

No-Reference Quality Assessment of Dermoscopic Images Using Minimal Expert Supervision

Original

No-Reference Quality Assessment of Dermoscopic Images Using Minimal Expert Supervision / Ferraris, Andrea; Branciforti, Francesco; Meiburger, Kristen M.; Veronese, Federica; Zavattaro, Elisa; Savoia, Paola; Salvi, Massimo. - In: APPLIED SCIENCES. - ISSN 2076-3417. - 16:4(2026). [10.3390/app16041682]

Availability:

This version is available at: 11583/3007468 since: 2026-02-10T07:06:22Z

Publisher:

MDPI

Published

DOI:10.3390/app16041682

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

Article

No-Reference Quality Assessment of Dermoscopic Images Using Minimal Expert Supervision

Andrea Ferraris ^{1,†}, Francesco Branciforti ^{1,†}, Kristen M. Meiburger ¹, Federica Veronese ², Elisa Zavattaro ³, Paola Savoia ³ and Massimo Salvi ^{1,*}

¹ Biolab, PoliToBIOMed Lab, Department of Electronics and Telecommunications, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy; andrea.ferraris@polito.it (A.F.); francesco.branciforti@polito.it (F.B.); kristen.meiburger@polito.it (K.M.M.)

² SCU Dermatologia, AOU Maggiore della Carità, C.so Mazzini 18, 28100 Novara, Italy; federica.veronese@med.uniupo.it

³ Department of Health Science, University of Eastern Piedmont, Via Solaroli 17, 28100 Novara, Italy; elisa.zavattaro@med.uniupo.it (E.Z.); paola.savoia@med.uniupo.it (P.S.)

* Correspondence: massimo.salvi@polito.it

† These authors contributed equally to this work.

Abstract

Background: Assessing image quality is critical in medical imaging to ensure diagnostic reliability. Traditional no-reference image quality assessment (IQA) metrics designed for natural images often fail to address the complexities of medical images. This study proposes DermaQA, a novel no-reference metric for dermoscopic images that aligns quality scores with clinical perception. **Methods:** We developed a degradation pipeline simulating realistic artifacts without requiring extensive manual labeling. From 812 expert-classified images, we generated a comprehensive dataset (>125,000 images) using controlled blur and compression techniques. An iterative ranking procedure converted these degradations into a continuous quality scale, which was used to train a vision transformer model. **Results:** The proposed IQA metric outperformed both heuristic and deep learning techniques, achieving 92% accuracy in distinguishing high-quality vs. low-quality images. The approach demonstrated robust generalization when tested on external datasets with different acquisition characteristics, confirming its relevance across varied imaging conditions. **Conclusions:** DermaQA represents the first dermatology-specific quality metric that minimizes expert annotation requirements while maintaining clinical relevance. This tool enhances workflows through real-time acquisition feedback and acts as a gatekeeper for AI diagnostic systems, ensuring only high-quality images are processed. The trained model and inference scripts are publicly available.

Keywords: image quality assessment; dermoscopy; medical imaging; quality controls; no-reference metrics

Academic Editors: George Pavlidis, Athanasios Nikolaidis and Charalampos Strouthopoulos

Received: 16 January 2026

Revised: 2 February 2026

Accepted: 4 February 2026

Published: 7 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Evaluating image quality is crucial in medical imaging, where it directly impacts the reliability of downstream analyses and clinical decisions [1–4]. The challenge is particularly significant in multi-centric and multi-device studies, where image harmonization becomes essential for standardizing appearances and ensuring consistent analysis across varied acquisition protocols and equipment [5]. In dermoscopic imaging,

image quality encompasses several aspects that directly impact diagnostic reliability: (1) focus quality, affecting the visibility of fine structures and patterns; (2) presence of artifacts such as hair, air bubbles, or compression distortions; (3) illumination adequacy, ensuring proper exposure; and (4) complete lesion visibility with appropriate margins.

Image quality assessment (IQA) metrics can be broadly categorized based on their reference requirements. Full-reference metrics compare test images against an ideal reference, providing direct measures of fidelity through methods like peak signal-to-noise ratio (PSNR) and structural similarity index metric (SSIM). Task-specific metrics, such as contrast-to-noise ratio (CNR) or signal-to-noise ratio (SNR), evaluate how well an image supports specific diagnostic tasks. However, these approaches have significant limitations in medical contexts. Critical evaluations have questioned their suitability [6], as they often fail to capture subtle yet diagnostically significant distortions across various modalities including CT, MRI, OCT, X-ray, and dermoscopic images [7]. Moreover, they are limited by their dependence on manual region of interest (ROI) selection and their lack of generalizability across different imaging contexts [8,9].

No-reference metrics address these limitations by assessing quality without requiring any reference image, instead relying on intrinsic statistical properties and learned features from the test image itself [10]. This capability makes them particularly valuable for clinical applications, especially in contexts like dermoscopy, where image acquisition conditions can vary significantly and reference standards are rarely available. However, developing effective no-reference metrics for medical imaging presents several challenges. Traditional no-reference metrics designed for natural images often fail to capture diagnostically relevant features and subtle quality variations that are crucial in medical contexts. Additionally, training such metrics typically requires extensive manual quality assessments by experts [11], creating a significant resource bottleneck.

In the specific context of digital dermatology, an emerging field that relies heavily on photographic imaging for diagnosis and monitoring of skin conditions, quality assessment becomes particularly crucial. The increasing adoption of varied acquisition devices, from professional dermoscopes to smartphone adapters used in telemedicine settings [1,12,13], creates an urgent need for robust quality control mechanisms that can optimize image acquisition protocols, provide feedback to clinicians, and filter suitable images for AI model training or inference [3].

To address these challenges, we propose *DermaIQA*, a novel no-reference methodology specifically designed for dermoscopic image quality assessment. Although binary classification (high vs. low quality) might seem sufficient, continuous quality scores offer significant advantages. Such scores enable quantitative assessment of image improvement or deterioration while providing precise feedback for optimizing acquisition protocols. Moreover, continuous scores allow for flexible thresholds that can be adjusted based on specific clinical requirements or downstream analysis needs. Our approach leverages the predictable relationship between synthetic degradations and quality perception, eliminating the need for extensive manual scoring by clinical experts. By combining controlled image degradation with a continuous ranking system, we generate comprehensive quality comparisons that closely approximate human perception while requiring minimal expert supervision.

1.1. Related Works on No-Reference IQA Metrics

In this work, we focus on the development of a no-reference IQA metric to evaluate image quality without requiring a ground-truth reference image. While metrics like *MANIQA* [14], *MUSIQ* [15], and *HYPERIQA* [16] have shown effectiveness for natural images by analyzing statistical and perceptual features [17–19]. However, these metrics fail to address the specific requirements of medical imaging, where domain-specific

diagnostic features (such as lesions, tissue patterns, and anatomical structures) require different evaluation criteria. As a result, conventional metrics often produce quality assessments that do not align with clinical utility.

A significant barrier to developing medical-specific no-reference IQA metrics is the difficulty in obtaining large-scale quality assessments. Traditional approaches require extensive manual labeling by domain specialists like radiologists or dermatologists, creating resource bottlenecks that limit dataset size and diversity. This constraint has driven recent research toward developing more efficient methodologies that can reduce dependence on manual annotation while maintaining clinical relevance.

Several approaches have emerged to address these challenges across various medical imaging modalities [20,21]. For instance, the Uno-QA framework [22] demonstrates the value of modality-specific solutions through its use of multi-scale pyramid pooling for OCTA image quality assessment. In the field of transvaginal ultrasound, researchers have shown success by fine-tuning transformers with multi-annotator labels [23]. For X-ray CT, local keypoint-based methods [24] have proven effective in addressing specific quality concerns in radiological imaging.

Recent advances have also focused on improving the interpretability of quality assessment methods. New IQA strategies incorporating saliency maps [25] help highlight quality-degrading regions. Hybrid frameworks combining subjective and objective evaluation have shown promise, as demonstrated by tools that integrate expert surveys with standard metrics to identify degraded medical images [26].

A general methodological framework has begun to emerge for developing domain-specific no-reference IQA metrics in medical imaging. This pipeline typically involves: (1) collecting domain-specific images, (2) classifying them by quality level through either expert annotation or analytical methods, (3) training specialized neural networks on these examples, and (4) validating the resulting metrics against clinical standards. This framework is adaptable across medical specialties while maintaining the essential focus on clinical relevance.

1.2. Proposed Approach and Main Contributions

Our research focuses on digital dermatology, where the increasing diversity of image acquisition devices (from professional dermoscopes to smartphone adapters) creates the need for robust quality assessment methods. The heterogeneity in image capture technology necessitates reliable quality control mechanisms that can:

- Optimize image acquisition protocols
- Provide immediate feedback to clinicians
- Filter suitable images for AI model training or inference
- Ensure consistency across different imaging devices and settings

For optical imaging modalities like dermoscopy, image quality degradation follows predictable patterns that can be objectively characterized. When an image exhibits blur, compression artifacts, or noise, its diagnostic quality is inherently reduced compared to its original counterpart. We leverage this relationship to develop *DermaIQA*, a no-reference quality assessment network that evaluates dermoscopic images without extensive expert supervision. The main contributions of our work are:

- We propose a novel data preparation methodology combining controlled synthetic degradation with objective quality scoring to create a comprehensive assessment system that aligns with clinical perception while capturing critical diagnostic patterns.
- We introduce a new training framework using a continuous ranking system that significantly reduces the need for expert annotations.

- Our metric achieves 92% accuracy in distinguishing between high- and low-quality dermoscopic images as classified by expert dermatologists, showing strong alignment with clinical quality perception.
- We provide a transferable framework that can be extended to other medical imaging modalities, offering a blueprint for developing domain-specific quality metrics with limited annotation resources.

