

Enhanced Dermatology Image Quality Assessment via Cross-Domain Training

Ignacio Hernández Montilla
Legit.Health
Bilbao, Spain
University of Deusto
Bilbao, Spain
ignaciohernandez@legit.health

Alfonso Medela
Legit.Health
Bilbao, Spain
alfonso@legit.health

Paola Pasquali
Pius Hospital de Valls
Tarragona, Spain
pasqualipaola@gmail.com

Andy Aguilar
Legit.Health
Bilbao, Spain
andy@legit.health

Taig Mac Carthy
Legit.Health
Bilbao, Spain
taig@legit.health

Gerardo Fernández
Legit.Health
Bilbao, Spain
gerardo@legit.health

Antonio Martorell
Legit.Health
Bilbao, Spain
Hospital de Manises
Valencia, Spain
antonio@legit.health

Enrique Onieva
University of Deusto
Bilbao, Spain
enrique.onieva@deusto.es

Abstract

Teledermatology has become a widely accepted communication method in daily clinical practice, enabling remote care while showing strong agreement with in-person visits. Poor image quality remains an unsolved problem in teledermatology and is a major concern to practitioners, as bad-quality images reduce the usefulness of the remote consultation process. However, research on Image Quality Assessment (IQA) in dermatology is sparse, and does not leverage the latest advances in non-dermatology IQA, such as using larger image databases with ratings from large groups of human observers. In this work, we propose cross-domain training of IQA models, combining dermatology and non-dermatology IQA datasets. For this purpose, we created a novel dermatology IQA database, Legit.Health-DIQA-Artificial, using dermatology images from several sources and having them annotated by a group of human observers. We demonstrate that cross-domain training yields optimal performance across domains and overcomes one of the biggest limitations in dermatology IQA, which is the small scale of data, and leads to models trained on a larger pool of image distortions, resulting in a better management of image quality in the teledermatology process.

CCS Concepts

• **Computing methodologies** → **Artificial intelligence**.



This work is licensed under a Creative Commons Attribution 4.0 International License.
ICBRA 2025, Prague, Czech Republic
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1580-8/25/09
<https://doi.org/10.1145/3774976.3774977>

Keywords

Image Quality Assessment, Teledermatology, Computer Vision, Dermatology

ACM Reference Format:

Ignacio Hernández Montilla, Alfonso Medela, Paola Pasquali, Andy Aguilar, Taig Mac Carthy, Gerardo Fernández, Antonio Martorell, and Enrique Onieva. 2025. Enhanced Dermatology Image Quality Assessment via Cross-Domain Training. In *The 12th International Conference on Bioinformatics Research and Applications (ICBRA 2025), September 19–21, 2025, Prague, Czech Republic*. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3774976.3774977>

1 Introduction

Over the last years, the use of teledermatology has become widely accepted, especially after the SARS-CoV-2 (COVID-19) pandemic. From its global validation [49], dermatology has benefited from the advancements in remote care, and it has become a broadly accepted and useful communication method between general practitioners and dermatologists [12][25].

There are three main types of teledermatology: real-time, or synchronous; store-and-forward, or asynchronous; and hybrid, which combines real-time and store-and-forward. The most utilized method is store-and-forward, in which pictures can be taken either by the healthcare practitioner or the patients themselves. This method has reportedly been considered useful by dermatologists, showing agreement with practitioners that perform conventional in-person visits [20]. In store-and-forward dermatology, patients with an inconclusive diagnosis or who require an in-person appointment or referral to a specialized consultation are contacted more quickly, resulting in a reduction of waiting lists.

Dermatology relies on visual inspection to diagnose conditions affecting the skin, hair, nails, and mucous membranes, which makes

digital photography an essential tool for many purposes. These include creating a visual record of the disease in patient’s history, showing evidence of treatment efficacy, keeping track of changes in a specific lesion, receiving second opinion other practitioners, or educational and scientific purposes [36]. The heavy use of image data in dermatology explains why the emergence of dermatology artificial intelligence (AI) has mostly occurred in the field of computer vision. Thanks to the current skin image databases, AI has been successfully applied to a variety of scenarios such as skin cancer screening, automatic severity assessment [29][14][27], and skin pathology recognition [15].

