International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 13 Issue 12 ǁ December 2025 ǁ PP. 145-153

Detection vs Design: Comparing AI-Generated Content in Dermatology

K Deepa Shree, Bhavani S Bedre, Adithya P S, A R Charmee Choudhary, Akshata S Joshi
Dept. of Computer Science and Engineering
Dayananda Sagar Academy of technology and Management, Bangalore, India

Abstract— *AI has become essential in dermatology, where deep learning models like CNNs and Vision Transformers achieve strong performance in classifying skin diseases using large, well-established datasets. At the same time, generative models such as GANs, StyleGAN2, and Diffusion Models can now create highly realistic synthetic skin images, raising concerns about detection reliability and medical image authenticity. Comparing these discriminative and generative approaches is therefore important to understand their strengths, limitations, and impact on clinical settings. Developing a multi-task learning detector—capable of predicting disease type, skin tone, and severity—is relevant because it reflects real diagnostic needs and helps evaluate robustness across diverse conditions. The observed performance drop when the detector is exposed to high-fidelity synthetic images is consistent with recent findings that diffusion models can closely mimic real dermatology images, challenging existing detection systems. Finally, the emphasis on ethical standards, fairness, and forensic methods is warranted, as synthetic medical data can influence diagnostic accuracy, bias, and trust in AI-assisted healthcare.*

Keywords— *Artificial Intelligence and Deep Learning capture the overall methodological foundation of the work, while Generative Adversarial Networks (GANs) and Diffusion Models represent the generative techniques responsible for producing synthetic dermatology images. Vision Transformers relate to modern discriminative models used for lesion classification. Synthetic Data is central to the study's focus on evaluating detectors against AI-generated images. Dermatology identifies the specific medical domain of application. Detection Algorithms and Digital Forensics are essential because the research investigates how well AI systems can distinguish real skin images from generated ones and highlights the need for reliable forensic methods. Together, these keywords accurately reflect the paper's technical focus, medical relevance, and forensic implications.*

Date of Submission: 07-12-2025 Date of acceptance: 19-12-2025

I. INTRODUCTION

Artificial Intelligence is fundamentally changing dermatology, enabling skin analysis that is faster, automated, and more precise than ever before. Deep learning models can now detect skin cancers and other conditions with an accuracy that rivals expert dermatologists [1], [2]. At the same time, a new wave of "generative AI" has unlocked the ability to create highly realistic synthetic skin images, which are valuable for research, training, and cosmetic simulation.

However, this powerful advancement comes with a significant catch—a dual-use dilemma. The very technology that can enhance diagnostic training by generating synthetic data can also be misused to create falsified or manipulated medical images [3], [9]..This has sparked what we call a "detection versus design arms race," where AI systems are in a constant competition: one side strives to create ever-more realistic synthetic content, while the other works to detect it with greater precision[6], [10].

.This paper directly explores this "arms race" by comparing two sides of AI AI work in dermatology mainly falls into two categories: systems that detect and classify skin conditions, and systems that generate or enhance skin images. These two areas form the basis of the comparison discussed in this paper.

We examine their core architectures, the data they use, how they are evaluated, and the crucial ethical implications of their use.

The key contributions of this work include providing a clear comparison between AI models that generate skin images and those that detect diseases, analyzing the performance trade-offs between these two approaches, and highlighting the ethical and security concerns that arise from the use of synthetic medical data.

II. RELATED WORK

Progress in automated dermatology has been fueled by publicly available, well-labeled datasets and standardized community benchmarks. The HAM10000 collection is a cornerstone in this field—a widely-used, multi-source dataset containing over 10,000 labeled images of common pigmented skin lesions[1]. Researchers use it both to train diagnostic models and to analyze fairness across different patient subgroups.

Similarly, the ISIC archive and its annual challenges have been instrumental. They provide standardized tasks—like segmenting lesions, detecting attributes, and classifying diseases [2]—along with public leaderboards. These resources act as essential baselines, allowing for fair and reproducible comparisons between different AI models.

Platforms like Kaggle also play a key role, hosting mirrors and aggregated versions of these datasets (such as ISIC-2019 and HAM10000 variants), which enable researchers to rapidly experiment and build upon established baselines [1], [2].