The remainder of this paper is organized as follows: Section 2 describes our dataset collection and the detailed implementation of the proposed metric, including the degradation pipeline and training strategy. Section 3 presents our experimental results, including benchmarks on the ISIC test set and comparisons with existing IQA metrics. Section 4 discusses the implications of our findings and current limitations, while Section 5 concludes the paper with final remarks and future directions.

2. Materials and Methods

The complete methodology for developing and validating DermaIQA is illustrated in Figure 1. We begin by collecting dermoscopic images from diverse sources, followed by expert-guided classification of a subset into high-quality (D_0) and low-quality (D_2) categories. We then apply synthetic degradation techniques to expand the quality spectrum, creating additional degraded versions (D_1 and D_3). Each image is divided into overlapping patches, and a continuous quality ranking is established using both degradation categories ($D_0 > D_1 > D_2 > D_3$) and perceptual metrics, specifically Peak Signal-to-Noise Ratio (PSNR). This hierarchical evaluation is then transformed through an Elo rating system [27], which provides a continuous quality score for each patch. These scores serve as training labels for an attention-based Vision Transformer (ViT) architecture [14].

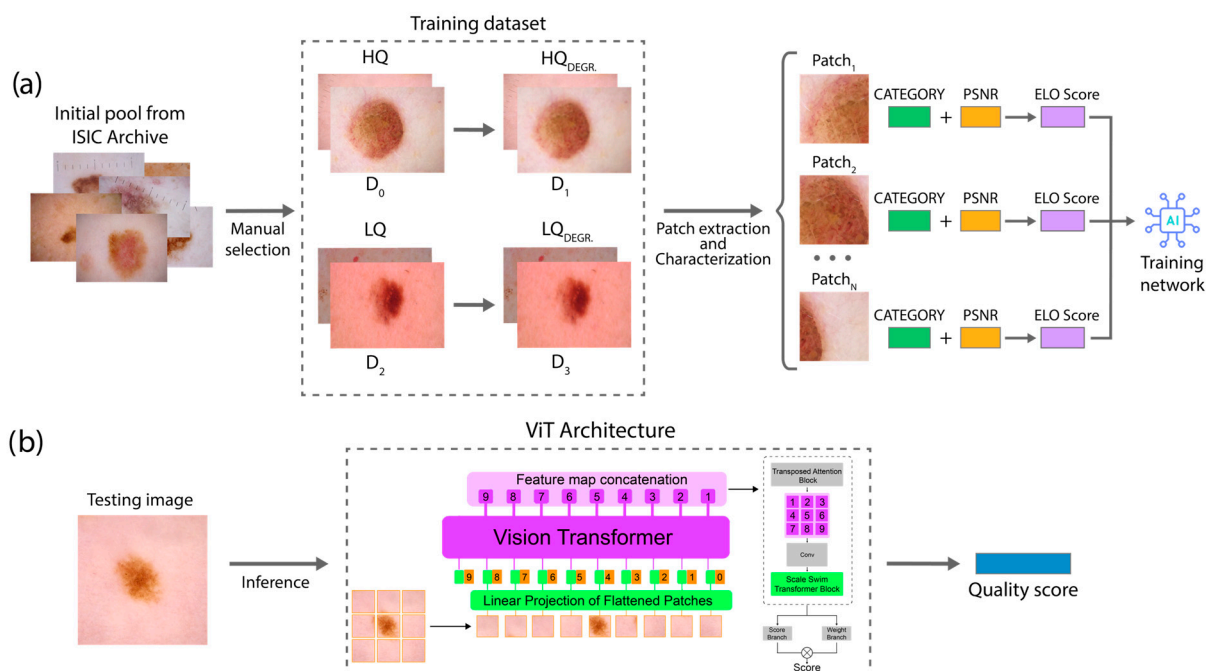


Figure 1. Overview of the proposed methodology. (a) Training stage: dermoscopic images are collected and classified by experts into quality categories (D_0 , D_2), then undergo synthetic degradation (D_1 , D_3). After patch extraction, a continuous quality ranking is computed and used to train an attention-based Vision Transformer (ViT) for no-reference IQA. (b) Testing stage: the trained network evaluates new dermoscopic images without requiring reference standards.

In the testing stage, the final model is evaluated through a comprehensive pairwise comparison against expert dermatologists' classifications. To promote reproducibility and facilitate further research in this area, we have made our inference code and test dataset publicly available through Mendeley Data (10.17632/2ryw3hpb6v.1), including labeled high-quality and low-quality dermoscopic images used in our evaluation.

2.1. Dataset

The initial dataset was collected from a pool of 15,857 dermoscopic images sourced from the ISIC Archive [28,29]. We selected only images with a minimum resolution of 1024×1024 pixels to ensure adequate spatial resolution and lesion visualization. From this collection, an experienced dermatologist manually selected 689 high-quality (HQ) images and 423 low-quality (LQ) images through visual inspection. This curated subset provided a manageable dataset while minimizing expert annotation requirements, with selection based on specific quality criteria. Images were labeled as 'HQ' if they met all the following requirements:

- Complete visualization of the lesion with appropriate margins (minimum 10% border area around the lesion)
- Absence of significant artifacts (less than 5% of the lesion area affected by hair, air bubbles, ruler markings, ink)
- Proper focus throughout the region of interest
- Adequate lighting with balanced exposure

On the other hand, images were labeled as 'LQ' if they exhibited one or more of the following deficiencies:

- Substantial artifacts obscuring diagnostically relevant features (>10% of lesion area affected);
- Poor focus rendering critical patterns indiscernible;
- Improper illumination (>15% pixels under- or overexposed);
- Incomplete lesion visualization (lesion extends beyond image boundaries).

Following these clinical criteria, DermaQA was designed to assess specific technical aspects of image quality that directly impact diagnostic utility. The metric evaluates image sharpness and detail preservation through blur detection, identifies compression artifacts and their impact on pattern analysis, and measures overall image fidelity through combined analysis of blur and compression effects.

The remaining images in the ISIC Archive were not included as they represented borderline quality cases requiring multiple expert evaluations or contained other complicating factors such as non-standard acquisition techniques. This selective approach was intentional: by training on clearly defined high- and low-quality examples, we aimed to establish robust and objective quality boundaries.

The selected images ($n = 1112$) were split into training and testing sets, with the training set comprising 539 HQ images (D_0) and 273 LQ images (D_2), and the testing set containing 150 HQ and 150 LQ images. To create a comprehensive quality spectrum, we generated additional quality levels using our custom degradation pipeline: D_1 (degraded versions of D_0 images) and D_3 (degraded versions of D_2 images). The complete composition of the primary dataset is presented in Table 1.

For external validation, we included three additional datasets:

- The PH² dataset [30]: 200 dermoscopic images acquired with professional equipment
- The PAD-UFES-20 dataset [31]: 2044 smartphone-based dermoscopic images (from an original set of 2298)
- A private dataset from Novara Hospital (Novara, Italy): 149 images collected using the Nurugo smartphone dermoscope [13]

Table 1. Distribution of dermoscopic images across different quality levels. D_0 : original HQ images, D_1 : degraded versions of D_0 , D_2 : original LQ images, D_3 : degraded versions of D_2 .

| Subset | Quality Level | Number of Images |
|----------------------|------------------------------|------------------|
| Training images | D_0 (original HQ) | 539 |
| | D_1 (degraded HQ) | 539 |
| | D_2 (original LQ) | 273 |
| | D_3 (degraded LQ) | 273 |
| Testing—ISIC Archive | D_0 | 150 |
| | D_2 | 150 |
| External validation | PH ² (dermoscope) | 200 |
| | PAD-UFES-20 (smartphone) | 2044 |
| | Novara (smartphone) | 149 |

2.2. Custom Degradation Pipeline

The degradation pipeline was implemented to create a dataset spanning the full spectrum of quality levels by simulating typical artifacts encountered in clinical dermoscopy. Starting from the expert-classified images, we created four quality-based subsets: original high-quality images (D_0), their degraded variants (D_1), original low-quality images (D_2), and their degraded variants (D_3). The entire degradation process is illustrated in Figure 2.

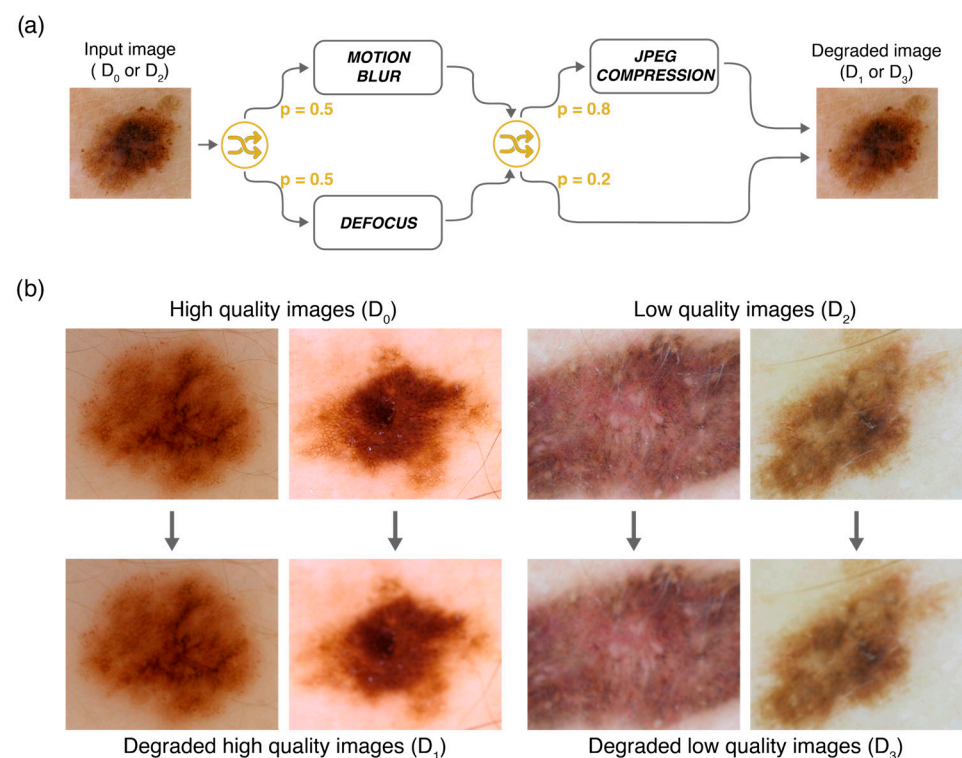


Figure 2. Overview of the synthetic degradation process and its effects. (a) Schematic representation of the degradation pipeline used to simulate real-world artifacts in dermatoscopic images, creating additional quality levels D_1 and D_3 from the original D_0 and D_2 images. (b) Examples showing the visual impact of the degradation pipeline, demonstrating the quality reduction from original to degraded images.

