LesionGen: A Concept-Guided Diffusion Model for Dermatology Image Synthesis

Jamil Fayyad[1⋆], Nourhan Bayasi[2], Ziyang Yu[3], and Homayoun Najjaran[1]

[1] University of Victoria, Canada
[2] University of British Columbia, Canada
[3] University of Toronto, Canada
jfayyad@uvic.ca

Abstract. Deep learning models for skin disease classification require large, diverse, and well-annotated datasets. However, such resources are often limited due to privacy concerns, high annotation costs, and insufficient demographic representation. While text-to-image diffusion probabilistic models (T2I-DPMs) offer promise for medical data synthesis, their use in dermatology remains underexplored, largely due to the scarcity of rich textual descriptions in existing skin image datasets. In this work, we introduce LesionGen, a clinically informed T2I-DPM framework for dermatology image synthesis. Unlike prior methods that rely on simplistic disease labels, LesionGen is trained on structured, concept-rich dermatological captions derived from expert annotations and pseudo-generated, concept-guided reports. By fine-tuning a pretrained diffusion model on these high-quality image-caption pairs, we enable the generation of realistic and diverse skin lesion images conditioned on meaningful dermatological descriptions. Our results demonstrate that models trained solely on our synthetic dataset achieve classification accuracy comparable to those trained on real images, with notable gains in worst-case subgroup performance. Code and data are available here.

Keywords: Synthetic Data Generation · Text-to-Image Diffusion Models · Dermatology · Skin Lesion Classification.

1 Introduction

Deep learning (DL) has revolutionized medical image analysis, offering unprecedented accuracy in disease detection [24] and classification [2,4,3,5,10,8]. However, these models are inherently data-hungry, requiring vast amounts of high-quality, diverse training images to generalize well [9,11]. In dermatology, this need is particularly acute because skin disease classification relies on nuanced visual cues that can vary across demographics, lighting conditions, and disease progression. Yet, real-world medical datasets are often limited due to privacy concerns, ethical restrictions, and the high cost of expert annotations. The scarcity

⋆ Corresponding author

of diverse, labeled skin disease images presents a significant barrier to developing robust DL models.

To address this issue, synthetic data generation has emerged as a promising solution, with generative models offering a powerful tool for augmenting medical datasets. Among these, diffusion probabilistic models (`DPMs`) [13] have demonstrated outstanding capabilities in generating high-fidelity images across diverse domains, including image synthesis [14,25], translation [16], classification [17] and segmentation [23]. Particularly noteworthy are text-to-image diffusion models (`T2I-DPMs`), which leverage natural language descriptions to generate realistic images aligned with specific conditions, achieving state-of-the-art performance. For instance, de Wilde et al. [7] adapted pre-trained `T2I-DPMs` to medical imaging through textual inversion, successfully generating diagnostically accurate images across a range of modalities. Similarly, Chen et al. [6] introduced EyeDiff, a `T2I-DPM` trained on a variety of ophthalmic datasets, demonstrating its effectiveness in generating images of rare eye diseases to address the critical issue of data imbalance in diagnostic models.

However, despite the promising potential of `T2I-DPMs` in medical imaging, their application to dermatology faces a significant limitation: unlike other medical modalities, where images are typically accompanied by detailed textual reports [22], skin disease datasets often lack such structured descriptions. This absence of rich textual metadata makes it challenging to fully leverage `T2I-DPMs` for data synthesis in dermatology. To date, only two studies have applied `T2I-DPMs` to skin lesion synthesis [19,1], but both relied on overly simplistic textual descriptions limited to disease labels; e.g., '`an image of <label-only>`', where `<label-only>` is replaced by the name of the corresponding skin disease. This approach is insufficient, as it fails to capture the rich, nuanced visual features that define each condition, such as variations in texture, color, shape, and progression. These features are critical for accurate lesion classification and diagnosis, and without them, the generated images lack the diversity and specificity needed to enhance model performance effectively.

In this work, we propose `LesionGen`, a novel `T2I-DPM` framework for concept-driven skin image synthesis, primarily aimed at improving worst-case performance across subgroups. Our method generates rich image–caption pairs by producing structured, clinically meaningful descriptions of dermatological images. Specifically, we employ two concept-based text generation strategies to create these captions: (1) a strategy based on expert dermatological descriptions, where each image is annotated with seven clinically relevant diagnostic attributes; and (2) a strategy using pseudo-generated dermatological descriptions, where a vision-language model is guided to produce detailed medical text conditioned on these dermatological concepts. These concept-grounded captions are paired with their corresponding images to form high-quality training data for `LesionGen`, enabling conditional image generation. Our results demonstrate that models trained solely on this synthetic data achieve competitive classification accuracy, with notable improvements in worst-case subgroup performance.

