

# 通过基于生成 AI 的图像合成促进针对基于 AI 的皮肤 病变分类器的公平性评估

Ko Watanabe, Stanislav Frolov, Adriano Lucieri, and Andreas Dengel

German Research Center of Artificial Intelligence (DFKI)  
ko.watanabe@dfki.de

**Abstract.** 深度学习的最新进展及其在边缘应用上的潜力，为像黑色素瘤这样的皮肤癌常规筛查的革命带来了巨大潜力。尽管这项技术预期会带来诸多好处，但不可预见和固有的偏见也带来了潜在的风险。因此，评估和提高此类系统的公平性至关重要。公平性评估中的一个关键挑战是确保评估数据集能充分代表不同的 Personal Identifiable Information (PII)（性别、年龄和种族）及其他少数群体。在这一挑战的背景下，本研究利用最先进的 Generative AI (GenAI) LightningDiT 模型评估公开可用的黑色素瘤分类器的公平性。结果表明，使用高度逼真的合成数据进行公平性评估是一个有前途的方向。然而，我们的研究表明，当用于评估的黑色素瘤检测模型是在与合成图像基础数据集不同的数据上训练时，验证公平性变得困难。尽管如此，我们提出，我们的方法为采用合成数据来衡量和增强医学成像 Artificial Intelligence (AI) 系统的公平性提供了一条有价值的新途径。

## 1 引言

黑色素瘤是所有皮肤癌中最致命的一种。2020 年，全球共报告了 325,000 个新病例（其中 174,000 名男性和 151,000 名女性）以及 57,000 例死亡（其中 32,000 名男性和 25,000 名女性） [3]。此外，预计到 2040 年，全球新的黑色素瘤病例数将达到 510,000（约增加 50 %），死亡人数估计将达到 96,000（增加 68 %）。然而，通过早期检测可以显著提高生存机会。例如，在德国的石勒苏益格-荷尔斯泰因州，进行了一项针对 360,288 名 20 岁及以上人群的皮肤癌筛查项目，与 2008 年和 2009 年相比，男性黑色素瘤的死亡人数减少了 47 %，女性减少了 49 % [21]。同样在澳大利亚，共有 2,452 名患者参与了一项筛查项目，黑色素瘤的死亡人数减少了 59 % [40]。这些事实表明，尽管预计黑色素瘤病例数量将增加，早期检测可以显著降低死亡风险。由于 AI 的最新发展以及智能手机和其他影像设备等技术的普及，日常生活中的自动化医疗诊断得到了显著促进。在临床常规筛查中，黑色素瘤的检测可以通过基于 AI 的决策支持系统轻松支持，这些系统能够以人类可理解的语言解释其决策 [24]。理论上，这项技术甚至可以在世界任何地方远程应用。基于智能手机的应用程序在临床数据上操作，可以使患者进行手动预筛查，而无需立即咨询医疗专业人员 [8, 31]。这些创新具有巨大的社会影响，不仅通过促进早期检测、缩短治疗时间、降低死亡风险，还可以潜在地减少医生的工作量、诊断错误率和医疗成本。另一方面，AI 的可信度正变得越来越重要，特别是随着欧盟人工智能法案 [12] 及相关立法的出现。在社会广泛应用的医疗领域中，实现 AI 预期利益的必要条件是这些系统的公平性 [1, 2, 5]。对 AI 系统公平性的适当评估需要以基础扎实的 PII 标注

数据为基础。在皮肤病变数据集领域, International Skin Imaging Collaboration (ISIC)<sup>1</sup> 通过创建一个高质量的皮肤病变图像及元数据合集 [15], 做出了重大贡献。ISIC 数据集 [9, 10, 15, 17, 33, 38] 包括患者诊断图像以及类似性别、年龄和种族 (Fitzpatrick 皮肤类型 [36]) 的信息。虽然 ISIC 做出了巨大贡献, 但数据集的不平衡仍然是一个问题。为不同属性拥有相同数量的评估图像仍然是公平性评估的一个挑战。最近在 GenAI 开发上的飞跃, 导致在图像合成领域取得了巨大进展。扩散模型允许基于提示轻松生成高保真图像。在本文中, 我们研究了 State-of-the-art (SOTA) 生成模型的潜力, 利用算法公平性评估来检测黑色素瘤。使用 ISIC 数据集和 SOTA LightningDiT 方法 [41], 我们训练了一个基础模型, 用于真实合成高保真皮损图像, 并以相关的 PII 属性为条件。利用这些生成的图像, 我们评估了面向公开可用的皮损分类器的公平性。具体而言, 我们提出了两个研究问题 (R1–R2):

