC 2020 [26]		32,542	584	512×512
Milk10k [19]	 	2,966	7,514	512×512
PAD20 [22]		1,209	1,089	512×512
Kaggle 1 [16]		5,000	4,605	256×256
Kaggle 2 [20]		6,289	5,590	256×256
DDI [6]		485	171	256×256
ISIC 2024 [18]		400,666	393	128×128

the International Skin Imaging Collaboration (ISIC) challenge datasets (2019-2024) [27,5,13,26,19], Derm12345 [33] and PAD20 [22]. Training dataset statistics are summarized in Table 1; sizes are reported before filtering. The collection spans clinical and dermoscopic imagery, capturing substantial variability in lesion appearance, skin tone, and acquisition conditions, and thus provides a diverse foundation for synthetic skin lesion generation. All images undergo unified pre-processing, including resolution standardization and removal of samples with minimum side length below 128 pixels. To accommodate varying source resolutions, the data are grouped into three training scales: 128×128 , 256×256 , and 512×512 . These scales contain 400,901, 22,140, and 85,303 samples, respectively, resulting in a total dataset size of 508,344 samples.

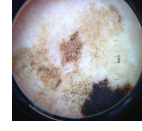
Synthetic caption generation. Original dataset metadata provides diagnostic labels but lacks detailed morphological descriptors. We therefore generate structured, attribute-level captions for each skin lesion image using a vision-language model (VLM). For every image, we prompt the VLM to describe lesion characteristics following the ABC components of the ABCDE dermatological criteria—namely, asymmetry, border irregularity, and color variation. We evaluated three VLMs on a subset of 100 images: Llama 3.2 [12], ChatGPT 5 [21], and Gemini [11]. All models generated comparable structured descriptions, thus Llama 3.2 was selected for its scalability and local deployment capabilities. Two clinical experts confirmed the medical plausibility of 100 captions, with representative benign and malignant examples shown in Fig. 2. The final dataset contains approximately 500k structured image-text pairs used to train DERMAFLUX.

2.2 Model Architecture & Adaptation

DERMAFLUX adapts the pretrained FLUX.1 rectified flow backbone [3] to dermatological lesion synthesis through Low-Rank Adaptation (LoRA) [15].



“The mole in the image is generally symmetrical, with a relatively smooth border. However, there are some areas where the border appears to be slightly irregular, with a few small notches and bumps. The color of the mole is primarily brown, with some areas of lighter and darker shades of brown. There are also some small, darker spots scattered throughout the mole, which may be pigmentation variations or small freckles. Overall, the mole has a relatively uniform color and texture, with some minor variations.”



“The mole exhibits a notable degree of asymmetry, with one half being significantly larger than the other. The border of the mole is irregular and jagged, with a rough, uneven texture. The coloration of the mole is also irregular, featuring a mix of dark brown and black patches scattered throughout the lesion. Additionally, there are areas of lighter skin tone visible, which may indicate a variation in pigmentation. Overall, the mole’s appearance suggests a potential malignancy, warranting further examination and diagnosis.”

Fig. 2. Given a lesion image and its label, benign (*left*) or malignant (*right*), Llama 3.2 generates a synthetic caption using the prompt: “This is an image containing a [label] lesion. Give me a description of this mole regarding its asymmetry, border irregularity, and color.”

FLUX.1 employs a transformer-based architecture to model rectified flows in latent space. It comprises an encoder $\mathcal{E}$, a decoder $\mathcal{D}$, dual text encoders (CLIP [23] and T5-XXL [24]) and a hybrid transformer with 19 double-stream and 38 single-stream blocks. Let $z_0 \sim p_0(z) = \mathcal{N}(\mathbf{0}, \mathbf{I})$ denote Gaussian noise and $z_1 \sim p_1(z)$ the latent data distribution. Given caption embeddings c , rectified flows define the linear interpolation path $z_t = (1 - t)z_0 + tz_1$, $t \in [0, 1]$, and learn a velocity field $v_\theta(z, t, c)$ that deterministically transports z_0 to z_1 . The latent representations z_t are tokenized into visual tokens and fused with text embeddings via cross-attention, enabling joint visual-semantic modeling. The final latent z_1 is then decoded to the synthesized lesion image $x_1 = \mathcal{D}(z_1)$ (Fig. 1).