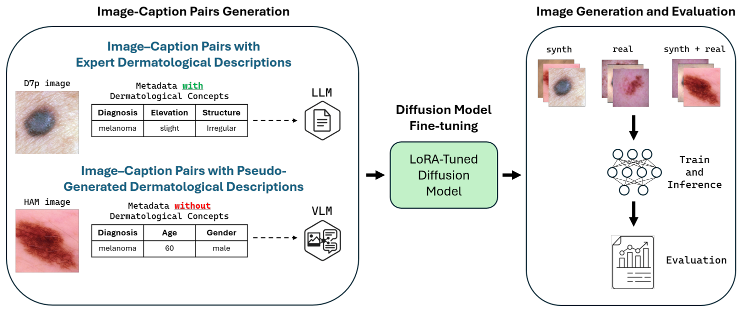

Fig. 1. Overview of LesionGen. Lesion images are paired with either expert dermatological descriptions transformed into text by an LLM, or pseudo-captions generated from image features using a VLM, forming a unified multimodal dataset. A LoRA-tuned diffusion model is trained on the multimodal dataset, and the trained model generates diverse, class-consistent images, which are then evaluated in the downstream classification task.

2 Methodology

We propose LesionGen, a dermatology image synthesis using a fine-tuned T2I-DPM. As shown in Fig. 1, the framework constructs image–caption pairs from expert annotations and pseudo-generated descriptions, then fine-tunes the diffusion model on these pairs for image generation and downstream evaluation.

2.1 Training Data Construction

Image–Caption Pairs with Expert Dermatological Descriptions. To construct clinically grounded pairs, we use the D7P dataset [15], which contains 1,926 dermoscopic images spanning six diagnostic classes: nv, mel, bcc, df, bkl, and vasc. Each image is paired with structured metadata describing seven clinically meaningful attributes (referred to as *concepts*), including pigmentation, lesion elevation, and structure. To transform this metadata into natural language, we prompt an LLM model with a structured template (Fig. 2–top), generating dermatologist-style captions that serve as conditioning text for diffusion training.

Image–Caption Pairs with Pseudo-Generated Dermatological Descriptions. The HAM10000 (HAM) dataset [21] is a widely used benchmark in dermatology, comprising over 10,000 dermoscopic images across seven diagnostic classes (the six from D7P, plus akiec). Although HAM includes basic metadata, such as patient age, gender, lesion type, and anatomical location, it lacks corresponding textual descriptions and, critically, does not provide annotations for key dermatological concepts, as in D7P, that are important to clinical diagnosis. To address this, we leverage a VLM model to generate structured dermatological descriptions that draw on both the available metadata (e.g., lesion label, patient age, gender) and the visual content of each image. Rather than prompting the

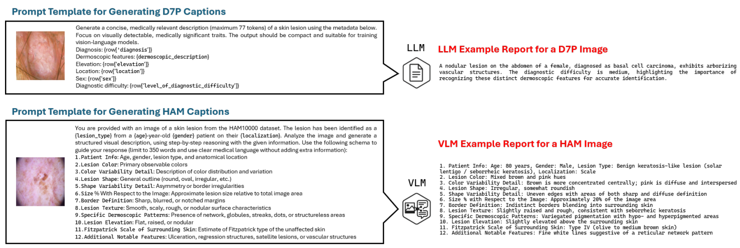

Fig. 2. Prompt templates and example outputs for dermatological caption generation from (top) expert-annotated D7P data and (bottom) pseudo-generated HAM data.

VLM to produce free-form clinical descriptions, we craft targeted prompts that explicitly instruct the model to describe each image using the seven diagnostic concepts defined in the D7P dataset (e.g., pigmentation, elevation, structure). Although these attributes are not explicitly annotated in HAM, our prompting strategy enables the VLM to generate clinically meaningful, concept-based descriptions that enhance the dataset's utility for diffusion model fine-tuning. Fig. 2–bottom illustrates the prompting strategy and an example output.