- R1 我们能否使用最新的生成图像合成方法来获得一个平衡的公平性评估数据集? 我们开发了一个协议, 使用基于扩散的图像合成来生成平衡的皮肤镜图像群体, 包括性别、年龄和 Fitzpatrick 皮肤类型。
- R2 这个合成数据集能否可靠地评估皮肤病变分类器的公平性? 我们将合成的图像应用于三个经过同行评审的训练好的皮肤病变分类模型。我们的分析量化模型的 Demographic Parity (DP)。

在本节中, 我们介绍了关于合成数据生成、黑色素瘤检测模型和皮肤病变分类中的公平性评估的相关工作。

### 1.1 合成数据生成

合成数据生成已经成为解决皮肤病变数据集 (如 ISIC) 中的数据稀缺性、不平衡和隐私问题的重要方法。早期的工作 [4] 探索了使用生成对抗网络 (GAN) 架构进行高分辨率皮肤病变合成, 包括 DCGAN [30] 和 LAPGAN [11], 即使在小型训练集上也展示了有前景的视觉保真度。DermGAN [14] 改编了 Pix2Pix [18] 以生成具有可控制病变大小、位置和肤色的临床风格皮肤图像, 从而在用于增强时提高了分类器在罕见恶性病症上的性能。Bissoto 等人 [6] 提出了一种基于风格的 GAN [20], 用于合成具有临床细节的皮肤镜像, 从而将分类准确率提高了几个百分点。StyleGAN2ADA [19] 被用于合成有条件 and 无条件图像, 进行了详细的保真度和多样性评估, 以及用于跨机构聚合数据的联邦学习场景 [7]。

除了生成对抗网络 (GANs), 基于扩散的生成模型最近已成为视觉生成模型的主流方法。稳定扩散模型 (Stable Diffusion) [32] 与 DreamBooth [34] 配合使用, 如在 Cancer-Net SCa-Synth [37] 中所示, 能够使用少量样本数据生成均衡的良性/黑色素瘤图像, 从而在微调后的实际数据上提高分类性能。同样, 潜在扩散方法的改进突出了控制病变特征和增强多样性的潜力, 特别是在不同肤色类型之间, 从而解决了病变分类中的公平性差距 [39]。

与之前的所有工作不同, 我们基于 LightningDiT [41] 训练了一个 SOTA 潜在扩散变换器 (DiT), 作为我们公平性评估的生成主干, 利用文本描述和所有可用的 ISIC 数据。

<sup>1</sup> <https://www.isic-archive.com/>

**Table 1.** 现有黑色素瘤检测模型的比较。“预训练模型”列（✓：可用，--：不可用）指示模型的可用性。

| Reference               | Architecture    | Dataset   | Pretrained Model | Performance      |
|-------------------------|-----------------|-----------|------------------|------------------|
| DeepGuide [26]          | DenseNet        | Derm7pt   | ✓                | 0.788 (AUC)      |
| MelaNet [42]            | VGG-GAP         | ISIC 2016 | ✓                | 0.811 (AUROC)    |
| Patch-Lesion [13]       | Dense121        | HAM10000  | –                | 0.855 (F1-Score) |
| SkinLesionDensenet [28] | DenseNet201     | ISIC      | ✓                | 0.878 (F1-Score) |
| EFFNet [25]             | EfficientNetV2  | HAM10000  | –                | 0.932 (F1-Score) |
| Sabir et al. [35]       | EfficientNet-B0 | HAM10000  | –                | 0.970 (F1-Score) |

## 1.2 黑色素瘤检测模型

Table 1 展示了现有黑色素瘤检测模型的比较。在研究黑色素瘤检测模型的对比时，各种深度学习模型已针对不同的数据集进行了验证。除了 ISIC 数据集，之前的研究还利用了 Derm7pt [22] 和 HAM10000 (HAM) [38] 数据集进行黑色素瘤检测模型的训练和验证。HAM 包含七种皮肤病变类型的 10,015 张皮肤病变图像。