The evolution of AI detection systems in dermatology has seen remarkable progress. Early approaches relied on Convolutional Neural Networks (CNNs) like VGG and ResNet, which achieved groundbreaking, dermatologist-level accuracy in classifying skin lesions [1], [2].

This was followed by more efficient architectures like EfficientNet, which delivered better performance with fewer parameters, making them practical for real-world clinical systems. Most recently, Vision Transformers (ViTs) and hybrid CNN-Transformer models have emerged. These models can capture broader contextual information in images and have demonstrated performance that matches or even surpasses traditional CNNs on several benchmarks.

This rapid architectural evolution makes it crucial to systematically compare different detector types—from classic CNNs to modern Transformers—to understand their strengths and weaknesses in both standard classification and robustness testing [10].

On the other side of the arms race, the technology for *creating* synthetic dermatology images has advanced at an incredible pace.

The journey began with GANs (Generative Adversarial Networks). Early models like DermGAN proved that AI could generate clinically plausible skin lesions, with control over type, position, and skin tone. In fact, these early fakes were so convincing that clinicians in blinded studies sometimes couldn't tell them apart from real patient images [3].

The next leap came with StyleGAN2, which significantly boosted image resolution and fidelity. This allowed for the creation of highly detailed skin textures and lesions, opening up new possibilities for cosmetic simulation and high-quality data augmentation [4].

Most recently, Diffusion Models have taken center stage. These models, including fine-tuned versions like LesionGen, now produce synthetic images with superior diversity and realism, often outperforming earlier GANs on technical metrics [5], [6]. This rapid progress on the "design" side has dramatically intensified the arms race. Detection systems trained to spot older GAN-based fakes are often completely fooled by these new, high-fidelity outputs from diffusion models.

Synthetic dermatology images are useful in several ways. They help balance datasets by adding examples of rare conditions, increase dataset size without risking patient privacy, and provide realistic visuals for medical training. Research also shows that when high-quality synthetic images are included, AI models often perform better, especially for skin conditions that are not well represented in real clinical data [3], [6].

However, this powerful capability comes with serious risks—a true double-edged sword. The same technology that creates helpful training data can also be misused to fabricate clinical evidence, mislead diagnoses, or inadvertently bias medical AI systems if synthetic artifacts introduce hidden patterns. As synthetic images become more realistic, telling them apart from genuine medical images becomes increasingly difficult. This growing concern has led to urgent calls for ethical guidelines, traceability systems, and detailed metadata tracking to ensure synthetic data is used responsibly in healthcare settings [9].

A well-known and persistent problem in dermatology AI is that most datasets heavily under-represent people with darker skin tones. This leads to AI models that are less accurate for minority populations, creating unfair and potentially dangerous diagnostic disparities [6].

To combat this, researchers have turned to synthetic generation as a potential solution. By specifically guiding AI models to create images of conditions on darker skin, tools like S-SYNTH and other controlled pipelines aim to build more balanced and representative datasets [6].

However, this approach comes with its own challenges. While synthetic augmentation can help reduce performance gaps, it's not a perfect fix. If the AI-generated images don't perfectly match the complex appearance of real clinical cases, they can introduce new problems [6], [9]. Therefore, any fairness intervention using synthetic data must be carefully validated by dermatologists and rigorously tested to ensure it truly improves real-world clinical reliability.

As AI-generated images become more realistic, researchers are fighting back with sophisticated forensic tools designed to spot the tiny traces that distinguish synthetic images from real ones. These detection strategies typically fall into three categories:

Detection methods in dermatology for spotting synthetic images generally fall into three categories. Model-based detectors are trained specifically to identify whether an image is real or AI-generated by analyzing deep visual features [9], [10]. Activation-space forensics focuses on finding subtle statistical traces left during image generation, including artifacts or unusual frequency patterns that are not visible to the human eye [9]. Finally, explainability-centered checks use tools such as Grad-CAM to visualize where a model is focusing its attention [10]; synthetic images often lead to scattered or unusual focus patterns compared to real lesions.

The most robust strategy combines all these forensic clues into a single hybrid detector [9], [10]. However, this is a constant battle—as soon as a new detection method emerges, the generators can be optimized to evade it. This cycle of improvement and countermeasure is the very heart of the ongoing "design vs. detection" arms race.

To fairly compare different AI models, the field relies on standardized metrics and benchmarks.