We specialize FLUX.1 to dermatological synthesis using LoRA. We fine-tune only the attention layers of the transformer while keeping backbone weights frozen. In this way, we preserve the pretrained model’s generative capacity while enabling efficient domain adaptation. For pre-trained weights $\mathbf{W} \in \mathbb{R}^{d \times k}$, LoRA learns a low-rank update $\mathbf{W}' = \mathbf{W} + \frac{\alpha}{r} \mathbf{B} \mathbf{A}$, where $\mathbf{A} \in \mathbb{R}^{r \times k}$, $\mathbf{B} \in \mathbb{R}^{d \times r}$, $r \ll \min(d, k)$, and α scales the update.

We adapt FLUX.1-dev (12B parameters) and set the LoRA hyperparameters to $r = 64$ and $\alpha = 64$. Under this configuration, the LoRA layers amount to approximately 612M additional trainable parameters. Training is performed on four NVIDIA A100 GPUs for five epochs at three resolutions (128×128 , 256×256 , 512×512), using only the predefined training splits with batch size of 32, and an Adam optimizer with $\beta = (0.9, 0.99)$ and learning rate 10^{-4} . At inference, images are generated with a flow-based ODE sampler using 20 interpolating steps. Fig. 3

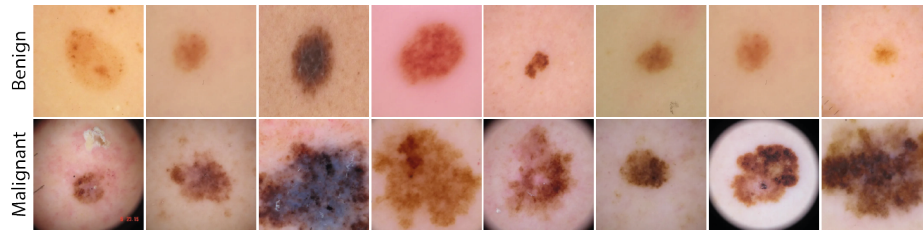


Fig. 3. Examples highlighting our method’s capacity to produce diverse and realistic skin lesion images in both benign (*top*) and malignant (*bottom*) categories.

Table 2. Classification accuracy of ResNeXt and ViT under our two training scenarios.

	Scenario	Derm-T2IM	DERMAFLUX
ResNeXt	(i)	0.6278	0.6398
	(ii)	0.6412	0.6525
ViT	(i)	0.6795	0.7698
	(ii)	0.7091	0.7438

Table 3. Comparison with state-of-the-art dermatology models.

Method	Accuracy
BiomedCLIP [34]	0.5474
DermLIP-ViT [30]	0.6603
MAKE [31]	0.7012
DermLIP-PanDerm [30]	0.7047
DERMAFLUX-ViT(Ours)	0.7804

illustrates examples of benign and malignant generated images, when we used the caption “This is an image containing a *label* lesion.”, where $label \in \{\text{benign}, \text{malignant}\}$.

3 Results

We evaluate DERMAFLUX through a series of downstream binary classification experiments designed to assess the quality and practical utility of the generated synthetic images. All evaluations use a held-out test set of approximately 9,100 benign and malignant lesions aggregated from the predefined test splits of three datasets in our corpus: Kaggle 1, Kaggle 2, and ISIC 2019 (Table 1). We train two representative classifiers, ResNeXt [29] and Vision Transformer (ViT) [7], to ensure model architecture diversity. Specifically, we use ResNeXt-50 ($32 \times 4d$), initialized with ImageNet-1K pretrained weights, and ViT-Base (Patch16-224), initialized with ImageNet-21K pretrained weights. Both classifiers are trained for 40 epochs with batch size 32. Base learning rates are 1×10^{-4} for ResNeXt and 2×10^{-5} for ViT, with a cosine annealing schedule.

First, we compare DERMAFLUX against the diffusion-based Derm-T2IM [8] under controlled real-to-synthetic training ratios to assess relative synthetic image quality (§3.1). Second, we perform a sensitivity analysis by varying the proportion of synthetic data to quantify its contribution to model generalization (§3.2). Finally, we benchmark our best-performing configuration against state-of-the-art dermatology classifiers (§3.3).