在检查模型性能时，我们发现它们在不同数据集上的表现并不一致。一些使用 F1 分数来衡量模型的性能，而另一些则使用 AUC 或 AUROC。因此，直接对模型进行比较并不简单。

我们还验证了预训练模型的可用性存在差异。例如，DeepGuide [26]、MelaNet [42] 和 SkinLesionDensenet [28] 对公众可用，但 Patch-Lesion [13] 和 EFFNet [25] 对公众不可用。在这项研究中，我们使用 DeepGuide [26]、MelaNet [42] 和 SkinLesionDensenet [28] 的预训练模型进行公平性评估。

在皮肤病变分类的公平性评估中已经进行了多种方法的研究。Ansari 等人提出了一种新颖的数据增强方法，以缓解类别不平衡和肤色偏见。该方法利用了一种通过结合两张不同图像生成合成图像的方法。其中一张是选定的肤色图像，另一张是诊断图像。而 BiaslessNAS 通过在 NAS 过程中每个阶段中加入公平性考虑来改善公平性。该方法可以识别出更加准确和公平的神经网络。与传统方法相比，准确性提高了 2.55 %，公平性提高了 65.50 %。

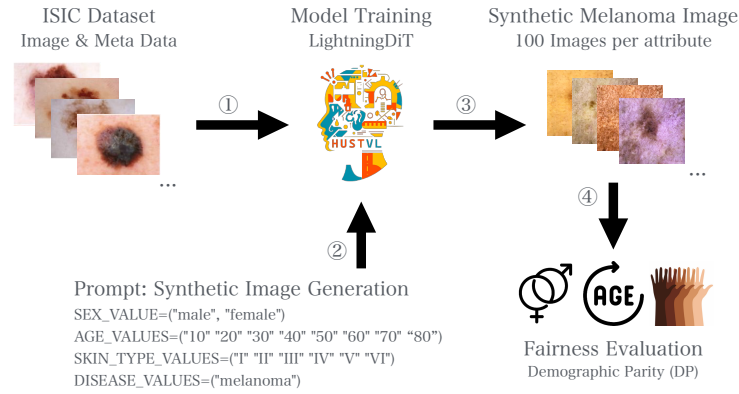
与现有的数据增强方法相比，我们的方法可以在提示基础上生成具有各种 PII 属性的合成图像。我们还注重评估对现有预训练模型 [26, 28, 42] 的公平性，这将使我们能够理解基于提示的合成数据生成在测试公平性方面的潜力。

## 2 方法论

Figure 1 展示了我们的公平性测试工作流程。在本节中，我们首先描述生成合成测试数据的过程。接下来，我们描述针对所选的预训练模型进行黑色素瘤检测公平性评估的过程。

要生成合成测试数据，我们首先需要生成模型。为此，我们首先训练一个 SOTA 潜在扩散变换器，使用最近的 LightningDiT [41] 主干<sup>2</sup>。LightningDiT 通过利用预训练的视觉基础模型来调整底层变分自编码器 (VAE)，提高了图像重建和生成性能。具体来说，与视觉基础模型对齐的 VAE (VA-VAE) 确保图像的潜在表示与预训练的视觉基础模型对齐，例如 DINOv2 [27] 和 MAE [16]。

<sup>2</sup> <https://github.com/hustvl/LightningDiT>



**Fig. 1.** 我们公平性测试工作流程的整体管道。该工作流程展示了针对 LightningDiT [41] 微调 ISIC 数据集的过程。使用系统提示生成的合成黑色素瘤测试图像。使用生成的图像，对预训练的黑色素瘤检测模型进行公平性评估（DP）。