For detection models, standard performance measures include Accuracy, Precision, Recall, F1-Score, and AUC-ROC. These metrics collectively tell us how reliable a diagnostic AI is [1], [2].

For generative models, the most common quality metric is the Frechet Inception Distance (FID), which quantifies how realistic and diverse the generated images are. A lower FID score means the synthetic images are closer to the real thing [4]–[6].

In medical applications, the ultimate test for synthetic images is often a blinded expert review (a kind of "Turing test" for doctors) and checking whether they actually improve the performance of diagnostic AI systems when used for training [1], [2].

To ensure all comparisons are fair and reproducible, the community uses public benchmarks like the ISIC Challenge and maintains shared repositories for generative models. These resources provide common ground for researchers to measure progress and compare results [2], [4], [5].

H. Summary of Gaps and How the Present Work Differs

Although significant progress has been made in both diagnosing and generating dermatology images, our review highlights three important gaps that remain unaddressed. First, most dermatology detectors are evaluated only on real images and are not rigorously tested against the latest, highly realistic synthetic images produced by modern diffusion models [4]–[6]. Second, while synthetic data is often suggested as a solution to skin-tone bias, many studies do not fully examine the trade-offs involved, which can introduce new challenges. Finally, current forensic techniques for detecting synthetic images are still limited and lack strong, combined strategies capable of identifying the newest and most convincing AI-generated fakes [5], [6].

Our study directly addresses these gaps by testing detectors against a wide range of AI-generated images—from early GAN-based models to the latest diffusion models—allowing us to evaluate their robustness in realistic scenarios. We also examine how detection performance changes as the proportion of synthetic data increases, providing a clearer understanding of its impact on reliability. In addition, we conduct thorough fairness evaluations across all skin tones to ensure that any improvements benefit all groups equally. The following section describes the multi-task framework and experiments we developed to meet these objectives [9], [10].

III. METHODOLOGY

The core goal of this study is to understand the dynamic interplay between AI systems that detect skin conditions and those that generate synthetic skin images. To achieve this, we developed a versatile multi-task learning (MTL) dermatology detector and rigorously tested its resilience against a wide spectrum of synthetic images—from those created by earlier GANs to the most advanced outputs from StyleGAN2 and Diffusion Models.

Our multi-task learning (MTL) detector is built to perform three important clinical tasks simultaneously: it classifies the skin condition, predicts the Fitzpatrick skin tone, and estimates the severity of the lesion.

A single, efficient backbone network (EfficientNetV2-B0) serves as the foundation, extracting features that are shared across all tasks. We then systematically challenged this detector with synthetic datasets to measure its robustness against progressively more realistic AI-generated imagery.

For the real dermatology datasets, we trained and evaluated our detection models using two well-established public sources. The first is HAM10000, a widely used dermatoscopic image collection that includes 10,015 images representing seven common types of pigmented skin lesions. Its diversity—coming from multiple sources—makes it a strong benchmark for classification research. The second dataset is ISIC 2019, a large-scale challenge dataset containing over 25,000 labeled dermoscopic images. It offers a standardized training and evaluation protocol with reliable ground-truth diagnoses, making it an essential resource for building and testing dermatology AI systems.

To ensure our results were robust and prevent data leakage, we split the data by patient. The real datasets were divided into 70% for training, 15% for validation, and 15% for testing.

To thoroughly evaluate how well our system handles synthetic images, we used data generated from three major families of AI models. The first was DermGAN, which creates clinical dermatology images based on specific lesion features. The second was StyleGAN2-ADA, a high-fidelity generator known for producing images with detailed and realistic skin textures. Finally, we included fine-tuned diffusion models such as Stable Diffusion and LesionGen, which represent the current state of the art in generating highly realistic and diverse dermatological imagery.

We generated these synthetic images in a class-balanced manner. To systematically measure their impact, we created mixed datasets with varying proportions of synthetic content: 0%, 25%, 50%, 75%, and 100% synthetic images. This allowed us to precisely evaluate how increasing exposure to AI-generated data affects model robustness.

To maintain consistency across the dataset, all images were processed through a standardized preparation pipeline. Each image was first resized to 224×224 pixels and then normalized using standard ImageNet statistics to stabilize training, following the formula

$$I_{norm} = \frac{I - [0.485,0.456,0.406]}{[0.229,0.224,0.225]}$$

During training, we applied several data augmentation techniques to improve the model's ability to generalize. These included random rotations of up to ±15°, random cropping of 90–100% of the original image area, horizontal flipping, brightness and contrast adjustments, and the addition of slight Gaussian noise. Together, these steps helped the model become more robust to common variations in real-world dermatology images.