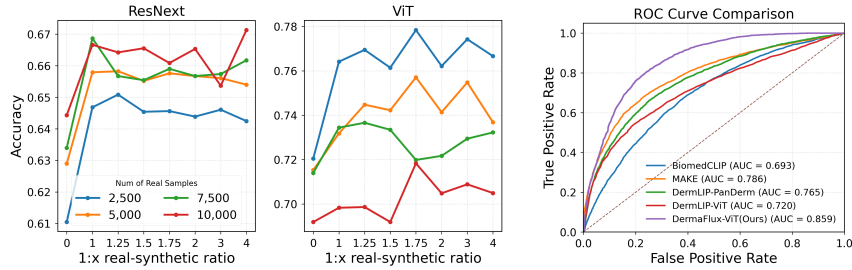


Fig. 4. Test accuracy scores for ResNeXt and **Fig. 5.** DERMAFLUX-ViT across varying real-to-synthetic training rates malignant and benign test data ratios. Results are averaged over five in-samples more reliably than competing models.

3.1 Downstream Evaluation of Synthetic Image Quality

We assess the quality of synthetic images generated by DERMAFLUX through a controlled downstream comparison with Derm-T2IM [8], a synthetic lesion generator built upon SD [25]. Derm-T2IM is designed to enhance dermatology classification in data-scarce settings and 6,000 synthetic lesions are publicly available. We consider two training scenarios: (i) classifiers trained exclusively on synthetic data: 6,000 images from Derm-T2IM and 6,000 from DERMAFLUX; (ii) classifiers trained on a mixture of 2,500 real images and 5,000 synthetic images from each method, with Derm-T2IM samples randomly drawn from the authors’ released dataset. In both scenarios, datasets are balanced between benign and malignant skin lesions, and ResNeXt and ViT are trained under identical training settings.

Across both architectures and scenarios, classifiers trained on DERMAFLUX-generated samples consistently outperform those trained on diffusion-generated samples. Table 2 shows that ResNeXt achieves gains of +1.20% and +1.13% in scenarios (i) and (ii), respectively. The improvement is more pronounced for ViT, gaining +9.03% and +3.47% under the same settings. These results indicate superior downstream utility of the proposed synthetic image generator.

3.2 Impact of Synthetic Data Augmentation

We evaluate synthetic data augmentation by progressively enriching balanced real-image training sets with DERMAFLUX-generated samples. For each experiment, a fixed number of real images (2,500, 5,000, 7,000, 10,000) is randomly sampled from our dataset with equal numbers of benign and malignant cases. Synthetic images are then incrementally added to construct varying real-to-synthetic ratios. Ratios are expressed as 1: x , meaning that for every n real images, $n \cdot x$ synthetic samples are added to the training set. We train ResNeXt and ViT and evaluate performance using Accuracy and ROC-AUC scores.

Synthetic augmentation improves generalization by enriching data diversity, particularly when real data are limited. Fig. 4 shows that, across both architec-

tures, incorporating synthetic data consistently improves performance over real-only training. The largest gains occur when moving to balanced real-synthetic configurations (e.g., 1:1). For example, with 5,000 real samples, ResNeXt improves from 63.2% to 64.68% accuracy when synthetic samples are added, while ViT increases from 71.3% to 76.1%. Improvements are more pronounced in lower-data regimes (2,500 and 5,000 real samples) and remain positive but smaller for larger training sets. We additionally evaluate purely synthetic training (0:1 ratio; 2,500 and 5,000 samples). Even without real images, DERMAFLUX improves accuracy by 0.7-2.3% for ResNeXt and 1.84-4.66% for ViT compared to the 1:0 ratio, suggesting that generated samples capture class-discriminative lesion characteristics beyond simple augmentation effects.

Consistent improvements in ROC-AUC mirror the accuracy trends: e.g., gains of +3.43% and +3.00% for ResNeXt and +1.97% and +4.68% for ViT when moving from a 1:0 to a 1:1 ratio with 2,500 and 5,000 real images, respectively. Since ROC-AUC measures the probability that a malignant lesion receives a higher confidence score than a benign one, these improvements indicate that training with DERMAFLUX-generated images enhances class separability.

3.3 Comparison with State-of-the-Art Classifiers

We evaluate the effectiveness of the proposed synthetic data generation framework by comparing our trained ViT model against state-of-the-art dermatology classifiers: BiomedCLIP [34], MAKE [31], and the DermLIP variants introduced in Derm1M [30]. Table 3 reports the binary classification accuracy for each method on the held-out test set.

Across all evaluated training configurations, DERMAFLUX-ViT consistently outperforms prior state-of-the-art dermatology classifiers. Using 2,500 real images and varying the real-to-synthetic ratio, accuracy ranges from 76% to 78.04%. The best performance is achieved with 4,375 synthetic samples, reaching 78.04% accuracy—an improvement of 8 percentage points over MAKE and DermLIP. These findings indicate that clinically structured synthetic data can effectively compensate for limited real-world supervision, enabling competitive performance with substantially reduced training data.