We selected defocus blur, motion blur, and JPEG compression as our primary degradation types based on a systematic analysis of common quality issues in clinical dermoscopic practice. Defocus blur frequently occurs due to incorrect focal distance or

mechanical instability of handheld dermatoscopes. Motion blur is common in clinical settings where patient movement or operator hand tremor can affect image acquisition. JPEG compression artifacts are increasingly relevant as telehealth platforms and electronic health records often automatically compress images for storage and transmission. While other degradations exist (such as specular reflections, uneven illumination, or color distortions), these three types of degradation represent fundamental technical issues that can significantly impact diagnostic quality in dermoscopic imaging.

Using the Albumentation library [32], our pipeline followed a parameter-driven approach to ensure reproducibility and precise control. For each image, the process began with either defocus blur or motion blur [33], randomly selected based on configuration settings. Following this initial degradation, JPEG compression [34] was subsequently applied with a probability of 0.8, further simulating quality loss introduced during storage or transmission.

2.2.1. Motion Blur Degradation

Motion blur was applied to simulate image quality loss caused by movement, which may result from either patient motion or handheld device instability. This effect was implemented by convolving the original image $I(x, y)$ with a linear motion blur kernel $K(x, y)$:

$$I_{\text{blurred}}(x, y) = (I * K)(x, y) = \sum_{u=-\lfloor \frac{k}{2} \rfloor}^{\lfloor \frac{k}{2} \rfloor} \sum_{v=-\lfloor \frac{k}{2} \rfloor}^{\lfloor \frac{k}{2} \rfloor} I(x - u, y - v)K(u, v) \quad (1)$$

where $I_{\text{blurred}}(x, y)$ represents the resulting blurred image, and k denotes the size of the blur kernel. The kernel size k was randomly chosen between 9 and 13 pixels, a range determined through preliminary experiments to simulate realistic motion artifacts observed in clinical dermoscopic imaging while avoiding excessive degradation that would render images diagnostically unusable.

The direction of the motion was controlled by an angle θ that was uniformly sampled from a range:

$$\theta \sim U(0^\circ, 360^\circ) \quad (2)$$

This random angle enabled the simulation of movement in any direction, effectively representing various clinical scenarios where hand tremors or patient movements can degrade image quality.

2.2.2. Defocus Blur Degradation

Defocus blur simulated the artifacts caused by incorrect focal distance during dermoscopic image capture. This degradation was implemented by convolving the original image with a disc-shaped kernel:

$$I_{\text{defocus}}(x, y) = (I * D_r)(x, y) \quad (3)$$

The disk kernel with radius r (D_r) was randomly chosen between 3 and 10, with smaller values producing subtle blurring that mimics minor focus issues, while larger values approximate more significant defocus artifacts commonly encountered during handheld dermoscopic imaging.

To ensure natural appearance and reduce aliasing artifacts, a Gaussian kernel was applied with standard deviation:

$$\sigma_{\text{alias}} \sim U(\sigma_{\text{min}}, \sigma_{\text{max}}) \quad (4)$$

where $\sigma_{\min} = 0.1$ and $\sigma_{\max} = 0.5$, controlling the level of smoothing applied to the edges of the defocus effect.

2.2.3. JPEG Compression Artifacts

JPEG compression was applied to simulate artifacts commonly introduced during image storage or transmission in clinical settings. The degradation applied compression at a randomly selected quality level Q between 80 and 90, intentionally focused on the higher quality range to simulate the compression artifacts typically present in clinical images rather than severely compressed images that would be immediately rejected in practice. This JPEG compression process involves five primary steps:

1. **Color Space Transformation:** The image, originally in RGB format, is first transformed into the YCbCr color space. This separates luminance (Y) from chrominance (Cb and Cr), allowing for more efficient compression since the human visual system is less sensitive to changes in chrominance compared to luminance.
2. **Block Splitting and Discrete Cosine Transform (DCT):** The image is divided into non-overlapping 8×8 blocks, and each block undergoes the Discrete Cosine Transform (DCT) [35] to convert spatial pixel information into frequency components:

$$C(u, v) = \frac{1}{4} \sum_{x=0}^7 \sum_{y=0}^7 I(x, y) \cos \left[\frac{(2x+1)u\pi}{16} \right] \cos \left[\frac{(2y+1)v\pi}{16} \right] \quad (5)$$

3. **Quantization:** The DCT coefficients are quantized by dividing each coefficient by a corresponding value from a quantization matrix $Q_{u,v}$, followed by rounding:

$$C_q(u, v) = \text{round} \left(\frac{C(u, v)}{Q_{u,v}} \right) \quad (6)$$

Lower quality values of Q lead to more aggressive quantization, which results in significant information loss and visible compression artifacts.

4. **Entropy Coding:** The quantized coefficients are then compressed through entropy coding techniques. In JPEG terminology, the first coefficient of each block (representing the average value) is called the DC coefficient, while the remaining 63 coefficients (representing increasingly fine details) are called AC coefficients. These coefficients are encoded differently: DC coefficients using Differential Pulse Code Modulation (DPCM) [36], and the AC coefficient using run-length encoding and Huffman coding [37] based on their patterns of zeros.
5. **Reconstruction:** During decompression, this process is reversed. The encoded data is decoded and then dequantized by multiplying with the same quantization matrix:

$$C_{\text{dequantized}}(u, v) = C_q(u, v) \cdot Q_{u,v} \quad (7)$$

The Inverse Discrete Cosine Transform (IDCT) is then applied to reconstruct the pixel values, and the image is transformed back to the RGB color space. The complete process introduces three characteristic artifacts: blocking (visible grid-like patterns at block boundaries), blurring (loss of fine details), and ringing (oscillations near sharp edges). These artifacts realistically simulate the quality degradation seen in compressed clinical images.

2.3. Patch Extraction and Characterization

Following the degradation process, we characterized image quality through local analysis. We first cropped 5% from each edge of all images to eliminate potential border artifacts. Using the Patchify library [38], each image was then divided into 384×384 pixel

patches with 50% overlap, ensuring continuity and minimizing information loss at boundaries. This process yielded 125,202 patches distributed across our quality levels: 45,168 from D_0 images, 45,168 from D_1 images, 17,433 from D_2 images, and 17,433 from D_3 images.

To quantify degradation between original and synthetically degraded pairs (D_0 vs. D_1 and D_2 vs. D_3), we calculated the Peak Signal-to-Noise Ratio (PSNR) [39] for each corresponding patch pair:

$$PSNR = 20 \times \log_{10} \left(\frac{MAX_{pixel}}{\sqrt{MSE}} \right) \quad (8)$$

where $MAX_{pixel} = 255$ for 8-bit images and MSE represents the mean squared error between the original and degraded patches. Each PSNR value calculated for a given pair of patches was subsequently assigned to both the original and degraded patches.

Elo Rating System for Quality Ranking

While traditional IQA metrics rely on thousands of human visual scores for training, our approach introduces a novel methodology that significantly reduces dependency on subjective evaluations. By adapting the Elo rating system [27] (originally developed for chess player rankings) to image quality assessment, we generate several quality comparisons from a limited set of expert-classified images. This represents a significant advance from conventional IQA training paradigms that require extensive human scoring datasets.

Our implementation evaluates relative quality through pairwise comparisons, with each image patch starting at a neutral rating of 3000. For any two images A and B with ratings $R_A(t)$ and $R_B(t)$ at iteration t , the expected match outcome probabilities are calculated as:

$$\begin{cases} E_A(t) = 20 \times \log_{10} \left(\frac{1}{1 + 10^{\frac{R_B(t) - R_A(t)}{400}}} \right) \\ E_B(t) = 1 - E_A(t) \end{cases} \quad (9)$$

After comparing the images, the ratings are updated for the next iteration ($t + 1$):

$$\begin{cases} R'_A(t + 1) = R_A(t) + K \times (S_A - E_A(t)) \\ R'_B(t + 1) = R_B(t) + K \times (S_B - E_B(t)) \end{cases} \quad (10)$$

where $K = 32$ is a scaling factor controlling adjustment sensitivity, and $S_A, S_B \in \{0, 0.5, 1\}$ represent the match results (loss, draw, win). Our implementation followed a systematic iterative process:

1. Initialization: All image patches were assigned an initial Elo rating of 3000.
2. Pairing: Image patches were paired using an adaptive strategy throughout the training process. In early iterations, patches were matched across broader rating differences to establish initial hierarchies. The pairing range then progressively narrowed as training advanced, starting from ± 100 points and gradually decreasing to ± 50 points in later iterations. This approach helped establish a stable global ranking while preventing local ranking cycles that could emerge from fixed-range comparisons.
3. Outcome determination: For each pair, we first compared their degradation category (D_0, D_1, D_2, D_3), with the established quality hierarchy $D_0 > D_1 > D_2 > D_3$. When comparing patches within the same degradation level, we employed PSNR as the final quality indicator. If all metrics showed differences below 0.1%, the match was declared a draw.
4. Rating update: After each comparison, we updated the Elo ratings using equation (10).
5. Convergence: We repeated steps 2–4 until the maximum absolute change in Elo ratings across all images fell below 10^{-3} , indicating stable relative quality rankings. We

repeated steps 2–4 until meeting two criteria: maximum absolute change in Elo ratings across all images below 10^{-3} and no rating inversions in the last 500 comparisons. This ensured convergence to a stable and reliable ranking system.