Among other technical limitations, teledermatology is subject to poor image quality, which can limit the effectiveness of consultations as low-quality images may oblige practitioners to ask for new pictures and can eventually lead to referrals to live consultations, unnecessarily increasing the patient’s time to receive a diagnosis [35]. Image quality issues also pose a challenge for novel AI-powered image-based diagnostic support tools, which rely on the correct visual content of an image to provide accurate predictions [28].

Image Quality Assessment (IQA) is the area of research that develops quality measures to detect distorted and low-quality images. Such measures which can be broadly divided into three approaches. The most classical IQA methods require the original image and the distorted version to make a comparison, which is called full-reference IQA (FR-IQA). Reduced-reference IQA (RR-IQA) measures do not require the original or "pristine" image, but assume that some partial information of it is available. Finally, no-reference or blind IQA measures (NR-IQA) rely exclusively on the distorted image, with no additional information [33]. For each approach, the measures can be either hand-crafted, such as BM [9], SSIM, NIQE, BLI-INDS, DIIVINE or BRISQUE [31], or learned by a machine learning model such as HOSA [54] or DeepBIQ [6].

To the best of our knowledge, the research on dermatology IQA is sparse and always has been in the no-reference domain due to the predominance of the store-and-forward teledermatology model. There have been recent works that support the use of deep learning models for blind image visual quality assessment, such as TrueImage [52], ImageQX [18], and DIQA [32]. In all cases, the development of domain-specific image datasets and the annotation of IQA labels was required, which is time-consuming.

This raises some questions: *given the plethora of natural IQA datasets available, is it possible to train an IQA model on natural, everyday image quality datasets and use it in the dermatology domain? To what extent is it required to develop domain-specific datasets for dermatology image quality assessment? Can dermatology IQA models benefit from adding image quality databases outside the dermatology domain? Can a single model perform well across all domains?*

In our preliminary work [32], we demonstrated the potential of deep learning models as a tool to evaluate dermatology images in a store-and-forward teledermatology scenario, using a combination of natural and dermatology IQA datasets. However, some aspects remained unexplored, such as the performance of models trained separately on each natural IQA dataset with no additional dermatological fine-tuning, and the performance at higher resolutions.

In the present work, we present a study of model performance on dermatology IQA comparing the usage of natural image datasets

with no additional dermatological data to the combination of natural and dermatology images for training. We began by training several deep learning models on natural image datasets and evaluating their performance on a new dermatology IQA dataset to discern whether it is possible to use already existing IQA databases without depending on dermatological data. As part of this analysis, we also explored the relationship between input image resolution and model performance: the importance of choosing the right resolution for a deep learning-based IQA model has been previously reported [16], and we aimed to extrapolate it to the dermatology domain. Finally, we show the benefits of cross-domain training of IQA models: training simultaneously on many IQA datasets, including dermatology images, provides consistent performance across all natural IQA datasets and improves the performance of dermatology IQA models, as they are trained on a larger pool of image distortions than the ones that could be captured in a smaller, dermatology-only image dataset.

2 Related work

2.1 Image quality control in telemedicine

There has been an increase in teledermatology guidelines during the last years, many of them including recommendations for dermatological photography [7]. Despite the availability of this material, image quality is still perceived as the most crucial disadvantage in teledermatology [39]. Fortunately, thanks to the widespread adoption of digital photography, together with the rapid evolution of hardware, most cameras can fulfill the requirements for good dermatological pictures, which bridges the gap between professionals with and without experience in photography and contributes to a better overall image quality [36].

However, while healthcare practitioners have shown to be proficient at dermatological photography [8], patients may struggle at taking pictures. In fact, the review of guidelines conducted in [7] reveals that dermatological photography recommendations are aimed at healthcare practitioners. This excludes patients from learning how to take good photographs, despite they have become an important player in the clinical process, specially after the COVID-19 pandemic [48].

With the widespread adoption of smartphone photography in everyday life, patients are likely to use these devices for teledermatology consultations; although specifications are continuously improving, these devices are not designed with the purpose of dermatology imaging [2]. In addition, patients are most likely to take pictures at their homes, which are non-standardized environments, in non-standardized ways (such as using the front camera of a smartphone), increasing the chances of generating low-quality image data. To overcome these limitations, it is possible to analyze the most critical aspects of medical photography and define some basic principles that clinicians and patients should understand to take good quality pictures, even with a smartphone [57]. Educating patients for optimal skin photography poses an unsolved challenge, though, as even with proper training they may fail to generate good-quality images [17]. This knowledge gap can even be widened in some populations due to the generational digital divide, as patients of old age are less familiarized with healthcare technology and apps [24].

