Advanced visual Turing test for validation and responsible use of synthetic dermoscopic images of melanoma and atypical nevi in medicine

1. Background

Malignant melanoma (MM) is the most aggressive form of skin cancer, and early detection is vital for effective treatment. However, acquiring large, high-quality, annotated image datasets is hindered by privacy constraints, variability in skin types, and the rarity of certain lesion subtypes. Synthetic image generation using Generative Adversarial Networks (GANs), such as StyleGAN2-ADA, offers a promising solution to augment datasets and support training of AI diagnostic models.

Despite improvements in GAN architectures, the validation of synthetic images remains an open challenge. Traditional automated metrics like FID, KID, and precision-recall offer insights into image fidelity and diversity, but fall short of capturing clinical realism and diagnostic utility. As shown in recent studies only a few works have attempted comprehensive validation combining both quantitative metrics and expert dermatological evaluation.

To address this gap, a structured validation framework has been proposed and implemented. This includes a web-based evaluation platform and a robust assessment protocol that collects expert feedback across diagnostic, qualitative, and morphological criteria to gauge the usefulness and realism of GAN-generated dermoscopic images. These synthetic images are generated based on an extended real-world dermoscopic dataset of histopathologically-confirmed MM lesions that serves as a ground-truth [1,2].

[1] Luschi, A., Tognetti, L., Cartocci, A. et al. Advancing synthetic data for dermatology: GAN comparison with multi-metric and expert validation approach. Health Technol. 15, 553–562 (2025). https://doi.org/10.1007/s12553-025-00971-x

[2] Luschi, A., Tognetti, L., Cartocci, A. et al. Design and Development of a Systematic Validation Protocol for Synthetic Melanoma Images for Responsible Use in Medical Artificial Intelligence. Biocybernetics and Biomedical Engineering. 45(4), 608-616 (2025). https://doi.org/10.1016/j.bbe.2025.09.001

[3] Luschi, A., Tognetti, L., Cevenini, G., Rubegni, P., Iadanza, E. (2025). A Region-Specific GAN-Based Solution for Data Augmentation in Dermatology, Lecture Notes in Computer Science, 2025, vol 15735, pp. 240–244, https://doi.org/10.1007/978-3-031-95841-0_45

[4] Luschi, A. et al., "Enhancing Dermatological Diagnostics: A GAN-Based Approach for Synthetic Skin Lesion Image Generation", IFMBE Proceedings, 2025, vol 118, pp. 97-84, https://doi.org/10.1007/978-3-031-86323-3_12

2. Aims and Scope

The main objectives of this protocol are:

- To assess the diagnostic realism of synthetic dermatological images by expert dermatologists.

- To compare synthetic and real images through blinded evaluation across key clinical features.
- To identify potential limitations or artefacts in synthetic data that could affect clinical applicability.
- To contribute toward the standardisation of synthetic image validation in dermatology and related medical fields.

This protocol applies to both research and clinical validation settings and is specifically designed for body-localised melanocytic lesions (excluding face, palms, and soles).

3. Instructions for Evaluators

3.1 Platform Access and Login

- Access the validation platform at: https://webser.dcb.unisi.it

- Register using your institutional email. Approved emails will receive a password to access the evaluation tool.

- Login sessions are valid for 24 hours. After this time, re-authentication is required.

3.2 Evaluation Workflow

Each evaluator will be randomly assigned 60 dermoscopic images (a mix of real and synthetic images). For each image, the following tasks must be performed:

- A. Diagnostic Classification
    - - Label the lesion: Benign (nevus) or Malignant (melanoma)
    - - Identify the image source: Real or Synthetic
- B. Morphological Pattern Recognition
    - Select one or more of the following 16 standard dermatological patterns (as needed):

ID	Pattern Name	Description
1	Atypical network (AN)	Irregularly meshed pigmented network irregularly distributed through the lesion
2	Hypopigmented areas (HA)	Defined interruption of the pigmented network/dermoscopic structures, resulting in structureless areas
3	Irregular dots and globules (IDG)	Sharply circumscribed, round to oval, brown to black structures of various sizes and irregular

		distribution
4	Irregular streaks (IS)	Peripheral brownish to black lines/pseudopods of variable thickness and length, not combined with pigment network lines
5	Irregular pigmented blotches (IPB)	Brown to black irregularly pigmented area circumscribed that precludes recognition of subtler dermoscopic structure/ irregular hyperpigmented structureless areas
6	Blue with veil	Irregular, superficial veil consisting of a confluent, white-blue to whitish-blue pigmentation
7	Blue-grey globules (BGG)	Oval structures of a blue to bluish appearance
8	Blue-grey peppering (BGP)	Fine blue/grey/blue-grey pepper-like structures/granularity
9	White scar-like areas (WSA)	White areas with scar-like appearance/ White scar-like depigmentation
10	Shiny white streaks (SWS)	Linear structures with shiny white colour within a structured area / Chrysalis structures
11	Atypical vascular pattern (AVP)	The presence of two or more of the following types of vessel: hairpin, dotted, linear, irregular, corkscrew, polymorphic vessels
12	Pink areas (PA)	Pink/pinkish shade
13	Reticular pattern	A benign pattern composed of a typical pigment network throughout the entire lesion
14	Globular pattern (GP)	A benign pattern composed of regular globules
15	Homogenous pattern (HP)	Usually seen in congenital nevi, blue nevi or feomelanine phenotype patients
16	Multicomponent pattern	A multi-component pattern consists of a combination of reticular, globular and

		homogenous patterns in the same lesion. If the multi-component pattern is symmetric throughout the lesion, follow-up is recommended

3.3 Navigation and Completion

- After assessing an image, click 'Save' to submit your responses. The next image will load automatically.
- Once saved, an evaluation cannot be modified.

- The evaluation ends after all 60 images have been assessed.

For technical assistance, please contact: alessio.luschi@unisi.it