2.2 Diffusion Model Fine-Tuning

We fine-tune a pretrained `DPM` using a combined set of expert and pseudo-generated image–caption pairs from D7P and HAM. The resulting model, `LesionGen`, is trained to generate realistic and diverse skin lesion images conditioned on dermatological descriptions. To improve alignment between medical language and visual features, we leverage CLIP for text conditioning and apply Low-Rank Adaptation (LoRA) for efficient fine-tuning. Formally, `LesionGen` samples an image x_0 given a text embedding c as $x_0 \sim p_\theta(x_0 \mid c)$, where c encodes the structured dermatological description.

3 Experiments and Results

We evaluate how synthetic skin lesion images generated by `LesionGen` affect downstream classification performance. Specifically, we ask: (1) How effective are these synthetic images for training a CNN classifier (e.g., ResNet18 [12])? and (2) How does prompt design influence the utility of generated images?

3.1 Experimental Setup

We fine-tune a pretrained Stable Diffusion v1.4 model [18] on our multimodal dataset of image–caption pairs (see Section 2), which makes our `LesionGen` model. In our main experimental setting, we use **rich and balanced prompts** as input to `LesionGen` during image generation. We refer to this configuration as `LesionGen-R&B`. Rich refers to the prompts being identical to those used during diffusion model fine-tuning, whereas balanced refers to addressing class imbalance during image generation by ensuring each lesion class is associated with an equal number of prompts. Since some classes are underrepresented in the original dataset, we use GPT-4o to generate paraphrased versions of existing prompts for those classes. These paraphrases retain the original clinical meaning while introducing linguistic variation; e.g., turning '`A nodular melanoma featuring diffuse irregular pigmentation, irregular dots`' into '`A nodular melanoma showing uneven pigmentation, scattered dots`'. Using `LesionGen-R&B`, we generate a class-balanced synthetic dataset with 500 samples per class. A ResNet18 classifier is then trained on different combinations of the synthetic and real data, and evaluated on held-out real test sets from D7P and HAM.

3.2 Baseline and Competing Method

We compare `LesionGen-R&B` against two baselines. The first is a **real-only (upper bound)** configuration, where the ResNet18 classifier is trained solely on real images. The second is a prior SOTA method [1], referred to as **p-SOTA**, in which the diffusion model is both **fine-tuned and sampled** using only static prompts that only include the label (e.g., `melanoma`) without any descriptive captions.

3.3 Implementation Details

Prompt Generation. We use GPT-4o in two modes: text-only for D7P (from structured metadata) and vision–language for HAM (using base64-encoded images with structured instructions). Prompts are generated with temperature 0.3 and a 77-token limit to comply with CLIP's tokenization limits.
Diffusion Model Fine-Tuning. We use Stable Diffusion v1.4, pretrained on the LAION-2B dataset [20], and fine-tune it using LoRA (Low-Rank Adaptation) on our image-caption dataset. The LoRA training uses a rank of 64, learning rate of 1e-5, and a constant scheduler. Training is performed for 15,000 steps using mixed precision (fp16), gradient checkpointing, and image augmentations like cropping and horizontal flipping. The output resolution is 256×256. We qualitatively monitor progress by evaluating validation prompts every five epochs.
Downstream Classification. The downstream classifier is a ResNet18 trained from scratch. Input images are resized to 224×224, normalized with a mean and standard deviation of 0.5, and randomly flipped during training. We use stochastic gradient descent with momentum 0.9, initial learning rate 0.01, and step decay (factor 0.1 every 10 epochs). Training uses a batch size of 32 and early stopping with patience of 5 epochs. Our code is implemented in PyTorch using the HuggingFace Diffusers library and OpenAI APIs.

3.4 Main Results

We train the ResNet18 model on a combined dataset consisting of up to 250 real images per class (when available), supplemented with synthetic images generated by `LesionGen-R&B` to reach a total of 500 training samples per class. This

Table 1. Classification performance on the D7P test set. We report overall accuracy and per-class precision. The best results are **bolded** in green-shaded cells.

Experiment	Overall Accuracy	bcc	bkl	df	mel	nv	vasc
p-SOTA (synth+real)	0.450	**0.333**	0.125	0.250	0.627	0.712	0.250
real-only	0.650	0.214	**0.333**	0.000	0.636	0.678	0.000
Ours							
`LesionGen-R&B` (synth+real)	**0.653**	0.188	0.316	**0.500**	**0.658**	**0.721**	**0.251**

Table 2. Classification performance on the HAM test set. We report overall accuracy and per-class precision. The best results are **bolded** in green-shaded cells.