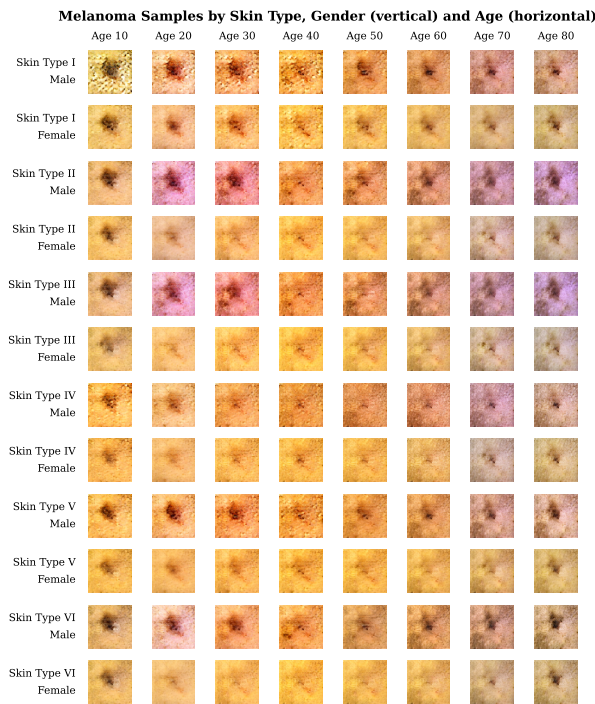
我们首先使用 VA-VAE 在所有切分（2016 年、2017 年和 2018-2024 年）上提取真实 ISIC 图像的潜表示，这些图像总共大约有 500,000 张。我们使用 256 的图像分辨率，并通过现有的注释为每张图像配对相应的文本描述。具体来说，我们将对应于性别、年龄、大小、肤色和诊断的标签串联到一个文本描述中。我们使用 CLIP [29] 产生文本嵌入。接下来，我们用 675M 的参数和标准 LightningDiT 设置训练潜扩散变压器，进行 80,000 步，使用 8 个 A100 GPU，以全球批量大小 1024 和混合精度，采用 AdamW [23] 和 0.0002 的学习率。这个过程大约需要 8 小时。

最后，我们可以通过对随机噪声进行采样并将其基于文本提示进行条件化，利用经过训练的模型生成合成测试数据，使用目标标签进行公平性评估。我们使用 250 步的欧拉求解器的标准设置和无分类器引导比例为 10 来采样图像。

如 Figure 1 所示，我们通过提示生成合成图像。为了公平测试，我们为每个 PII 属性生成 100 张图像。模式被分为性别（2 种）、年龄（8 种）、皮肤类型（7 种）和疾病病例（1 种）。总共，我们为测试生成了 11,200 个（ $2 \times 8 \times 7 \times 1 \times 100$ ）图像。

为了评估预训练模型的公平性，我们选择人口差异（DP）。DP 也称为 Accuracy Parity (AP)，是不同子群体之间准确性差异。DP 特别关注不同人口（PII）群体之间的 AP。使用平衡的 PII 考虑的合成测试图像，我们针对预训练模型评估 DP。

在本研究中，我们选择了三个预训练模型，分别是 DeepGuide [26]、MelaNet [42] 和 SkinLesionDensenet [28]。如前所述，我们选择这些模型是因为有预训练的黑色素瘤检测模型可用。这三个模型之间最显著的区别在于 MelaNet 和 SkinLesionDensenet 都在 ISIC 上训练，而 DeepGuide 使用 HAM 数据集进行训练。此外，DeepGuide 和 SkinLesionDensenet 的模型结构基于 DenseNet，而 MelaNet 基于 VGG-GAP 进行训练。



**Fig. 2.** LightningDiT [41] 生成的合成黑色素瘤图像数据。纵轴显示的是费兹帕特里克皮肤类型 (I–VI) 及其性别 (男性或女性)，而横轴显示的是年龄 (10–80)。该网格展示了跨不同人口群体的合成黑色素瘤图像，以评估皮肤病变分类的公平性。