This approach transformed our discrete quality levels (D_0 , D_1 , D_2 , D_3) into a continuous quality spectrum, with Elo scores ranging from 2000 to 4000. These scores were then min-max normalized to $[0, 1]$ range for training the ViT network. The continuous nature of these normalized scores enabled the network to learn fine-grained quality distinctions that aligned well with human perception of image quality.

2.4. No-Reference Metric Network Training

Our implementation builds upon the ViT architecture provided by the GitHub repository of MANIQA (Multi-dimensional Attention Neural IQA) [14]. This framework offers a complete pipeline for training IQA models on continuous quality scores. The model uses a transformer-based architecture with a Swin Transformer [40] backbone, modified with a multi-dimensional attention mechanism that analyzes both spatial and channel relationships in feature maps.

Our dataset partitioning strategy was designed to optimize model performance and generalization. The training set included samples from all quality categories (D_0 , D_1 , D_2 , D_3). Including degraded variants in the training set acted as an effective augmentation technique, exposing the model to a broader spectrum of quality variations. The validation set, instead, contained only original high-quality (D_0) and low-quality (D_2) image patches, ensuring specific assessment of the network's ability to distinguish between genuine quality differences.

The network was trained with a learning rate of 2×10^{-5} using cosine annealing, a batch size of 8, and the AdamW optimizer with weight decay 0.01. Training was set for 50 epochs with early stopping if no improvement was observed after 5 epochs. The training process concluded at epoch 15, with both correlation coefficients reaching their peak values at epoch 7 on the validation set: Spearman's Rank Correlation Coefficient (SRCC) at 0.82 and Pearson Linear Correlation Coefficient (PLCC) at 0.89. These results determined the selection of the final model weights (Figure 3). The entire training process was conducted on an NVIDIA A6000 GPU and required approximately 3 h to complete.

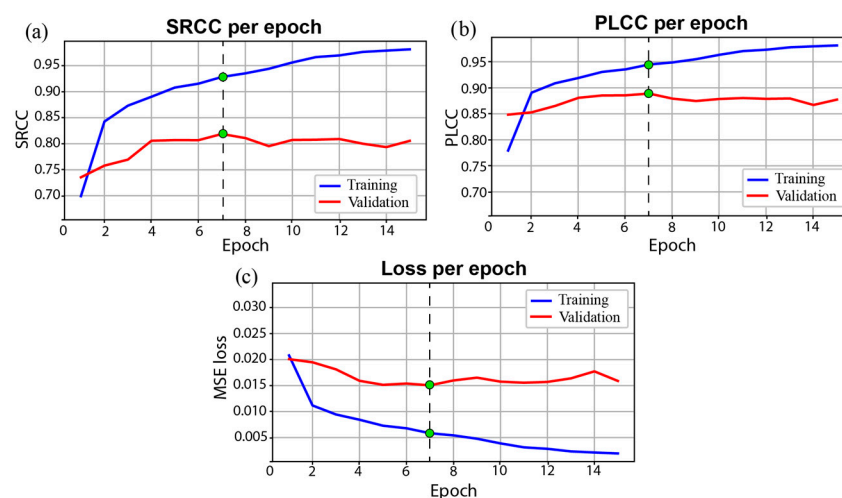


Figure 3. Training and validation metrics across epochs. The plots show (a) Spearman's Rank Correlation Coefficient (SRCC), (b) Pearson Linear Correlation Coefficient (PLCC), and (c) Mean Squared Error (MSE) loss for both training and validation sets. Both correlation coefficients reached their maximum values at the seventh epoch, after which they maintained consistent performance. Green circles highlight the chosen epoch (7th).

2.5. Inference Strategy

Our method employs a patch-based approach for dermoscopic image quality evaluation. Each input image is decomposed into overlapping 384×384 pixel patches with 50% stride. The trained ViT network processes each patch independently, generating patch-level quality scores that are then aggregated through arithmetic mean to produce a single image-level quality score.

Unlike the training phase, where 5% border regions were excluded to eliminate potential artifacts, our inference strategy intentionally includes these border areas. This design choice ensures that no clinically relevant information at image boundaries is overlooked, particularly important when lesions extend to the edges of the frame. The patch-based approach provides three key advantages: (i) it captures localized quality variations across the image, (ii) enables processing of dermoscopic images of arbitrary dimensions without relying on resizing operations, and (iii) remains computationally efficient even for high-resolution images. The inference procedure of DermalQA, implemented in Python 3.9.16 using PyTorch 2.2.1, follows a systematic approach (Algorithm 1) that ensures efficient processing of dermoscopic images of any size.

Algorithm 1. DermalQA Inference Strategy.

Input: Dermoscopic image I of arbitrary size

Output: Quality score Q and optional patch-level analysis

1. Image preprocessing:
 - Load and normalize image to [0, 1] range
 - Convert to RGB format
 2. Patch extraction:
 - Split image into overlapping patches (384×384 pixels, 50% overlap)
 - Process patches regardless of input image dimensions
 3. Quality assessment:
 - For each patch:
 - Apply trained Vision Transformer model
 - Compute patch-level quality score
 4. Score aggregation:
 - Compute final image quality as mean of patch scores
-

2.6. Evaluation Criteria

We structured our evaluation to assess both the performance and generalization capabilities of DermalQA through four main analyses. First, we evaluated the metric's ability to distinguish between quality levels using the ISIC testing set described in Section 2.1, comprising 150 HQ and 150 LQ dermoscopic images classified by experienced dermatologists. We analyzed the distribution of quality scores across both classes to assess the metric's ability to create separable quality distributions.

Second, we performed a thorough pairwise evaluation, generating all possible combinations between LQ and HQ images ($150 \times 150 = 22,500$ unique pairs). For each pair (HQ, LQ), we calculated quality scores for both images and determined if the metric correctly ranked the HQ image higher. This 'discrimination accuracy' represents the percentage of correct rankings, directly assessing the metric's alignment with expert clinical judgment.

Third, we compared DermalQA with existing approaches to evaluate the advantages of our domain-specific training. We evaluated state-of-the-art no-reference IQA metrics including MANIQA [14], TRES [17], MUSIQ [15], HYPERIQA [16], UNIQUE [41], ILNIQE [19], and NIQE [18], using pre-trained weights from the PyIQA library [42]. We also included established heuristic methods that assess different image quality aspects:

Tenengrad [43] (measuring sharpness through gradient magnitude), Laplacian Variance [44] (capturing focus quality), Entropy [45] (quantifying information content), and Edge Density [46] (evaluating structural clarity).

Finally, we assessed generalization capabilities across different acquisition devices and settings using three external datasets. Specifically, we evaluated performance on professional dermoscopic images using the PH² dataset [30], containing 200 high-quality dermoscopic images. To assess generalization to smartphone-based dermoscopy, we used two additional datasets: the public PAD-UFES-20 dataset [31] comprising 2298 images, and a private dataset of 149 images acquired at Novara Hospital using the Nurugo smartphone dermoscope [13].

3. Results

3.1. Benchmark on the ISIC Test Set

Figure 4 illustrates the performance of DermaIQA in distinguishing between LQ and HQ dermoscopic images on the ISIC test set. The Kernel Density Estimate (KDE) plots in Figure 4a show the distribution of metric scores for both LQ ($n = 150$) and HQ ($n = 150$) images. The metric achieves an Area Under the ROC Curve (AUC) of 0.91. Based on the ROC analysis on the validation set, we selected a threshold of 0.4 as the operating point, balancing sensitivity and specificity. When applied to the test set, this threshold effectively separates the two quality groups, with 84.00% of LQ images scoring below it and 89.33% of HQ images scoring above it (Figure 4b).

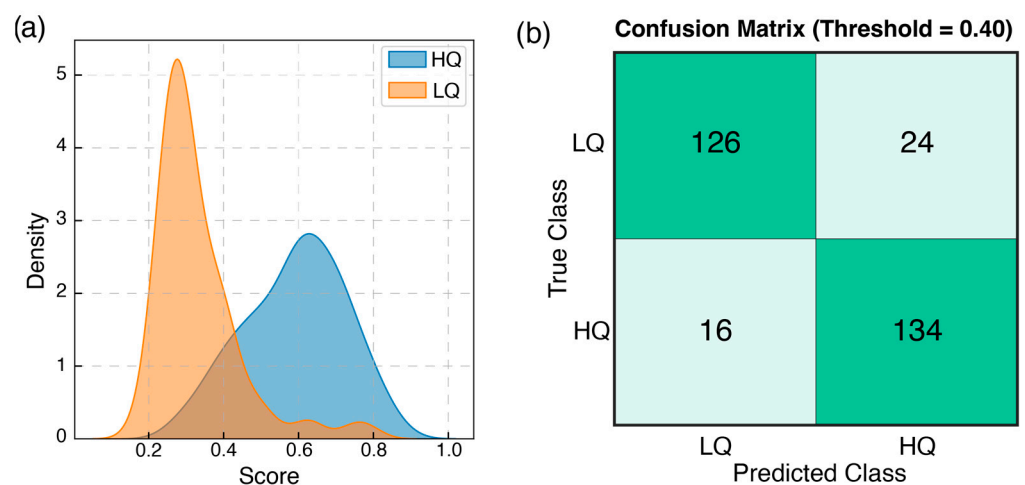


Figure 4. Performance evaluation on the ISIC test set. (a) Kernel Density Estimate (KDE) plots of the proposed no-reference IQA metric scores for LQ and HQ images. (b) Confusion matrix computed using a threshold of 0.4, which differentiates the two distributions.