Experiment	Overall Accuracy	akiec	bcc	bkl	df	mel	nv	vasc
p-SOTA (synth+real)	0.587	0.231	0.364	0.343	0.077	0.313	**0.954**	0.434
real-only	0.737	0.273	**0.488**	0.527	0.000	0.594	0.812	**1.000**
Ours								
`LesionGen-R&B` (synth+real)	**0.756**	**0.375**	0.421	**0.538**	**0.418**	**0.622**	0.807	0.315

setup is referred to as (**synth+real**). The classification performance measured by overall accuracy and per-class precision on the D7P and HAM test sets is reported in Tables 1 and 2, respectively.

On the **D7P test set** (Table 1), the classifier trained on `LesionGen-R&B`-generated data achieves the highest overall accuracy (65.3%), slightly outperforming the real-only model (65.0%) and significantly surpassing the p-SOTA baseline (45.0%). Beyond overall accuracy, our method yields major improvements in worst-class performance. Notably, the precision for the df class increases from 0.000 (real-only) and 0.250 (p-SOTA) to 0.500 with `LesionGen-R&B`, demonstrating that our rich and balanced generation strategy enables the classifier to recover performance on previously underrepresented classes. Additionally, `LesionGen-R&B` provides the highest precision in 4 out of 6 classes, including mel and nv, which are critical in clinical diagnosis.

On the **HAM test set** (Table 2), the classifier achieves the best overall accuracy (75.6%) with our approach, outperforming the real-only (73.7%) and p-SOTA (58.7%) scenarios. The improvements in rare or challenging classes are particularly notable: df improves from 0.000 (real-only) and 0.077 (p-SOTA) to 0.418 with our method, and akiec rises from 0.273 to 0.375. Our method also leads to the best precision in 5 out of 7 classes, showing that `LesionGen-R&B`-generated data improves class-wise consistency without overfitting to majority classes. While vasc and nv show a slight drop compared to real-only, they remain strong overall, and this trade-off results in a more balanced and robust classifier.

3.5 Ablation Studies

We conduct ablation studies to answer three questions: (A) Can synthetic data alone yield strong performance? (B) What is the impact of prompt balancing

Table 3. Ablation study results on the D7P test set, reporting overall accuracy and per-class precision. The best results in each ablation are **bolded** in green-shaded cells.

Experiment	Overall Accuracy	bcc	bkl	df	mel	nv	vasc
Ablation A: Impact of Using Synthetic Data Alone							
p-SOTA (synth only)	0.446	0.056	0.113	0.000	0.636	**0.715**	**0.111**
LesionGen-R&B (synth only)	**0.551**	**0.100**	**0.143**	0.000	**0.652**	0.690	0.056
Ablation B: Impact of Removing Prompt Balancing							
LesionGen-R (synth only)	0.577	0.000	**0.250**	**0.059**	**0.623**	0.651	0.050
LesionGen-R (synth+real)	**0.611**	**0.250**	0.067	0.000	0.596	**0.670**	**0.125**
Ablation C: Impact of Using Static, Label-Only Prompts							
LesionGen-S (synth only)	0.324	0.000	0.047	0.028	0.280	0.692	0.067
LesionGen-S (synth+real)	**0.637**	**0.077**	**0.222**	**0.300**	**0.597**	**0.733**	**0.300**

Table 4. Ablation study results on the HAM test set, reporting overall accuracy and per-class precision. The best results in each ablation are **bolded** in green-shaded cells.

Experiment	Overall Accuracy	akiec	bcc	bkl	df	mel	nv	vasc
Ablation A: Impact of Using Synthetic Data Alone								
p-SOTA (synth only)	0.315	0.009	0.126	0.130	0.000	0.236	**0.978**	**0.091**
LesionGen-R&B (synth only)	**0.524**	**0.145**	**0.188**	**0.210**	**0.045**	**0.489**	0.681	0.080
Ablation B: Impact of Removing Prompt Balancing								
LesionGen-R (synth only)	0.428	0.000	0.078	0.169	0.000	0.164	0.927	**1.000**
LesionGen-R (synth+real)	**0.585**	**0.230**	**0.389**	**0.312**	0.000	**0.298**	**0.938**	0.475
Ablation C: Impact of Using Static, Label-Only Prompts								
LesionGen-S (synth only)	0.176	0.011	0.211	0.068	0.000	0.214	**1.000**	0.000
LesionGen-S (synth+real)	**0.608**	**0.239**	**0.370**	**0.405**	**0.065**	**0.326**	0.942	**0.512**

on class-wise and overall accuracy? and (C) How much does prompt enrichment improve results over static, label-only prompts?