Figure 5 illustrates the discriminative capability of DERMAFLUX-ViT via ROC analysis. Our best-performing ViT achieves the highest AUC (0.859) among all compared models. Notably, its ROC curve approaches a true positive rate (TPR) close to 1.0 at a false positive rate (FPR) of 0.75, while prior models do not reach similar sensitivity levels. This operating characteristic enables DERMAFLUX-ViT to achieve very high sensitivity while maintaining higher specificity. From a clinical perspective, this reduces both the risk of missed malignant lesions and the number of unnecessary referrals and follow-up examinations.

4 Discussion and Conclusion

We introduced DERMAFLUX, a rectified flow-based text-to-image framework for dermatological lesion synthesis. By fine-tuning FLUX.1 with LoRA and con-

ditioning on structured Llama 3.2 captions, DERMAFLUX produces clinically coherent lesion images with improved attribute control and text–image alignment. In downstream evaluation, DERMAFLUX improves binary classification accuracy by up to 6% when augmenting limited real datasets and by up to 9% over diffusion-based synthetic images. With only 2,500 real images and 4,375 synthetic samples, DERMAFLUX-ViT achieves 78.04% accuracy and an AUC of 0.859, exceeding the next best dermatology model by 8%. Overall, clinically structured synthetic data reduce reliance on large-scale dataset curation while improving generalization and class separability in data-scarce settings.

While DERMAFLUX focuses on benign and malignant lesion generation, it can be extended to multi-class settings through caption conditioning on lesion sub-types. However, this requires further expert validation of sub-type annotations. Future efforts will also focus on enriching the synthetic captions with expert descriptions and additional metadata, including image type, pigmentation characteristics, and other clinically relevant information.

Acknowledgments. This work was supported by the EPSRC Turing AI Fellowship (Grant Ref: EP/Z534699/1): Generative Machine Learning Models for Data of Arbitrary Underlying Geometry (MAGAL).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Akrouf, M., Gyepesi, B., Holló, et al.: Diffusion-Based Data Augmentation for Skin Disease Classification: Impact Across Original Medical Datasets to Fully Synthetic Images. In: Deep Generative Models: Third MICCAI Workshop, DGM4MICCAI 2023, Held in Conjunction with MICCAI 2023, Vancouver, BC, Canada, October 8, 2023, Proceedings. https://doi.org/10.1007/978-3-031-53767-7_10
2. Bissoto, A., Perez, F., Valle, E., Avila, S.: Skin lesion synthesis with generative adversarial networks. In: OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis. Springer International Publishing (2018)
3. Black Forest Labs: FLUX.1: Text-to-Image Generation Model. <https://huggingface.co/black-forest-labs/FLUX.1-dev> (2024)
4. Cancer Research UK: Melanoma Skin Cancer Statistics. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer> (2026)
5. Codella, N.C.F., Gutman, D., et al.: Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC). arXiv preprint arXiv:1710.05006 (2017), <https://arxiv.org/abs/1710.05006>
6. Daneshjou, R., Vodrahalli, K., et al.: Disparities in dermatology AI performance on a diverse, curated clinical image set. *Science Advances* **8**(32), eabq6147 (2022). <https://doi.org/10.1126/sciadv.abq6147>
7. Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. ICLR (2021)