Figure 5 presents qualitative examples of the metric's performance. Figure 5a shows correctly classified cases where the metric assigns appropriate scores to both LQ and HQ images, demonstrating alignment with expert dermatologists' assessments. Figure 5b shows misclassification examples, where an LQ image received a high-quality score and an HQ image received a low-quality score. These cases highlight specific scenarios where the metric could benefit from further refinement to improve its reliability across the full spectrum of image qualities.

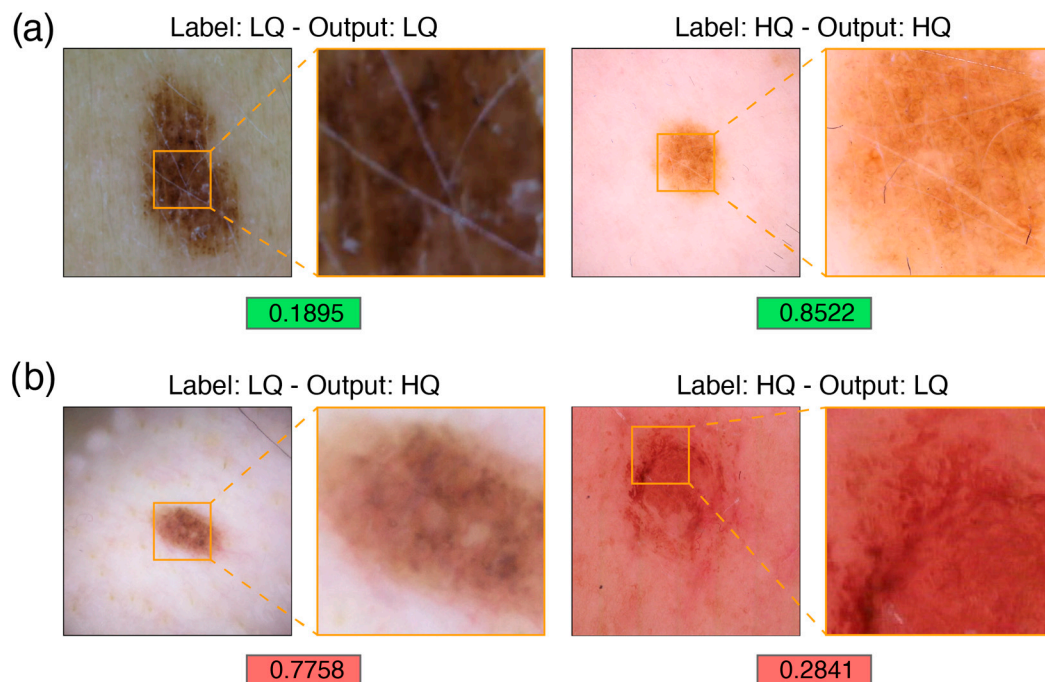


Figure 5. Performance evaluation on the ISIC test set. (a) Examples of correct classifications. (b) Examples of misclassifications by the proposed metric.

To validate our degradation pipeline and evaluate the consistency of quality ordering across our four-level classification (D_0 , D_1 , D_2 , D_3), we performed a detailed analysis of the quality score distributions. Figure 6 illustrates how our metric responds to synthetic degradations. The KDE plots (Figure 6a) demonstrate a good level of separation between the four quality levels, with D_0 and D_1 distributions showing similar shapes but distinct shifts in quality scores, mirroring the pattern observed between D_2 and D_3 . Quantitative analysis of these relationships (Figure 6b) confirms the statistical significance of quality differences between adjacent levels. Paired t -tests revealed significant differences between D_0 (mean score: 0.59 ± 0.13) and D_1 (mean score: 0.38 ± 0.14) with $p < 0.001$, and between D_2 (mean score: 0.33 ± 0.11) and D_3 (mean score: 0.27 ± 0.09) with $p < 0.001$.

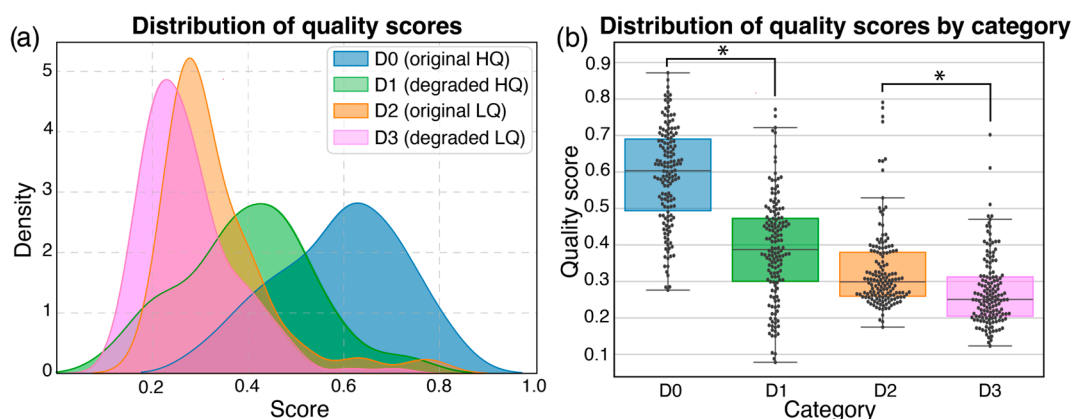


Figure 6. Validation of quality level separation and degradation effects. (a) Kernel density estimation plots showing the distribution of quality scores across all four quality levels (D_0 , D_1 , D_2 , D_3). (b) Box plots with overlaid data points showing the distribution of quality scores for each quality level. Asterisks indicate statistical significance ($* p < 0.001$) in paired t -tests between adjacent quality levels (D_0 – D_1 and D_2 – D_3).

3.2. Comparison with IQA Metrics

We compared our metric against both traditional heuristic methods and state-of-the-art (SOA) IQA metrics using all 22,500 possible LQ-HQ image pairs from our test set. All SOA metrics were used with their pre-trained weights from the PyIQA library, allowing us to assess how general-purpose IQA solutions perform when directly applied to dermoscopic images. This comprehensive evaluation assessed each metric’s ability to correctly rank image quality across diverse dermoscopic characteristics.

The comparison with heuristic methods (Figure 7a) shows that DermaIQA achieves 92% accuracy in correctly ranking image pairs, significantly outperforming traditional approaches. The closest competitor, Tenengrad, achieves 85% accuracy, while other heuristic methods like Laplacian Variance, Entropy, and Edge Density show lower performance. Similarly, when compared to SOA no-reference IQA metrics (Figure 7b), our approach maintains its 92% accuracy, substantially exceeding the performance of established metrics like ILNIQE (80%), NIQE (83%), and TRES (72%).

To better understand these performance differences, we analyzed four representative image pairs of increasing complexity (Figure 7c). For the first pair, showing obvious quality differences, most metrics (7 out of 11) correctly identified the higher quality image. However, metric performance degraded significantly with more challenging cases. The second pair, featuring red-toned images, was correctly classified by only three metrics. For the third pair, containing subtle artifacts, only the Entropy metric among competitors succeeded in correct classification. The fourth and most challenging pair, where the low-quality image contained sharp hairs that could be misinterpreted as high-quality features, was correctly classified only by our proposed method.