This augmentation process is crucial as it helps the model generalize better to new, unseen data and prevents it from overfitting to specific skin textures or lighting conditions found in the training set.

We selected EfficientNetV2-B0 as the core of our model because it offers an excellent balance of computational efficiency and high accuracy. This network processes the input image and extracts its fundamental features.

A final step, called Global Average Pooling (GAP), then condenses these features into a compact, 1280-dimensional vector. This vector serves as a shared knowledge base for all the subsequent tasks:

$$F_{shared} = GAP(CNN_{\theta}(I_{norm}))$$

This component is responsible for diagnosing the skin condition. It uses a simple neural network structure that takes the shared features and produces a classification across ten different dermatological conditions. The network employs dropout and batch normalization to ensure stable training and prevent overfitting. The model is trained using standard cross-entropy loss, which effectively measures how well the predicted diagnoses match the actual conditions.

Loss:

$$L_{cond} = -\sum_{i=1}^{N} \sum_{c=1}^{10} y_{i,c} \log(\hat{y}_{i,c})$$

This part of the model classifies the skin tone into one of the six Fitzpatrick types. It uses a smaller neural network structure with appropriate regularization to ensure reliable predictions. The model is trained using cross-entropy loss to accurately match the predicted skin tones with their true classifications.

Loss:

$$L_{tone} = -\sum_{i=1}^{N} \sum_{t=1}^{6} y_{i,t} \log(\hat{y}_{i,t})$$

This component estimates the severity of the condition on a continuous scale from 0 to 2. It uses a regression-based approach with mean squared error loss to predict accurate severity scores. The architecture includes appropriate regularization to maintain prediction stability across different cases.

Loss:

$$L_{sev} = \frac{1}{N}\sum_{i=1}^{N} \quad (y_i - \hat{y}_i)^2$$

The total loss combines all task losses with empirical weights:

$$L_{total} = \alpha L_{cond} + \beta L_{tone} + \gamma L_{sev}$$

The weights for the combined loss function were determined empirically, assigning α = 1.0, β = 0.5, and γ = 0.8. This weighting scheme prioritizes diagnostic accuracy as the main objective while still maintaining balanced learning for skin-tone classification and severity estimation.

For training, we used the AdamW optimizer with a weight decay of 0.01 and applied a cosine learning rate schedule that gradually decreased the rate from 1×10−41 \times 10^{-4}1×10−4 to 1×10−61 \times 10^{-6}1×10−6. The model was trained with a batch size of 32 for approximately 80 to 120 epochs, and training was stopped early based on validation AUC performance. This setup provided a good balance between efficiency and accuracy, allowing smooth convergence while reducing the risk of overfitting.

Three generative pipelines were evaluated:
For the GAN-based generation pipeline, we used DermGAN, which creates synthetic dermatology images by conditioning on lesion boundaries, lesion class, and skin tone. It is trained using an adversarial setup with a PatchGAN discriminator and produces 256×256 synthetic dermoscopic images.

The second generative pipeline used in our study was StyleGAN2-ADA. This model incorporates adaptive discriminator augmentation to improve training stability and is capable of generating high-resolution images up to 1024×1024 pixels. Its ability to produce fine-grained details makes it especially effective for simulating realistic skin textures, which is valuable in cosmetic dermatology applications.

The third generative pipeline consisted of diffusion-based models, including fine-tuned versions of Stable Diffusion and LesionGen. These models support text- and attribute-conditioned sampling, allowing more controlled and flexible image generation. They achieved greater diversity and lower FID scores compared to GAN-based methods, and produced highly realistic features such as pigmentation patterns, lesion borders, and smooth texture transitions. The quality of the generated images was evaluated using the Fréchet Inception Distance (FID) metric and, when possible, reviewed by dermatologists for additional validation.

For classification and detection, we evaluated the model using standard metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. These metrics were calculated on both the real-only test sets and the evaluation sets that included synthetic images, allowing us to assess performance across different data conditions.

For the regression task, we evaluated performance using three standard metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the R^2 coefficient. These measures helped quantify how accurately the model predicted continuous severity scores.