2.2 IQA methods in dermatology

The effect of low-quality images on the performance of dermatology AI models has been previously explored by Maier et al. [28]. Although it is a rather preliminary work limited by their model architecture and distortion choices, it points out that, even when the model is trained with data augmentation to increase robustness, there will always be the risk of encountering either extremely low-quality images or images affected by unexpected distortions, resulting in out-of-domain inputs with inaccurate outputs.

Most research on dermatology IQA has been focused on detecting common artifacts that produce low-quality dermatology pictures. Vodrahalli et al. developed an ensemble of deep learning and classical machine learning models to assess image quality via binary (good/bad) assessments on overall quality, blur, lighting and zoom, but they used a rather small input image size (between 128^2 and 256^2 px²) and obtained mixed results [52]. Similarly, Jalaboi et al. trained a deep learning model to detect common causes of poor quality (framing, lighting, blur, resolution, and distance) with equally mixed results, even though its performance was comparable to that of their dermatologists and used a much larger dataset [18]. The main disadvantage of both works is that they were designed for explaining what is causing the image to be unsuitable for teledermatology consultations, and did not follow the standardized approach in IQA research, which is rating images using absolute category rating (ACR) scales and measuring correlation between objective scores (the model’s predicted quality score) and true subjective scores.

Our previous work [32] presented an initial exploration of the usage of non-medical datasets for dermatology IQA, but it lacked a thorough evaluation of every IQA dataset and used a small dermatology image dataset of real distortions (934 images), which may limit models’ ability to learn image distortions.

2.3 IQA databases

The landscape of IQA datasets outside dermatology IQA is continuously expanding with new image databases. Over the last twenty years, many have been created to characterize the adequateness and performance of quality metrics. The earliest dataset released were IVC [3], LIVE [41], TID2008 [38], and CSIQ [19], which consisted of small selections of pristine images to which several distortions were applied. Later on, the TID2008 was later expanded to TID2013 [37], followed by the CID:IQ dataset [23]. The CID2013 dataset proposed by Virtanen et al. [51] introduced the novelty of including pictures with concurrent distortions taken with different imaging devices. Ghadiyaram et al. introduced another dataset with authentic distortions, LIVE in the Wild (LIVE-ItW)[11], which has become one of the most utilized datasets for benchmarking IQA methods. Their main limitation of these datasets, however, is their small size: as annotating subjective quality scores require large boards of observers, it is a costly and time-consuming process. A more recent dataset, NITS-IQA, provides a larger set of pristine images, but is still limited to a small sample of artificial distortions [40].

Recent datasets have addressed this size issue by providing a larger sets of images: Kadid-10k and KADIS-700k [21] provide a large collection of artificially distorted images, while KonIQ-10k

Table 1: Summary of natural IQA datasets used in this work

Dataset	Number of images	Score range	Distortion type
KonIQ-10k	10073	[1, 5]	Authentic
GFIQA-20k	20000	[0, 1]	Authentic
SPAQ	11125	[0, 100]	Authentic
BIQ2021	8000	[0, 1]	Authentic
LIVE-ItW	1169	[0, 100]	Authentic
Kadid-10k	10125	[1, 5]	Artificial

[16], BIQ2021 [1], and SPAQ [10] offer images with naturally occurring distortions.

3 Materials and methods

3.1 Natural IQA datasets

After reviewing the current landscape of IQA datasets, we selected Kadid-10k [21], KonIQ-10k [16], GFIQA-20k [45], SPAQ [10], BIQ2021 [1], and LIVE In the Wild [11] for our experiments. Kadid-10k, KonIQ-10k, GFIQA-20k, SPAQ and BIQ2021 comprise several thousands of images of common everyday scenarios, which made them suitable for our goal of training deep learning models.