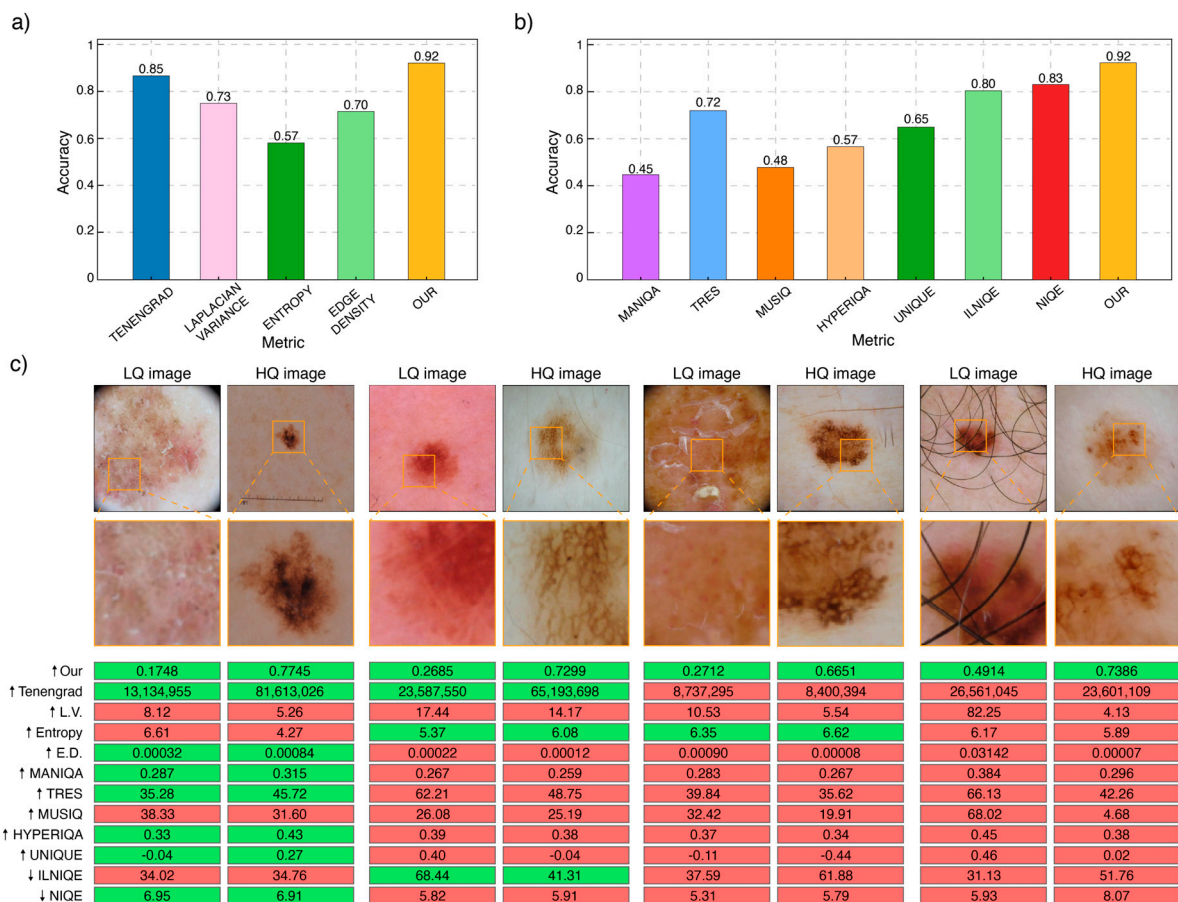


Figure 7. Performance comparison with existing metrics. (a) Accuracy of heuristic methods versus our proposed metric. (b) Accuracy of state-of-the-art no-reference IQA methods. (c) Four LQ-HQ

image pairs with quality scores from all compared metrics. Arrows indicate score interpretation: \uparrow (higher scores = better quality) or \downarrow (lower scores = better quality). L.V.: Laplacian Variance; E.D.: Edge Density.

3.3. Validation on the External Test Sets

We first evaluated our metric's generalization capabilities on the PH² dataset, containing 200 dermoscopic images acquired with professional equipment. Our team of dermatologists (F.V., E.Z., and P.S.) confirmed through visual inspection that these images consistently exhibited high and uniform quality, comparable to the HQ category of our ISIC test set.

The performance of DermaIQA on this external dataset aligned well with expert assessment. Figure 8a shows the quality score distributions for the PH² dataset alongside the ISIC test set's HQ and LQ distributions. The PH² images received consistently high scores with a narrow distribution centered at 0.58, positioned between the ISIC HQ (centered at 0.74) and LQ (centered at 0.42) distributions. The PH² distribution's standard deviation of 0.07 was notably smaller than both ISIC distributions (HQ: 0.12, LQ: 0.14), reflecting the uniform high quality of these images.

Figure 8b provides a visual validation through representative examples. We show a PH² image and an ISIC HQ image that received identical quality scores (0.64), demonstrating their comparable visual quality. For contrast, we included an ISIC LQ image that received a similarly high score (0.63), despite most LQ images scoring around 0.3. Magnified sections highlight the similar visual characteristics that led to comparable scores, confirming the metric's consistency across different image sources.

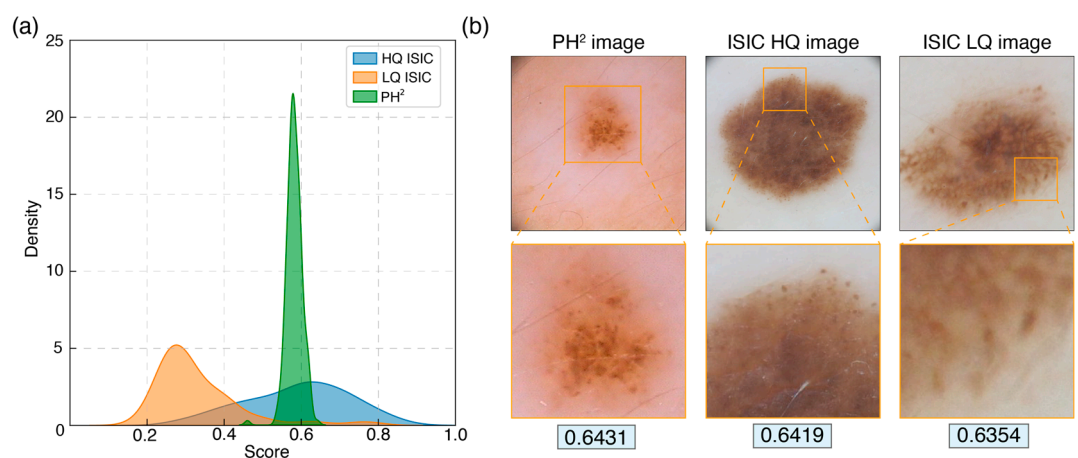


Figure 8. Validation of the proposed IQA metric on the PH2 Dataset. (a) Distribution of quality scores for the PH2 dataset compared to the ISIC testing set's high-quality (HQ) and low-quality (LQ) distributions. (b) Visual comparison of three dermoscopic images: one from the PH2 dataset, one HQ image from the ISIC test set, and one LQ image from the ISIC test set.

After validating DermaIQA on professional dermoscopic images, we extended our evaluation to smartphone-based dermoscopy datasets (Figure 9). From the original PAD-UFES-20 dataset of 2298 images, we processed 2044 images that met our minimum input size requirement (384×384 pixels). This subset shows a broad distribution of quality scores ranging from 0.226 to 0.841, with a mean score of 0.524. The Novara dataset, despite using a specific smartphone dermoscope (Nurugo), also exhibits significant quality variations, with scores ranging from 0.303 to 0.749 and a mean of 0.625. These distributions demonstrate how smartphone-based acquisition introduces greater variability compared to professional dermoscopic imaging (PH2 dataset). Figure 9 illustrates this variability through examples of high- and low-scoring images from each dataset, highlighting how

DermaIQA captures quality differences across different smartphone-based acquisition systems.

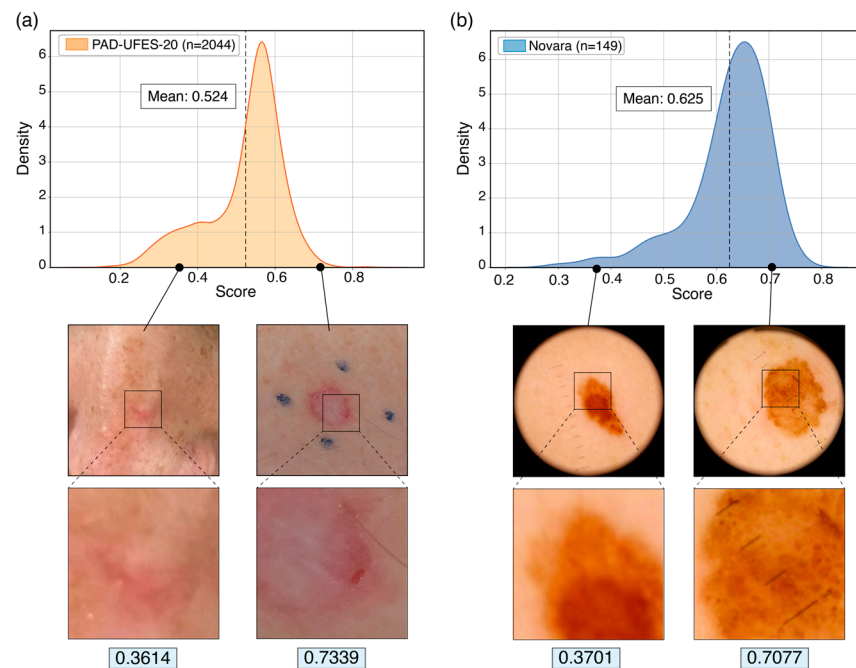


Figure 9. Evaluation of smartphone-based dermoscopy datasets. (a) Distribution of quality scores for the PAD-UFES-20 dataset with examples of low and high-scoring images (magnified sections shown). (b) Distribution of quality scores for the Novara dataset with representative low and high-quality examples (magnified sections shown). The wide range of scores in both datasets reflects the quality variability inherent in smartphone-based dermoscopic imaging.

4. Discussion

Reliable image quality assessment is crucial in dermatological imaging, yet to the best of our knowledge, no specialized no-reference IQA metric has been developed for dermoscopic images. Our research introduces both a novel IQA metric specifically designed for dermoscopic images and a methodological framework that can be extended to other medical imaging domains.

The key innovation in our framework lies in our approach to training data preparation and the adaptation of the Elo rating system for generating quality rankings. Unlike several IQA metrics that require thousands of human-provided scores, our methodology generates comprehensive quality comparisons through an iterative process requiring minimal human intervention. Starting with just 539 high-quality and 273 low-quality images classified by dermatologists, we developed a custom degradation pipeline that creates controlled quality variations among the images in our dataset. This combination of synthetic degradation and Elo-based ranking enabled us to create a comprehensive training dataset with over 125,000 images that closely approximates human perception. The approach reduced our data requirements from potentially millions of human evaluations to just over 800 binary classifications, representing orders of magnitude improvement in annotation efficiency.

The effectiveness of domain-specific training is particularly evident when comparing our results with existing approaches. While traditional heuristic methods and state-of-the-art no-reference IQA metrics achieved accuracies between 72% and 85% in our pairwise evaluations, DermaIQA reached 92% accuracy.

From a computational perspective, DermaIQA operates efficiently, processing full dermoscopic images in less than one second on standard GPU hardware. This near real-

time performance makes it suitable for integration into clinical workflows and AI pipelines for dermatological image analysis [47,48]. While we implemented our framework using an existing transformer-based architecture, our primary contribution lies in the novel combination of synthetic degradation with Elo-based quality ranking—an approach that could be applied with various neural network architectures.