在本节中，我们使用生成的合成测试数据评估皮肤病变分类器的公平性。我们首先展示该方法生成的图像样本。然后，我们使用合成测试数据评估皮肤病变分类器的公平性能。

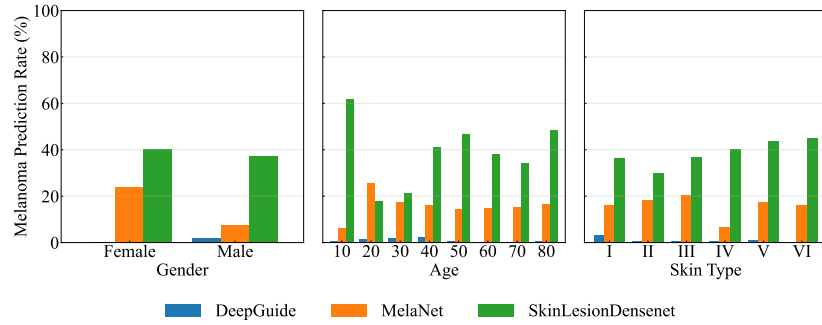
### 2.1 生成的合成皮肤病变图像

Figure 2 展示了由 LightningDiT 生成的图像样本。纵轴表示皮肤类型及性别差异，横轴表示年龄变化。生成的合成图像之间的差异是明显的。然而，一些图像，例如男性、年龄 10 岁、皮肤类型 I 的图像，似乎不真实。这可能是由于模型对真实训练数据集的接触有限。生成图像的性能评估是另一个重要主题，将在 ?? 中讨论。

### 2.2 公平性表现评估

Figure 3 展示了我们皮肤病变分类模型在不同人口群体中公平性评估结果。结果证实，预训练模型将测试图像预测为黑色素瘤。

值得注意的是，DeepGuide 的性能略低于 MelaNet 和 SkinLesion-Densenet。这可能是由于训练数据集的差异造成的。DeepGuide 使用 HAM



**Fig. 3.** 皮肤病变分类器在不同人口统计群体中的表现。该图展示了不同的性别、年龄和 Fitzpatrick 皮肤类型群体。

**Table 2.** 每个模型在不同人口属性（性别、年龄、肤色类型）上的最大（Max）、最小（Min）和人口平等（DP）准确率。更高的 DP 表示模型在相应属性上的表现差距更大。

| Model              | Sex    |        |        | Age    |        |        | Skin Type |        |        |
|--------------------|--------|--------|--------|--------|--------|--------|-----------|--------|--------|
|                    | Max    | Min    | DP     | Max    | Min    | DP     | Max       | Min    | DP     |
| DeepGuide          | 0.0179 | 0.0031 | 0.0148 | 0.0258 | 0.0017 | 0.0241 | 0.0312    | 0.0038 | 0.0274 |
| MelaNet            | 0.2402 | 0.0762 | 0.1640 | 0.2575 | 0.0633 | 0.1942 | 0.2062    | 0.0662 | 0.1400 |
| SkinLesionDensenet | 0.4040 | 0.3708 | 0.0332 | 0.6200 | 0.1792 | 0.4408 | 0.4494    | 0.3012 | 0.1482 |

数据集进行训练，而 MelaNet 和 SkinLesionDensenet 使用 ISIC 数据集进行训练。我们的合成测试数据生成基于 ISIC 数据集，这可能导致 DeepGuide 的性能较低。不仅要评估不同 PII 组的 DP 公平性，还要评估不同数据集的公平性。结果表明，合成测试数据生成可用于评估训练数据集的公平性。通过使用各种训练数据集创建不同的合成数据生成模型，可以评估预训练模型处理中数据集的公平性。

Table 2 表示 DP 的实际最大值。结果表明，在 SkinLesionDensenet 模型中，最大值的 DP 是年龄在 10 到 20 岁之间的范围。我们确认 DeepGuide 针对性别的最低 DP 表现，但这通常是由于整体准确性较低。我们验证了合成数据生成可以作为工具来估计同一训练数据集中 PII 群体的公平性，这适用于预训练模型和合成数据生成模型。然而，如果模型是在不同的数据集上训练的，对于在不同数据集上训练的预训练模型，合成数据可能会导致性能略有下降。因此，合成测试数据可以首先验证预训练模型的鲁棒性，然后评估其在不同的 PII 群体之间的公平水平。

在本文中，我们验证了 section 1 中介绍的两个研究问题（R1–R2）。我们验证了这项研究中的几个要点。对于 R1，使用 LightningDiT 生成的合成图像测试数据表现良好。我们使用 ISIC 作为训练数据集，并将其应用于其他图像，如 HAM 和 Derm7pt。因此，改进合成图像生成模型的下一步是考虑不同的数据集或用于捕捉图像的不同硬件。通过在皮肤病学中生成合成图像，这种方法能实现公平性和鲁棒性测试。对于 R2，我们验证了合成图像测试数据生成可以作为评估预训练模型公平性的工具。例如，检查像 SkinLesionDensenet 这样的预训