The last dataset, LIVE In the Wild, is one of the most common IQA datasets for benchmarking IQA measures and models. Despite its much smaller size, we included it in our study as a separate test set to assess the performance of the models and provide results that are easy to compare to existing literature. A summary of the natural datasets used in this study is presented in Table 1.

3.2 Dermatology IQA dataset

3.2.1 Data collection. For the evaluation of our IQA models on dermatology image quality assessment (DIQA), we created a new dataset, named Legit.Health-DIQA-Artificial. This dataset was generated by selecting 100 pristine dermatology images and applying a set of artificial distortions (Figure 1). Each distortion was applied with different levels of strength, resulting in 1800 additional images in total. The distortions were inspired by that used for Kadid-10k [21]: JPEG compression (3 levels), Gaussian blur (1), pixelation (2), sharpening (3), brightness (4), color (2), and contrast (3).

To create the dataset, we started by collecting images from well-known dermatology atlases and skin image websites (Atlas Dermatologico, Hellenic Dermatological Atlas, Meddean Dermatology Atlas, Black & Brown Skin). Then, we expanded our search to other skin images datasets, and added PAD-UFES-20 [34], ACNE04 [53], PH2 [30], and SMARTSKINS [50]. We also repurposed two biometric datasets related to skin structure: the Transient Biometric Nails datasets [4][5] and the AMI Ear Database [13]. Finally, we expanded the selection to 100 pristine images using with images from internal sources. For most sources, we only used a small portion of images since image content was very similar, whereas dermatology atlases presented higher variability in terms of image content.

A complete list of the image sources can be found in Table 2. By combining images from different sources, we achieved a high variability in terms of acquisition settings, imaging devices, and



Figure 1: A pristine image (top-left image) of the Legit.Health-DIQA-Artificial dataset and some distorted views. From top left to bottom right: pristine image, JPEG compression, Gaussian blur, pixelation, sharpening, brightness, color, and contrast distortion. Image source: Atlas Dermatologico (<https://www.atlasdermatologico.com.br>).

image content, while ensuring the representation of light and dark skin tones.

Table 2: Summary of sources used to create the Legit.Health-DIQA-Artificial dataset

Source	Images in pristine selection
Atlas Dermatologico	40
Hellenic Dermatological Atlas	13
Black & Brown Skin	4
Meddean Dermatology Atlas	2
PAD-UFES-20	6
ACNE04	4
PH2	3
SMARTSKINS	3
Transient Biometrics Nails	1
AMI Ear Database	1
Internal	23
Total	100

3.2.2 Data annotation. The dataset was exclusively annotated for image quality assessment. Each image was rated with a score in the range $[1, 10]$, where 1 and 10 correspond to poor and excellent perceived visual quality, respectively. Prior to annotation, we resized all images so that the largest side was 512 pixels, making them suitable for any display.

We followed the ITU-T P.910 recommendation from the International Telecommunications Union (ITU) and recruited a large board of 40 non-expert observers. This resulted in 40 quality ratings per image, which were averaged to obtain the mean opinion score (MOS). The distribution of MOS for this dataset is presented in Figure 2.

Although this may be considered a medical image dataset, perceptual image quality assessment should not require prior expert knowledge in dermatology, as quality features exist beyond the semantic content of an image. Although perceptual visual quality can be content dependent [42], we believe that common distortions such as overexposure, underexposure, blur, or compression can be easily detected by any observer in an image, regardless of their medical background.

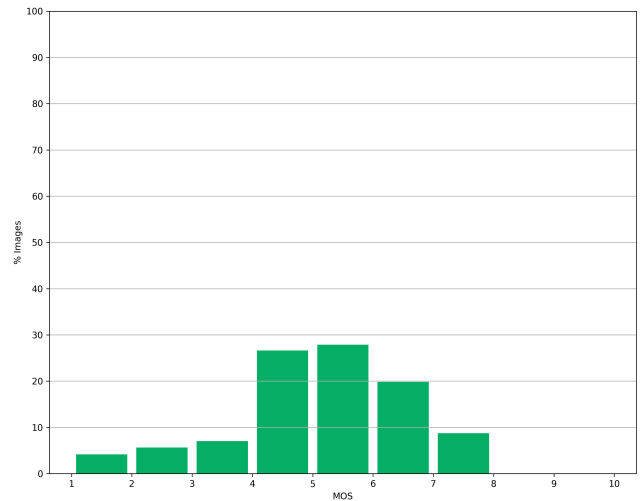


Figure 2: Distribution of the Mean Opinion Scores (MOS) of the Legit.Health-DIQA-Artificial dataset.