Despite these promising results, our current implementation has several limitations that suggest directions for future improvement. First, our degradation pipeline relies on empirically determined probability parameters (0.5 for initial blur selection and 0.8 for JPEG compression) to create a diverse quality spectrum. While these values proved effective, a systematic optimization of these parameters could potentially enhance the method's performance. Furthermore, our patch-based approach uses simple arithmetic mean for aggregating local quality scores, treating all image regions equally. In clinical practice, however, the lesion itself is the primary focus of diagnostic attention [49]. This limitation becomes particularly relevant in cases where critical diagnostic features (such as border irregularities, color variations, or dermoscopic structures) are concentrated in specific regions of the image. Integrating a lesion segmentation network to weight patch scores based on their diagnostic relevance could provide more clinically meaningful quality assessments. Such an approach could assign higher importance to patches containing the lesion boundary or distinctive dermoscopic patterns, better aligning the quality assessment with clinical requirements. Additionally, our overlapping patch strategy could be leveraged to generate saliency heatmaps, providing dermatologists with visual feedback on spatial quality variations and guiding image acquisition improvements.

As shown in Figure 4, we observed some overlap between LQ and HQ score distributions, with occasional cases where LQ lesions received higher scores than HQ ones (e.g., Figure 5b). This phenomenon typically occurs when evaluating images with different lesion-to-background ratios or varying image dimensions. The simple arithmetic mean aggregation currently employed may not fully capture these complexities, particularly when critical diagnostic features are concentrated in specific regions. The decision to minimize expert supervision requirements led us to rely on quality assessment from a single dermatologist, which may have introduced inherent biases in the labeling process (i.e., the definition of LQ and HQ image) that propagate through the network's learning process. Future implementations would benefit from a multi-expert consensus approach for quality labeling, potentially reducing these classification inconsistencies.

Further developments could also explore the relationship between image resolution and quality scores, potentially establishing minimum size requirements for reliable quality assessment. Expanding validation on other datasets would enhance generalization capabilities, particularly important given the growing adoption of mobile dermoscopy in clinical practice.

While our degradation pipeline was specifically designed for dermoscopic imaging, the underlying framework could be extended to other medical imaging domains where artifacts can be reliably simulated. For example, in digital pathology, scanning artifacts such as blur, compression, and color shifts similarly affect diagnostic quality and could be quantitatively modeled. This approach offers significant advantages over methods requiring extensive manual scoring [21,22].

5. Conclusions

In this work, we have developed DermaQA, a novel no-reference image quality assessment metric specifically tailored for dermoscopic images. Our approach demonstrates that domain-specific quality assessment substantially outperforms general-purpose metrics in specialized medical imaging contexts. By achieving 92% accuracy in distinguishing between high- and low-quality images and demonstrating robust generalization across

both professional and smartphone-based dermoscopy datasets, our metric provides a reliable tool for quality control in dermatological workflows. The methodological framework, combining limited expert assessment with synthetic degradation and Elo-based quality ranking, dramatically reduces manual labeling requirements while maintaining high performance. DermaIQA could enhance both clinical workflows and AI-based diagnostic systems by ensuring only high-quality images enter diagnostic pipelines. Future work will focus on generating quality heatmaps for visual feedback, prioritizing lesion regions through integration with automated segmentation, and extending the framework to additional medical imaging modalities.

Author Contributions: Conceptualization, M.S.; Data Curation, A.F. and F.B.; Formal Analysis, A.F., F.B., F.V., E.Z. and P.S.; Investigation, F.V., E.Z. and P.S.; Methodology, A.F. and F.B.; Resources, K.M.M. and M.S.; Software, A.F.; Supervision, M.S.; Validation, F.B., F.V., E.Z. and P.S.; Visualization, A.F. and F.B.; Writing—Original Draft, F.B. and M.S.; Writing—Review and Editing, A.F., K.M.M., F.V., E.Z. and P.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (Comitato Etico Internaziendale) (protocol 173/18, approved on 12 December 2018).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study

Data Availability Statement: The trained DermalQA model and inference scripts are publicly available at DOI: 10.17632/2ryw3hpb6v.1.

Acknowledgments: ChatGPT (GPT-5.2) was used only to improve grammar and clarity of the English language. No content was generated by the AI. All text was carefully reviewed and validated by the author, who is fully responsible for the final version.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|------|--|
| AI | Artificial Intelligence |
| AUC | Area Under the ROC Curve |
| CNR | Contrast-to-Noise Ratio |
| DCT | Discrete Cosine Transform |
| HQ | High-Quality |
| IQA | Image Quality Assessment |
| LQ | Low-Quality |
| MSE | Mean Squared Error |
| PLCC | Pearson Linear Correlation Coefficient |
| PSNR | Peak Signal-to-Noise Ratio |
| ROI | Region Of Interest |
| SRCC | Spearman Rank Correlation Coefficient |
| SSIM | Structural Similarity Index Metric |
| ViT | Vision Transformer |

References

1. Celebi, M.E.; Codella, N.; Halpern, A. Dermoscopy Image Analysis: Overview and Future Directions. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 474–478. <https://doi.org/10.1109/JBHI.2019.2895803>.

2. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature* **2017**, *542*, 115–118. <https://doi.org/10.1038/nature21056>.
3. Gessert, N.; Sentker, T.; Madesta, F.; Schmitz, R.; Kniep, H.; Baltruschat, I.; Werner, R.; Schlaefel, A. Skin Lesion Classification Using CNNs with Patch-Based Attention and Diagnosis-Guided Loss Weighting. *IEEE Trans. Biomed. Eng.* **2020**, *67*, 495–503. <https://doi.org/10.1109/TBME.2019.2915839>.
4. Tofighi, N.J.; Elfkir, M.H.; Imamoglu, N.; Ozcinar, C.; Erdem, A.; Erdem, E. Omnidirectional Image Quality Assessment with Local–Global Vision Transformers. *Image Vis. Comput.* **2024**, *148*, 105151. <https://doi.org/10.1016/j.imavis.2024.105151>.
5. Seoni, S.; Shahini, A.; Meiburger, K.M.; Marzola, F.; Rotunno, G.; Acharya, U.R.; Molinari, F.; Salvi, M. All You Need Is Data Preparation: A Systematic Review of Image Harmonization Techniques in Multi-Center/Device Studies for Medical Support Systems. *Comput. Methods Programs Biomed.* **2024**, *250*, 108200. <https://doi.org/10.1016/j.cmpb.2024.108200>.
6. Breger, A.; Karner, C.; Selby, I.; Gröhl, J.; Dittmer, S.; Lilley, E.; Babar, J.; Beckford, J.; Else, T.R.; Sadler, T.J.; et al. A Study on the Adequacy of Common IQA Measures for Medical Images. In *Proceedings of 2024 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2024), Manchester, UK, 19-21 November 2024*; Su, R., Frangi, A.F., Zhang, Y., Eds.; Lecture Notes in Electrical Engineering; Springer Nature: Singapore, 2025; Volume 1372, pp. 451–462.
7. Breger, A.; Biguri, A.; Landman, M.S.; Selby, I.; Amberg, N.; Brunner, E.; Gröhl, J.; Hatamikia, S.; Karner, C.; Ning, L.; et al. A Study of Why We Need to Reassess Full Reference Image Quality Assessment with Medical Images. *J. Imaging Inform. Med.* **2025**, *38*, 3444–3469. <https://doi.org/10.1007/s10278-025-01462-1>.
8. Yu, S.; Dai, G.; Wang, Z.; Li, L.; Wei, X.; Xie, Y. A Consistency Evaluation of Signal-to-Noise Ratio in the Quality Assessment of Human Brain Magnetic Resonance Images. *BMC Med. Imaging* **2018**, *18*, 17. <https://doi.org/10.1186/s12880-018-0256-6>.
9. Fei, X.; Xiao, L.; Sun, Y.; Wei, Z. Perceptual Image Quality Assessment Based on Structural Similarity and Visual Masking. *Signal Process. Image Commun.* **2012**, *27*, 772–783. <https://doi.org/10.1016/j.image.2012.04.005>.
10. Adhikari, A.; Lee, S.-W. AM-BQA: Enhancing Blind Image Quality Assessment Using Attention Retractable Features and Multi-Dimensional Learning. *Image Vis. Comput.* **2024**, *147*, 105076. <https://doi.org/10.1016/j.imavis.2024.105076>.
11. Chow, L.S.; Paramesran, R. Review of Medical Image Quality Assessment. *Biomed. Signal Process. Control* **2016**, *27*, 145–154. <https://doi.org/10.1016/j.bspc.2016.02.006>.
12. Hosny, K.M.; Kassem, M.A.; Foad, M.M. Skin Cancer Classification Using Deep Learning and Transfer Learning. In *Proceedings of the 2018 9th Cairo International Biomedical Engineering Conference (CIBEC), Cairo, Egypt, 20–22 December 2018*; IEEE: New York, NY, USA, 2018; pp. 90–93.
13. Veronese, F.; Tarantino, V.; Zavattaro, E.; Biacchi, F.; Airoidi, C.; Salvi, M.; Seoni, S.; Branciforti, F.; Meiburger, K.M.; Savoia, P. Teledermoscopy in the Diagnosis of Melanocytic and Non-Melanocytic Skin Lesions: Nurugo™ Derma Smartphone Microscope as a Possible New Tool in Daily Clinical Practice. *Diagnostics* **2022**, *12*, 1371. <https://doi.org/10.3390/diagnostics12061371>.
14. Yang, S.; Wu, T.; Shi, S.; Lao, S.; Gong, Y.; Cao, M.; Wang, J.; Yang, Y. MANIQA: Multi-dimension Attention Network for No-Reference Image Quality Assessment, Github Repository. Available online: <https://github.com/IIGROUP/MANIQA> (accessed on 3 February 2026).
15. Ke, J.; Wang, Q.; Wang, Y.; Milanfar, P.; Yang, F. Musiq: Multi-Scale Image Quality Transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021*; IEEE: New York, NY, USA, 2021; pp. 5148–5157.
16. Su, S.; Yan, Q.; Zhu, Y.; Zhang, C.; Ge, X.; Sun, J.; Zhang, Y. Blindly Assess Image Quality in the Wild Guided by a Self-Adaptive Hyper Network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Waikoloa, HI, USA, 3–8 January 2022*; IEEE: New York, NY, USA, 2020; pp. 3667–3676.
17. Golestaneh, S.A.; Dadsetan, S.; Kitani, K.M. No-Reference Image Quality Assessment via Transformers, Relative Ranking, and Self-Consistency. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2022*; IEEE: New York, NY, USA, 2022; pp. 1220–1230.
18. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a “Completely Blind” Image Quality Analyzer. *IEEE Signal Process. Lett.* **2013**, *20*, 209–212. <https://doi.org/10.1109/LSP.2012.2227726>.
19. Zhang, L.; Zhang, L.; Bovik, A.C. A Feature-Enriched Completely Blind Image Quality Evaluator. *IEEE Trans. Image Process.* **2015**, *24*, 2579–2591. <https://doi.org/10.1109/TIP.2015.2426416>.
20. Rodrigues, R.; Lévêque, L.; Gutiérrez, J.; Jebbari, H.; Outtas, M.; Zhang, L.; Chetouani, A.; Al-Juboori, S.; Martini, M.G.; Pinheiro, A.M.G. Objective Quality Assessment of Medical Images and Videos: Review and Challenges. *Multimed. Tools Appl.* **2024**, *84*, 29915–29948. <https://doi.org/10.1007/s11042-024-20292-x>.