Ablation A: Impact of Using Synthetic Data Alone. In ablation study A, we remove all real training images and train the ResNet18 model solely on synthetic data generated by either p-SOTA or LesionGen-R&B. As shown in Tables 3 and 4, this leads to a clear drop in both overall accuracy and worst-class precision compared to the synth+real setting (Tables 1 and 2). For instance, on the HAM dataset, training on p-SOTA synthetic images results in just 31.5% accuracy, whereas synthetic images from LesionGen-R&B yield a significantly higher 52.4%. While performance still falls short of configurations that include real data, these results indicate that our concept-driven generation approach produces higher-quality and more informative samples than prior methods. Overall, this ablation confirms that synthetic data alone is not sufficient, but can be highly effective when used in combination with real samples.

Ablation B: Impact of Removing Prompt Balancing. In this ablation, we generate synthetic data using rich, concept-guided prompts but without applying prompt balancing for underrepresented classes. As shown in Tables 3 and 4, removing balancing results in a decline in overall accuracy and highly inconsistent class-wise performance, particularly when compared to the full LesionGen-R&B setting (Tables 1 and 2) where balancing was applied. Furthermore, the classifier

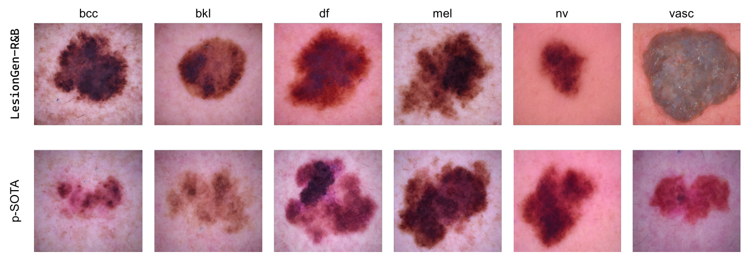

Fig. 3. Examples of synthetic skin lesion images generated from `LesionGen-R&B` (top) and p-SOTA (bottom) across six classes.

performs well on majority classes; e.g., on HAM, precision for nv reaches 93.8% in the synth+real setting, and vasc reaches 100% even with synthetic data alone. However, performance on minority classes could collapse entirely; e.g., df.

Ablation C: Impact of Using Static, Label-Only Prompts. Finally, we test the necessity of the rich prompts by replacing them with static, label-only prompts. In the synth-only setting, performance deteriorates markedly, with accuracy dropping to 32.4% on D7P (Table 3) and 17.6% on HAM (Table 4), with multiple classes exhibiting zero or near-zero precision. Even in the synth+real setting, performance lags behind the rich and balanced prompts configuration (Tables 1 and 2), demonstrating that label-only prompts lack the semantic richness needed to guide the diffusion model toward clinically meaningful generation.

3.6 Visualization Results

Fig. 3 shows samples from `LesionGen-R&B` (top) and p-SOTA (bottom). Despite visual similarity, our method's superior classification performance suggests it captures subtle, clinically relevant features beyond human perception.

4 Conclusion

In this work, we demonstrate the effectiveness of combining text-to-image diffusion models with concept-guided dermatological prompts for generating high-quality synthetic skin lesion images. Unlike prior approaches that rely on simple label-based conditioning, our method leverages rich, structured descriptions aligned with clinical concepts, and balances class representation through prompt paraphrasing. This design enables the generation of semantically diverse and class-balanced datasets that complement real-world dermatology benchmarks. Training a ResNet18 on `LesionGen`'s outputs significantly boosts classification performance, especially for underrepresented groups, outperforming the baselines. Future work includes expanding to skin tone diversity, interactive refinement, and multi-modal conditioning.