3.3 Data preparation

To make all datasets comparable, we scaled the scores of the natural IQA datasets to the $[1, 10]$ range. The distribution of scores for each natural dataset is presented in figure 3. Lastly, we made the training, validation, and test sets for 5 random splits. For the dermatology images, we were able to group the images manually into subjects during the curation of the Legit.Health-DIQA-Artificial dataset. The entire list of subjects was then divided into five training/validation/test splits; this means that if a pristine image is in the training set, all its distorted views will also be put in that set. This stratification strategy prevented data leakage, which may have led to unreliable results.

For the natural images, we applied different splitting strategies depending on the constraints of each dataset. All image clusters of Kadid-10k (i.e. the pristine image and its distorted views) are independent from each other, which made it possible to divide the dataset into five train/validation/test splits; similarly, all the images from Kadid-10k, KonIQ-10k, GFIQA-20k, and SPAQ are independent samples, so we randomly created five training/validation/test splits. On the contrary, BIQ2021 contained image clusters that were already split into a training and validation set by the original authors, which forced us to use this dataset only for training and validation. Lastly, LIVE In the Wild was always included in the test set of every split, as we stated previously.

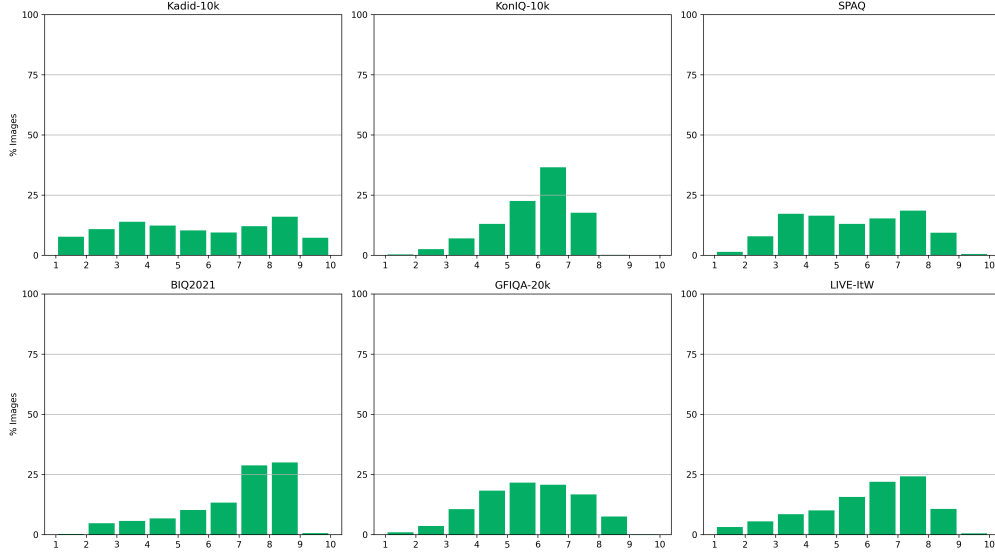


Figure 3: Distribution of Mean Opinion Scores (MOS) for each natural IQA dataset.

3.4 Model training

We conducted all the experiments with convolutional neural networks (CNNs), using the EfficientNet architecture. EfficientNet is a family of deep CNNs designed to be both accurate and efficient, using a compound scaling method that uniformly scales the network’s width, depth, and input resolution with a fixed ratio. From a base network (EfficientNet-B0), the creators applied such scaling method to generate a collection of models with increasing size and accuracy (up to EfficientNet-B7)[47].

In this work, we used three model sizes (B0, B3, and B5) and adjusted all datasets to their corresponding input sizes (224^2 , 288^2 , and 448^2 px², respectively). We used backbones pre-trained on the ImageNet dataset, adding two fully-connected layers (each with a decreasing number of neurons, a dropout rate of 0.50, and separated by a ReLU activation layer) for the prediction of a single scalar value. Model weights were updated at the training stage using the Mean Squared Error loss function (MSE, eq. 1), and validation and test performance was monitored using Pearson’s linear correlation coefficient (PLCC, eq. 2). Similarly to other works, we also computed the Spearman’s rank correlation coefficient (SROCC, eq. 3) for the test set:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

$$PLCC = \frac{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

$$SROCC = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (3)$$

where x and y correspond to the predicted and ground truth scores, respectively, and d_i is the difference between the ranks of predicted and ground truth scores. Both metrics are the most common IQA

metrics for measuring the relationship between objective predicted scores and the true subjective scores.