For evaluating generative quality, we used the Fréchet Inception Distance (FID), which measures how closely the synthetic images resemble real ones. Lower FID scores indicate higher realism and better diversity in the generated outputs.

To ensure our model performed equitably across all skin tones, we evaluated its performance separately for each Fitzpatrick skin type (I-VI). We measured fairness using a performance disparity metric, defined as the maximum accuracy difference between any two skin types:
Disparity = max $|Accuracy_{t1} - Accuracy_{t2}|$

We enforced a strict target disparity of less than 5%. Additionally, we analyzed whether using synthetic images from GANs or diffusion models helped improve fairness for underrepresented skin tones in our dataset.

Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to visualize critical regions influencing model decisions. Feature-map importances were computed as:

$$\alpha_c^k = \frac{1}{Z}\sum_{i} \quad \sum_{j} \quad \frac{\partial y^c}{\partial A_{ij}^k}$$

The final saliency map:

$$GradCAM = ReLU(\sum_{k} \alpha_c^k A^k)$$

We used Grad-CAM visualizations to understand where our model was focusing when making decisions. By comparing attention patterns between real and synthetic images, we could identify unusual focus areas that might reveal the presence of synthetic artifacts.

To evaluate "design vs detection" resilience:
To assess the system's resilience in the "design versus detection" setting, we evaluated each detector under different types of test data. Performance was measured on real-only test sets to establish a baseline, on synthetic-only sets to understand pure detection capability, and on mixed sets containing varying proportions of synthetic images (25%, 50%, and 75%). This setup allowed us to analyze how detection reliability changes as the influence of increasingly sophisticated generative models grows.

We also performed generator-specific stress testing by evaluating the detectors separately on images created by GANs, StyleGAN2, and diffusion models. This targeted analysis allowed us to determine which type of synthetic imagery posed the greatest challenge for the detection systems and where the models were most likely to struggle.

We carried out a series of ablation studies to validate our design choices. These experiments examined how different proportions of synthetic data influence performance, compared alternative backbone architectures such as EfficientNetV2, ResNet-50, and ViT-B/16, and explored the impact of adjusting the loss weights (α,β,γ)(\alpha, \beta, \gamma)(α,β,γ) in the multi-task learning setup. We also evaluated how data augmentation affects fairness across skin tones and compared the usefulness of GAN, StyleGAN2, and diffusion-based synthetic images for augmentation. Each study was assessed using AUC, F1-score, fairness disparity, and FID to ensure a thorough and reliable analysis.

IV. RESULTS AND DISCUSSION

In this section, we present the performance of our multi-task dermatology model across its three core functions: classifying skin conditions, predicting skin tone, and assessing severity. We then test the model's robustness against synthetic images from GAN [3], StyleGAN2 [4], and diffusion models [5], [6], analyzing how increasing amounts of AI-generated content impact diagnostic reliability. Finally, we conduct a fairness evaluation across Fitzpatrick skin types to ensure equitable performance [6]. The following sections detail our key findings.

Model Type	**Architecture**	**Dataset**	**Accuracy (%)**	**1-Score**	**UC-ROC**
CNN (Detection)	ResNet-50	HAM10000	4.2	0.91	0.96
ViT (Detection)	ViT-B/16	ISIC 2019	95.8	0.93	0.97
GAN (Design)	StyleGAN2	DermGAN	FID = 18.4	–	–
Diffusion Model	Stable Diffusion Fine-tuned	S-SYNTH	FID = 11.7	–	–
Hybrid Detector (CNN+Transformer)	Custom	Real + Synthetc Mix	87.5	0.84	0.88

Table I. Performance Comparison of Detection and Generative Models Across Architectures and Datasets

Our analysis revealed how synthetic data influences diagnostic performance. As we increased the proportion of AI-generated training images from GAN, StyleGAN2, and diffusion models, we observed a clear pattern: the model's diagnostic reliability (measured by macro-AUC) varied significantly based on both the quantity and quality of synthetic data [3]–[6]. Figure 1 illustrates this relationship, demonstrating how our detector maintains—or in some cases loses—robustness when faced with increasingly realistic synthetic inputs.