References

1. Akrout, M., Gyepesi, B., Holló, P., Poór, A., Kincső, B., Solis, S., Cirone, K., Kawahara, J., Slade, D., Abid, L., et al.: Diffusion-based data augmentation for skin disease classification: Impact across original medical datasets to fully synthetic images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 99–109 (2023)
2. Bayasi, N., Du, S., Hamarneh, G., Garbi, R.: Continual-gen: Continual group ensembling for domain-agnostic skin lesion classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 3–13 (2023)
3. Bayasi, N., Fayyad, J., Bissoto, A., Hamarneh, G., Garbi, R.: Biaspruner: Debiased continual learning for medical image classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 90–101. Springer (2024)
4. Bayasi, N., Hamarneh, G., Garbi, R.: Culprit-prune-net: Efficient continual sequential multi-domain learning with application to skin lesion classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 165–175. Springer (2021)
5. Bayasi, N., Hamarneh, G., Garbi, R.: Continual-zoo: Leveraging zoo models for continual classification of medical images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4128–4138 (2024)
6. Chen, R., Zhang, W., Liu, B., Chen, X., Xu, P., Liu, S., He, M., Shi, D.: Eyediff: text-to-image diffusion model improves rare eye disease diagnosis. arXiv preprint arXiv:2411.10004 (2024)
7. De Wilde, B., Saha, A., de Rooij, M., Huisman, H., Litjens, G.: Medical diffusion on a budget: textual inversion for medical image generation. Medical Imaging with Deep Learning (MIDL) (2023)
8. Elkhayat, M., Mahmoud, M., Fayyad, J., Bayasi, N.: Foundation models as class-incremental learners for dermatological image classification. arXiv preprint arXiv:2507.14050 (2025)
9. Fayyad, J.: Out-of-distribution detection using inter-level features of deep neural networks. Ph.D. thesis, University of British Columbia (2023)
10. Fayyad, J., Alijani, S., Najjaran, H.: Empirical validation of conformal prediction for trustworthy skin lesions classification. Computer Methods and Programs in Biomedicine **253**, 108231 (2024)
11. Fayyad, J., Gupta, K., Mahdian, N., Gruyer, D., Najjaran, H.: Exploiting classifier inter-level features for efficient out-of-distribution detection. Image and Vision Computing **142**, 104897 (2024)
12. 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)
13. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems (NeurIPS) **33**, 6840–6851 (2020)
14. Jiang, L., Mao, Y., Wang, X., Chen, X., Li, C.: Cola-diff: Conditional latent diffusion model for multi-modal mri synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 398–408 (2023)
15. Kawahara, J., Daneshvar, S., Argenziano, G., Hamarneh, G.: Seven-point checklist and skin lesion classification using multitask multimodal neural nets. IEEE journal of biomedical and health informatics **23**(2), 538–546 (2018)

16. Kui, X., Liu, B., Sun, Z., Li, Q., Zhang, M., Liang, W., Zou, B.: Med-lvdm: Medical latent variational diffusion model for medical image translation. Biomedical Signal Processing and Control **106**, 107735 (2025)
17. Prusty, M.R., Sudharsan, R.M., Anand, P.: Enhancing medical image classification with generative ai using latent denoising diffusion probabilistic model and wiener filtering approach. Applied Soft Computing **161**, 111714 (2024)
18. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models (2022)
19. Sagers, L.W., Diao, J.A., Melas-Kyriazi, L., Groh, M., Rajpurkar, P., Adamson, A.S., Rotemberg, V., Daneshjou, R., Manrai, A.K.: Augmenting medical image classifiers with synthetic data from latent diffusion models. arXiv preprint arXiv:2308.12453 (2023)
20. Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., Coombes, T., Katta, A., Mullis, C., Wortsman, M., et al.: Laion-5b: An open large-scale dataset for training next generation image-text models. Advances in neural information processing systems **35**, 25278–25294 (2022)
21. 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**(1), 1–9 (2018)
22. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2097–2106 (2017)
23. Wu, J., Fu, R., Fang, H., Zhang, Y., Yang, Y., Xiong, H., Liu, H., Xu, Y.: Medsegdiff: Medical image segmentation with diffusion probabilistic model. In: Medical Imaging with Deep Learning (MIDL). pp. 1623–1639 (2024)
24. Xie, J., Chen, X., Zhao, Y., Meng, Y., Zhao, H., Nguyen, A., Li, X., Zheng, Y.: Multi-disease detection in retinal images guided by disease causal estimation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. vol. LNCS 15001, pp. 743–753. Springer Nature Switzerland (October 2024)
25. Yellapragada, S., Graikos, A., Prasanna, P., Kurc, T., Saltz, J., Samaras, D.: Pathldm: Text conditioned latent diffusion model for histopathology. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 5182–5191 (2024)