We trained all models for 20 epochs, using the AdamW optimizer [26] with weight decay $\lambda = 10^{-5}$ and a one-cycle learning rate policy [44]. Our learning rate tests, similar to the one proposed in [43], showed that a learning rate $\gamma = 2 \times 10^{-4}$ was optimal for all experiments.

To overcome the imbalance of quality ratings, we followed a simple heuristic to assign a weight to each sample. First, we rounded every quality score, resulting in a set of 10 possible quality levels $l \in \{1 \dots 10\}$. For a given training dataset \mathcal{D} of images x_i and their corresponding rounded MOS l_i , each class weight w_l was calculated as:

$$w_l = \frac{|\mathcal{D}|}{N \cdot |\{x_i \in \mathcal{D} \mid l_i = l\}|} \quad (4)$$

where $N = 10$ is the number of levels and $|\mathcal{D}|$ is the size of a training dataset \mathcal{D} . Each image was then assigned its corresponding class weight w_l based on its rounded MOS. This method assigned a larger weight to the least represented classes, balancing the training.

Data augmentation is highly restricted in the context of IQA as even the slightest distortion would break the relationship between an image and its MOS label. We followed a similar strategy to the one used to train the NIMA model [46], which consists of resizing followed by random horizontal flipping and a small amount of random cropping. This does not significantly modify the content of the image but produces some additional variability. In our experiments, we set the initial resizing step to make the image 12.5% larger than the expected input size. At validation and test time, the images are directly resized to the model’s input size, followed by a center crop.

For the cross-domain training experiment, we did not require any further processing because the MOS of all datasets were scaled to the same $[1, 10]$ range. With this experiment, we wanted to test

if a model is capable of generalizing to all domains given enough examples of good and bad quality images from very different sources.

4 Results

The aggregated PLCC and SROCC metrics of all our cross-dataset experiments are presented in Table 3 and 4 (Legit.Health-DIQA-Artificial has been shortened to DIQA for clarity). In terms of average correlation across datasets, the best performing model was the EfficientNet-B5 trained on all datasets. Among all models trained on a single natural IQA dataset, the models trained on the SPAQ dataset achieved the highest performance on the dermatology dataset. As we mentioned in the previous section, the BIQ2021 dataset was only used for training and validation; therefore, we could not get any test metrics on this dataset.

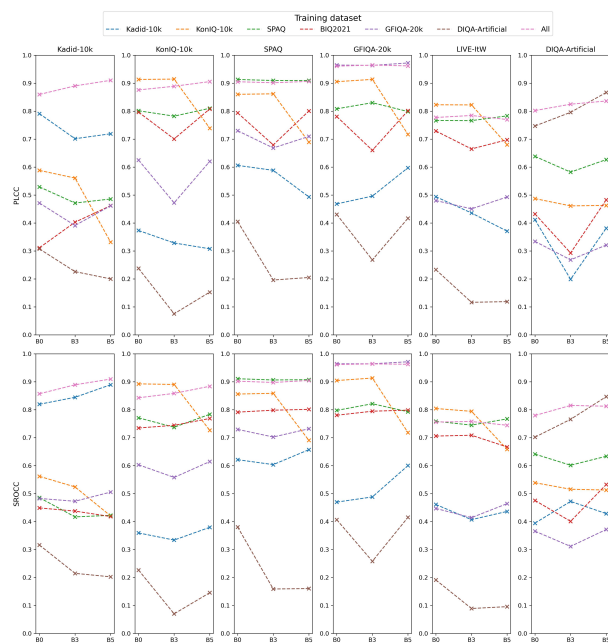


Figure 4: Analysis of the impact of input image size on correlation metrics (PLCC and SROCC).