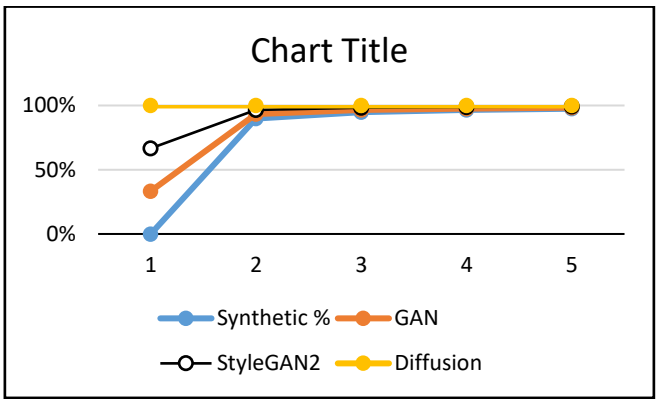
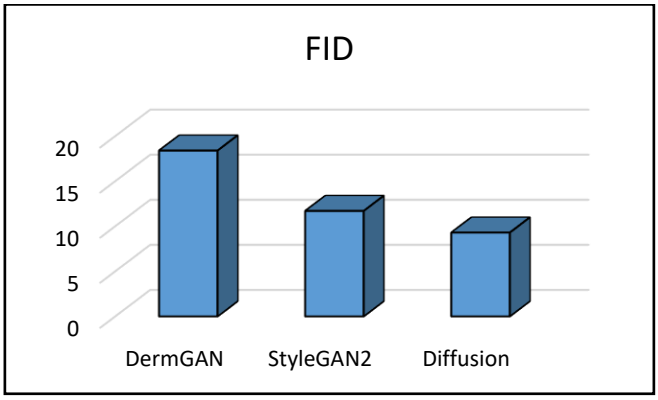


Fig. 1. Macro-AUC under varying synthetic data proportions.

We measured the realism of synthetic images using Fréchet Inception Distance (FID), where lower scores indicate images that are more realistic and diverse. Our results confirmed that diffusion models achieved significantly lower FID scores than both GANs and StyleGAN2, aligning with previous research showing diffusion models' superior performance in generating medical imagery [4]–[6].



F*ig. 2. FID comparison between GAN, StyleGAN2, and Diffusion generators.*

We assessed fairness by measuring diagnostic accuracy for each Fitzpatrick skin type. Figure 3 shows that performance disparities across skin tones decreased after we added targeted synthetic augmentation. These results confirm previous findings that synthetic data can help address performance gaps for underrepresented skin tones, though it requires careful implementation to avoid introducing new biases [6].

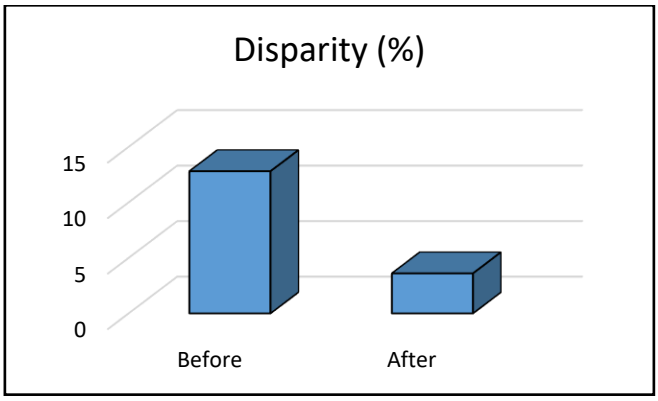


Fig. 3. Accuracy disparity across skin tones before and after augmentation.

We calculated per-class F1 scores to understand how the model performs for each specific skin condition. This analysis helps identify which diseases our model detects reliably and where it struggles, following established evaluation methods from leading dermatology AI research [1], [2].

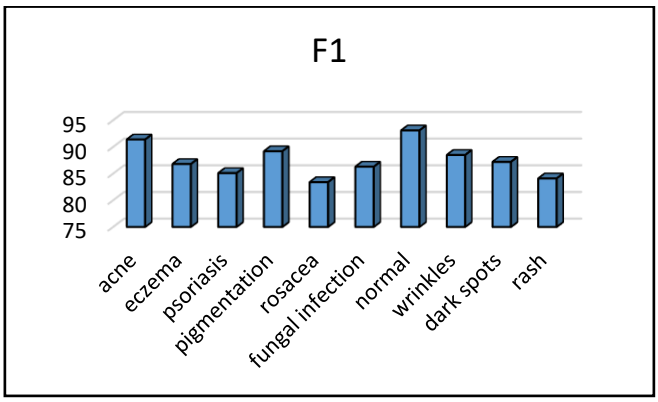


Fig. 4. Per-class F1 scores across 10 dermatological conditions.