练模型的性能，确认了黑色素瘤的预测。下一步是应用基于概念的公平性评估，以理解公平性问题的实际原因。

在本研究中，我们使用 LightningDiT 生成合成的黑色素瘤测试图像。我们验证了合成图像可以作为测试数据用于评估预训练模型的公平性。我们还发现，当分类器和生成器在不相交的数据集上训练时，性能会适度下降，这突显了数据集偏移的影响。这意味着，严格的隐私保护、无 PII 的公平性审查在生成器和检测器来自相同数据分布时最为可靠。除了公平性之外，合成测试集被证明对于压力测试模型的鲁棒性很有用，这表明数据驱动的生成可以作为医学影像中 AI 的通用诊断工具。

## Bibliography

- [1] Aayushman, Gaddey, H., Mittal, V., Chawla, M., Gupta, G.R.: Fair and accurate skin disease image classification by alignment with clinical labels. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 394–404. Springer (2024)
- [2] Ansari, F., Chakraborti, T., Das, S.: Algorithmic fairness in lesion classification by mitigating class imbalance and skin tone bias. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 373–382. Springer (2024)
- [3] Arnold, M., Singh, D., Laversanne, M., Vignat, J., Vaccarella, S., Meheus, F., Cust, A.E., De Vries, E., Whiteman, D.C., Bray, F.: Global burden of cutaneous melanoma in 2020 and projections to 2040. *JAMA dermatology* **158**(5), 495–503 (2022)
- [4] Baur, C., Albarqouni, S., Navab, N.: Melanogans: High resolution skin lesion synthesis with gans. arXiv preprint arXiv:1804.04338 (2018)
- [5] Bevan, P.J., Atapour-Abarghouei, A.: Detecting melanoma fairly: Skin tone detection and debiasing for skin lesion classification. In: MICCAI Workshop on Domain Adaptation and Representation Transfer. pp. 1–11. Springer (2022)
- [6] Bissoto, A., Perez, F., Valle, E., Avila, S.: Skin lesion synthesis with generative adversarial networks. arXiv preprint arXiv:1902.03253 (2019)
- [7] Carrasco Limeros, S., Majchrowska, S., Zoubi, M.K., Rosén, A., Suvilehto, J., Sjöblom, L., Kjellberg, M.: Assessing gan-based generative modeling on skin lesions images. In: Machine Intelligence and Digital Interaction Conference. pp. 93–102. Springer Nature Switzerland Cham (2022)
- [8] Chao, E., Meenan, C.K., Ferris, L.K.: Smartphone-based applications for skin monitoring and melanoma detection. *Dermatologic clinics* **35**(4), 551–557 (2017)
- [9] Codella, N., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., et al.: Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1902.03368 (2019)
- [10] Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., 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). In: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). pp. 168–172. IEEE (2018)
- [11] Denton, E.L., Chintala, S., Fergus, R., et al.: Deep generative image models using a laplacian pyramid of adversarial networks. *Advances in neural information processing systems* **28** (2015)

- [12] European Union High-Level Expert Group on AI: Ethics guidelines for trustworthy ai (2019), <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- [13] Gessert, N., Sentker, T., Madesta, F., Schmitz, R., Knip, H., Baltruschat, I., Werner, R., Schlaef, A.: Skin lesion classification using cnns with patch-based attention and diagnosis-guided loss weighting. *IEEE Transactions on Biomedical Engineering* **67**(2), 495–503 (2020). <https://doi.org/10.1109/TBME.2019.2915839>
- [14] Ghorbani, A., Natarajan, V., Coz, D., Liu, Y.: Dermgan: Synthetic generation of clinical skin images with pathology. *arXiv preprint arXiv:1911.08716* (2019)
- [15] Gutman, D., Codella, N.C., Celebi, E., Helba, B., Marchetti, M., Mishra, N., Halpern, A.: Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1605.01397* (2016)
- [16] He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 16000–16009 (2022)
- [17] Hernández-Pérez, C., Combalia, M., Podlipnik, S., Codella, N.C., Rotemberg, V., Halpern, A.C., Reiter, O., Carrera, C., Barreiro, A., Helba, B., et al.: Bcn20000: Dermoscopic lesions in the wild. *Scientific data* **11**(1), 641 (2024)
- [18] Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1125–1134 (2017)
- [19] Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., Aila, T.: Training generative adversarial networks with limited data. *Advances in neural information processing systems* **33**, 12104–12114 (2020)
- [20] Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 4401–4410 (2019)
- [21] Katalinic, A., Waldmann, A., Weinstock, M.A., Geller, A.C., Eisemann, N., Greinert, R., Volkmer, B., Breitbart, E.: Does skin cancer screening save lives? an observational study comparing trends in melanoma mortality in regions with and without screening. *Cancer* **118**(21), 5395–5402 (2012)
- [22] 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)
- [23] Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. *International Conference on Learning Representations (ICLR)* (2019), <https://arxiv.org/abs/1711.05101>
- [24] Lucieri, A., Bajwa, M.N., Braun, S.A., Malik, M.I., Dengel, A., Ahmed, S.: Exaid: A multimodal explanation framework for computer-aided diagnosis