8. Farooq, M.A., Yao, W., Schukat, M., Little, M.A., Corcoran, P.: Derm-T2IM: Harnessing Synthetic Skin Lesion Data via Stable Diffusion Models for Enhanced Skin Disease Classification using ViT and CNN. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (2024). <https://doi.org/10.1109/EMBC53108.2024.10781852>
9. Fayyad, J., Bayasi, N., Yu, Z., Najjaran, H.: LesionGen: A concept-guided diffusion model for dermatology image synthesis. In: MICCAI Workshop on Deep Generative Models. pp. 3–12. Springer (2025)
10. Giotis, I., Molders, N., et al.: MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images. Expert Systems with Applications (2015). <https://doi.org/10.1016/j.eswa.2015.04.034>
11. Google: Gemini. <https://gemini.google.com/> (2025)
12. Grattafiori, A., Dubey, A., et al.: The Llama 3 Herd of Models (2024), <https://arxiv.org/abs/2407.21783>
13. Hernández-Pérez, C., Combalia, M., et al.: BCN20000: Dermoscopic Lesions in the Wild. Scientific Data **11**(1), 641 (2024). <https://doi.org/10.1038/s41597-024-03361-8>
14. Hospital Italiano de Buenos Aires: Hospital Italiano de Buenos Aires - Skin Lesions Images (2019–2022) (2023). <https://doi.org/10.34970/587329>
15. Hu, E.J., Shen, Y., et al.: LoRA: Low-Rank Adaptation of Large Language Models. In: International Conference on Learning Representations (2022), <https://openreview.net/forum?id=nZeVKeeFYf9>
16. Javid, M.H.: Melanoma Skin Cancer Dataset of 10000 Images (2022). <https://doi.org/10.34740/KAGGLE/DSV/3376422>
17. Kim, M., Yoo, J., Kwon, S., et al.: Diffusion-based skin disease data augmentation with fine-grained detail preservation and interpolation for data diversity. Plos one (2025)
18. Kurtansky, N., Rotemberg, V., et al.: ISIC 2024 - Skin Cancer Detection with 3D-TBP. <https://kaggle.com/competitions/isic-2024-challenge> (2024), kaggle
19. MILK study team: MILK10k. ISIC Archive dataset (2025). <https://doi.org/10.34970/648456>
20. Mittal, B.: Melanoma Cancer Image Dataset, <https://www.kaggle.com/datasets/bhavesmittal/melanoma-cancer-dataset/data>
21. OpenAI: ChatGPT (GPT-4.5 or GPT-4o version). <https://chat.openai.com/>
22. Pacheco, A.G.C., Lima, G.R., et al.: PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones (2020). <https://doi.org/10.17632/zr7vgbcyr2.1>
23. Radford, A., Kim, J.W., et al.: Learning Transferable Visual Models From Natural Language Supervision. In: ICLR. pp. 8748–8763 (2021), <https://proceedings.mlr.press/v139/radford21a.html>
24. Raffel, C., Shazeer, N., Roberts, A., et al.: Exploring the limits of transfer learning with a unified text-to-text transformer. J. Mach. Learn. Res. **21**(1) (Jan 2020)
25. Rombach, R., Blattmann, A., et al.: High-Resolution Image Synthesis With Latent Diffusion Models. In: CVPR. pp. 10684–10695 (June 2022)
26. Rotemberg, V., Kurtansky, N., Betz-Stablein, B., et al.: A patient-centric dataset of images and metadata for identifying melanomas using clinical context. Scientific Data **8**, 34 (2021). <https://doi.org/10.1038/s41597-021-00815-z>
27. Tschandl, P., Rosendahl, C., Kittler, H.: The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions. Scientific Data **5**, 180161 (2018). <https://doi.org/10.1038/sdata.2018.161>

28. Wang, M., Gao, X., Zhang, L.: Recent global patterns in skin cancer incidence, mortality, and prevalence. *Chinese Medical Journal* **138**(2), 185–192 (2025). <https://doi.org/10.1097/CM9.0000000000003416>, epub 2024 Dec 17
29. Xie, S., Girshick, R., et al.: Aggregated Residual Transformations for Deep Neural Networks. In: *CVPR* (2017)
30. Yan, S., Hu, M., et al.: Derm1m: A million-scale vision-language dataset aligned with clinical ontology knowledge for dermatology. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 12681–12690 (2025)
31. Yan, S., Li, X., Hu, M., et al.: MAKE: Multi-Aspect Knowledge-Enhanced Vision-Language Pretraining for Zero-shot Dermatological Assessment (2025), <https://arxiv.org/abs/2505.09372>
32. Yan, S., Yu, Z., Primiero, C., et al.: A multimodal vision foundation model for clinical dermatology. *Nature Medicine* pp. 1–12 (2025)
33. Yilmaz, A., Yasar, S.P., Gencoglan, G., Temelkuran, B.: DERM12345: A Large, Multisource Dermatoscopic Skin Lesion Dataset with 40 Subclasses. *Scientific Data* (2024). <https://doi.org/10.1038/s41597-024-04104-3>
34. Zhang, S., Xu, Y., Usuyama, N., et al.: A Multimodal Biomedical Foundation Model Trained from Fifteen Million Image–Text Pairs. *NEJM AI* **2**(1) (2024). <https://doi.org/10.1056/AIoa2400640>