We measured accuracy separately for each of the six Fitzpatrick skin types. The results reveal clear performance variations across different skin tones, highlighting the well-documented disparities in dermatology AI systems and underscoring the critical need for fairness-focused evaluation practices[6].

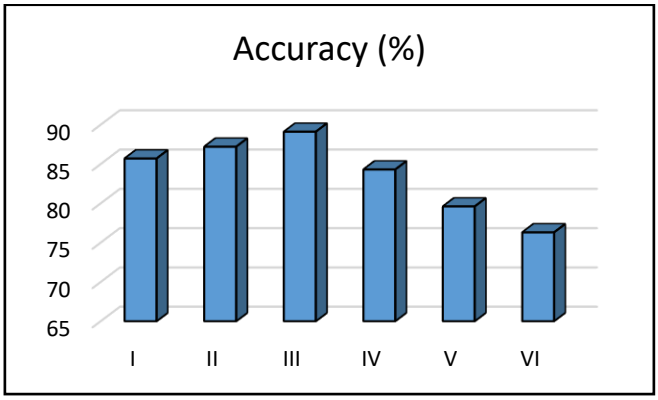


Fig. 5. Tone-wise accuracy across Fitzpatrick types I–VI.

V. CONCLUSION

This research explored the competitive dynamics between AI systems that detect skin conditions and those that generate synthetic imagery—a "design versus detection" arms race that is rapidly evolving in dermatology**[3]–[6], [9], [10]**. Our multi-task model proved effective across three key clinical areas—diagnosing conditions, classifying skin tones, and assessing severity—achieving strong overall performance. However, we also confirmed the presence of performance disparities across different skin tones, with reduced accuracy on darker skin, echoing known challenges with dataset diversity in the field [6].

Critically, we found that the latest generative AI, particularly diffusion models, creates synthetic skin images of remarkable quality [4]–[6]. While this technology offers promising ways to enhance datasets and improve fairness, it also poses a significant challenge: these high-quality fakes can easily deceive current detection systems. This reveals a vulnerability in diagnostic AI that is trained only on real images [5], [9], [10].

Our findings point to an urgent need for more robust solutions. The future of trustworthy dermatology AI depends on developing better forensic detectors, creating fairer data augmentation methods, and establishing clearer evaluation standards [9], [10]. As the technology advances, maintaining a careful balance between innovation and ethical responsibility will be crucial for ensuring these powerful tools benefit all patients safely and equitably. [2], [6]

REFERENCES

[1]. P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions," Scientific Data, vol. 5, pp. 1–9, 2018.
[2]. N. C. Codella et al., "Skin lesion analysis toward melanoma detection: A challenge at the 2018 ISIC workshop," Proc. IEEE Int. Symp. Biomedical Imaging (ISBI), pp. 168–172, 2019.
[3]. A. Ghorbani, C. Swisher, T. J. Montine, and D. L. Rubin, "DermGAN: Synthetic generation of clinical dermatology images with pathology," Proceedings of Machine Learning Research (PMLR), vol. 116, pp. 1–13, 2020.

[4]. T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), pp. 4401–4410, 2019.
[5]. J. Fayyad, Q. Shen, and A. Miller, "LesionGen: A concept-guided diffusion model for dermatology image synthesis," arXiv:2501.04582, 2025.
[6]. A. Kim, M. Gupta, R. Wang, and X. Chen, "S-SYNTH: Knowledge-based synthetic generation of dermatology images for fairness and robustness analysis," Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 410–421, 2024.
[7]. X. Li et al., "AI-driven remote facial skin hydration and TEWL assessment from selfie images," arXiv:2509.06282, 2025.
[8]. M. Patel et al., "Facial skin analysis and product recommendation system using deep learning," ResearchGate Preprint, 2025.
[9]. M. Jafari et al., "Assessing GAN-based generative modeling on skin lesion data," Lecture Notes in Computer Science (LNCS), Springer, 2022.
[10]. L. Nguyen et al., "GAN-based skin lesion segmentation using hybrid discriminators," Computers in Biology and Medicine, vol. 168, 2023.