21. Chen, B.; Solebo, A.L.; Bao, W.; Taylor, P. Medical Image Quality Assessment Based on Probability of Necessity and Sufficiency. In *Proceedings of the 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI), Houston, TX, USA, 14–17 April 2025*; IEEE: New York, NY, USA, 2025; pp. 1–4.
22. Chen, J.; Lin, L.; Cheng, P.; Huang, Y.; Tang, X. Uno-Qa: An Unsupervised Anomaly-Aware Framework with Test-Time Clustering for Octa Image Quality Assessment. In *Proceedings of the 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), Cartagena, Colombia, 18–21 April 2023*; IEEE: New York, NY, USA, 2023; pp. 1–5.
23. Petashvili, D.; Wang, H.; Deslandes, A.; Avery, J.; Condous, G.; Carneiro, G.; Hull, L.; Chen, H.-T. Learning Subjective Image Quality Assessment for Transvaginal Ultrasound Scans from Multi-Annotator Labels. In *Proceedings of the 2024 IEEE International Symposium on Biomedical Imaging (ISBI), Athens, Greece, 27–30 May 2024*; IEEE: New York, NY, USA, 2024; pp. 1–5.
24. Maruyama, S.; Watanabe, H.; Shimosegawa, M. An Image Quality Assessment Index Based on Image Features and Keypoints for X-Ray CT Images. *PLoS ONE* **2024**, *19*, e0304860. <https://doi.org/10.1371/journal.pone.0304860>.
25. Ozer, C.; Guler, A.; Cansever, A.T.; Oksuz, I. Explainable Image Quality Assessment for Medical Imaging. *arXiv* **2023**, arXiv:2303.14479.
26. Nikiforaki, K.; Karatzanis, I.; Dovrou, A.; Bobowicz, M.; Gwozdziewicz, K.; Díaz, O.; Tsiknakis, M.; Fotiadis, D.I.; Lekadir, K.; Marias, K. Image Quality Assessment Tool for Conventional and Dynamic Magnetic Resonance Imaging Acquisitions. *J. Imaging* **2024**, *10*, 115. <https://doi.org/10.3390/jimaging10050115>.
27. Elo, A.E. *The Rating of Chessplayers, Past and Present*, 2nd ed.; Arco Publishing Company, Inc.: New York, NY, USA, 1986.
28. Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions. *Sci. Data* **2018**, *5*, 180161. <https://doi.org/10.1038/sdata.2018.161>.
29. Codella, N.C.F.; 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 *Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018*; IEEE: New York, NY, USA, 2018; pp. 168–172.
30. Mendonca, T.; Ferreira, P.M.; Marques, J.S.; Marcal, A.R.S.; Rozeira, J. PH²—A Dermoscopic Image Database for Research and Benchmarking. In *Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013*; IEEE: New York, NY, USA, 2013; pp. 5437–5440.
31. Oztel, I.; Yolcu Oztel, G.; Sahin, V.H. Deep Learning-Based Skin Diseases Classification Using Smartphones. *Adv. Intell. Syst.* **2023**, *5*, 2300211. <https://doi.org/10.1002/aisy.202300211>.
32. Buslaev, A.; Iglovikov, V.I.; Khvedchenya, E.; Parinov, A.; Druzhinin, M.; Kalinin, A.A. Albuementations: Fast and Flexible Image Augmentations. *Information* **2020**, *11*, 125. <https://doi.org/10.3390/info11020125>.
33. Levin, A.; Weiss, Y.; Durand, F.; Freeman, W.T. Understanding and Evaluating Blind Deconvolution Algorithms. In *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009*; IEEE: New York, NY, USA, 2009; pp. 1964–1971.
34. Al-Ani, M.S.; Awad, F.H. The JPEG Image Compression Algorithm. *Int J Adv Eng Technol* **2013**, *6*, 1055–1062.
35. Ahmed, N.; Natarajan, T.; Rao, K.R. Discrete Cosine Transform. *IEEE Trans. Comput.* **1974**, *C-23*, 90–93. <https://doi.org/10.1109/T-C.1974.223784>.
36. Tomar, R.R.S.; Jain, K. Lossless Image Compression Using Differential Pulse Code Modulation and Its Application. In *Proceedings of the 2015 International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, India, 12–14 December 2015*; IEEE: New York, NY, USA, 2015; pp. 397–400.
37. Moffat, A. Huffman Coding. *ACM Comput. Surv.* **2020**, *52*, 1–35. <https://doi.org/10.1145/3342555>.
38. Weiyuan, W.; Verma, D.; Yang, W. Patchify Github Repository. GitHub Repository. Available online: <https://github.com/dovahcrow/patchify.py> (accessed on 3 February 2026).
39. Huynh-Thu, Q.; Ghanbari, M. Scope of Validity of PSNR in Image/Video Quality Assessment. *Electron. Lett.* **2008**, *44*, 800–801. <https://doi.org/10.1049/el:20080522>.
40. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021*; IEEE: New York, NY, USA, 2021; pp. 10012–10022.
41. Zhang, W.; Ma, K.; Zhai, G.; Yang, X. Uncertainty-Aware Blind Image Quality Assessment in the Laboratory and Wild. *IEEE Trans. Image Process.* **2021**, *30*, 3474–3486. <https://doi.org/10.1109/TIP.2021.3061932>.

42. Chaofeng, C- and Jiadi M. IQA-PyTorch: PyTorch Toolbox for Image Quality Assessment, GitHub Repository. Available online: <https://github.com/chaofengc/IQA-PyTorch> (accessed on 3 February 2026).
43. Krotkov, E. Focusing. *Int. J. Comput. Vis.* **1988**, *1*, 223–237. <https://doi.org/10.1007/BF00127822>.
44. Pech-Pacheco, J.L.; Cristobal, G.; Chamorro-Martinez, J.; Fernandez-Valdivia, J. Diatom Autofocusing in Brightfield Microscopy: A Comparative Study. In *Proceedings of the 15th International Conference on Pattern Recognition. ICPR-2000, Barcelona, Spain, 3–7 September 2000*; IEEE: New York, NY, USA, 2000; Volume 3, pp. 314–317.
45. Liu, L.; Liu, B.; Huang, H.; Bovik, A.C. No-Reference Image Quality Assessment Based on Spatial and Spectral Entropies. *Signal Process. Image Commun.* **2014**, *29*, 856–863. <https://doi.org/10.1016/j.image.2014.06.006>.
46. Al-refai, G.; Qadan, S.; Al-refai, M.; Elmoaqet, H. Introducing Edges Density and Texture Contrast Scores as Quality Metrics for Enhanced Low Light Images. In *Proceedings of the 2024 22nd International Conference on Research and Education in Mechatronics (REM), Amman, Jordan, 24–26 September 2024*; IEEE: New York, NY, USA, 2024; pp. 1–5.
47. Salvi, M.; Branciforti, F.; Molinari, F.; Meiburger, K.M. Generative Models for Color Normalization in Digital Pathology and Dermatology: Advancing the Learning Paradigm. *Expert Syst. Appl.* **2024**, *245*, 123105. <https://doi.org/10.1016/j.eswa.2023.123105>.
48. Salvi, M.; Branciforti, F.; Veronese, F.; Zavattaro, E.; Tarantino, V.; Savoia, P.; Meiburger, K.M. DermoCC-GAN: A New Approach for Standardizing Dermatological Images Using Generative Adversarial Networks. *Comput. Methods Programs Biomed.* **2022**, *225*, 107040. <https://doi.org/10.1016/j.cmpb.2022.107040>.
49. Yuan, Y.; Chao, M.; Lo, Y.-C. Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks With Jaccard Distance. *IEEE Trans. Med. Imaging* **2017**, *36*, 1876–1886. <https://doi.org/10.1109/TMI.2017.2695227>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.