- sis of skin lesions. *Computer Methods and Programs in Biomedicine* **215**, 106620 (2022)
- [25] Ma, X., Shan, J., Ning, F., Li, W., Li, H.: Effnet: A skin cancer classification model based on feature fusion and random forests. *Plos one* **18**(10), e0293266 (2023)
  - [26] Mallya, M., Hamarneh, G.: Deep multimodal guidance for medical image classification. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 298–308. Springer (2022)
  - [27] Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khaidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: DINOv2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193* (2023)
  - [28] Panda, S., Tiwari, A.S., Prusty, M.R.: Comparative study on different deep learning models for skin lesion classification using transfer learning approach. *International Journal of Scientific and Research Publications* **11**(1), 219–32 (2021)
  - [29] Radford, A., Kim, J.W., Hallacy, L., Ramesh, A., Goh, G., Agarwal, S., Sasstry, G., Aspell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision. In: *Proceedings of the 38th International Conference on Machine Learning (ICML)* (2021)
  - [30] Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434* (2015)
  - [31] Rat, C., Hild, S., Rault Sérandour, J., Gaultier, A., Quereux, G., Dreno, B., Nguyen, J.M.: Use of smartphones for early detection of melanoma: systematic review. *Journal of medical Internet research* **20**(4), e135 (2018)
  - [32] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 10684–10695 (2022)
  - [33] Rotemberg, V., Kurtansky, N., Betz-Stablein, B., Caffery, L., Chousakos, E., Codella, N., Combalia, M., Dusza, S., Guitera, P., Gutman, D., et al.: A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Scientific data* **8**(1), 34 (2021)
  - [34] Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., Aberman, K.: Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 22500–22510 (2023)
  - [35] Sabir, R., Mehmood, T.: Classification of melanoma skin cancer based on image data set using different neural networks. *Scientific Reports* **14**(1), 29704 (2024)
  - [36] Sachdeva, S.: Fitzpatrick skin typing: Applications in dermatology. *Indian journal of dermatology, venereology and leprology* **75**, 93 (2009)
  - [37] Tai, C.e.A., Ding, O., Wong, A.: Cancer-net sca-synth: An open access synthetically generated 2d skin lesion dataset for skin cancer classification. *arXiv preprint arXiv:2411.05269* (2024)

- [38] 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)
- [39] Wang, J., Chung, Y., Ding, Z., Hamm, J.: From majority to minority: A diffusion-based augmentation for underrepresented groups in skin lesion analysis. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 14–23. Springer (2024)
- [40] Watts, C.G., McLoughlin, K., Goumas, C., Van Kemenade, C.H., Aitken, J.F., Soyer, H.P., Peñas, P.F., Guitera, P., Scolyer, R.A., Morton, R.L., et al.: Association between melanoma detected during routine skin checks and mortality. *JAMA dermatology* **157**(12), 1425–1436 (2021)
- [41] Yao, J., Yang, B., Wang, X.: Reconstruction vs. generation: Taming optimization dilemma in latent diffusion models. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 15703–15712 (2025)
- [42] Zunair, H., Hamza, A.B.: Melanoma detection using adversarial training and deep transfer learning. *Physics in Medicine & Biology* (2020)