The effect of using bigger input sizes is presented in Figure 4. Based on these results, we observed that the effect is heavily dataset-dependent, with no clear overall benefit shown when using larger models and single-domain training. However, for cross-domain training, using larger models led to better performance on most datasets.

5 Discussion

Our results are consistent with previous cross-dataset evaluations of IQA models, which showed that IQA models trained on one dataset did not perform well when tested on others [55][22]. We have discovered that this also occurs when cross-testing on medical image datasets. However, our work presents a limitation similar to previous attempts at cross-domain training, which is not being able to perform MOS scaling. Without proper and dataset-specific

scaling, images from different datasets with significantly different quality features may be scaled to a similar MOS, which can reduce the effectiveness of combining IQA datasets [56]. In future developments, we will include images from several IQA datasets to enable MOS scaling for a more precise training and validation, instead of naively rescaling all datasets to [1, 10].

In our single-domain experiments, the models trained on the SPAQ dataset yielded the best individual performance on the Legit.Health-DIQA-Artificial dataset, showing strong correlation (PLCC and SROCC greater than 0.6).

The models trained on the Kadid-10k dataset ranked the lowest positions in terms of correlation on the dermatology dataset; conversely, the models trained on Legit.Health-DIQA-Artificial performed poorly on Kadid-10k. We find these results surprising since they are the only two artificial datasets used in this study, and they contain similar distortions. This may suggest that the models learned the image content instead of the artificial distortions, leading to poor generalization. This may also be happening to the models trained on KonIQ-10k, SPAQ, and GFIQA-20K, as they score the highest correlation metrics in their corresponding test sets.

Another cause of the domain gap may have been the quality rating procedure of each dataset, as perceived visual quality is content dependent to some extent [42]. In dermatology IQA, we suspect that image quality is deeply entangled with visual aesthetics and clinical utility. When rating dermatology images, annotators may be influenced by the underlying purpose of the images, which is to capture a disease or a lesion; inner judgment of how well the image captures such clinical information may be affecting the rating of visual quality features. This behavior would explain why not a single image from the dataset obtained a MOS higher than 8 (Figure 2). Moreover, visual aesthetics of medical images are entirely different from a natural image, which may be adding some difficulty to this task for non-expert observers, even after receiving proper training.

Despite these limitations, we conclude that cross-domain training of an IQA model for dermatology IQA tasks provides better results than training exclusively on dermatology data. In tables 3 and 4, we can observe that the performance of the cross-domain models on the DIQA dataset is superior in most cases, with the exception of the EfficientNet-B5 variants, which may be suffering from overfitting.

In conclusion, we believe that it is not reliable to use non-dermatology IQA models individually for the assessment of teler dermatology images. Instead, models should be trained using a cross-domain approach, including curated dermatology IQA datasets, preferably annotated by dermatologists.

6 Conclusion

In this work, we covered the use of clinical and non-clinical image data to overcome the main limitations of current approaches in dermatology image quality assessment (DIQA), namely limited training data and nonstandard definitions of perceived visual quality assessment. We followed the common approach for IQA development, leveraging not only an already validated methodology but also a variety of natural IQA databases released in recent years.

Table 3: Pearson linear correlation coefficients of the IQA models on all test sets (mean \pm standard deviation).

Model	Dataset	Kadid-10k	KonIQ-10k	SPAQ	GFIQA-20k	LIVE-ItW	DIQA
B0	Kadid-10k	0.7911 \pm 0.0200	0.3733 \pm 0.0567	0.6060 \pm 0.0400	0.4684 \pm 0.0548	0.4939 \pm 0.0162	0.4116 \pm 0.0951
B3	Kadid-10k	0.7014 \pm 0.1682	0.3286 \pm 0.0557	0.5883 \pm 0.0160	0.4962 \pm 0.0469	0.4359 \pm 0.0118	0.1989 \pm 0.2543
B5	Kadid-10k	0.7191 \pm 0.3770	0.3078 \pm 0.2213	0.4933 \pm 0.3058	0.5977 \pm 0.0527	0.3708 \pm 0.2427	0.3813 \pm 0.0602
B0	KonIQ-10k	0.5881 \pm 0.0170	0.9134 \pm 0.0061	0.8602 \pm 0.0038	0.9054 \pm 0.0046	0.8228 \pm 0.0067	0.4872 \pm 0.0518
B3	KonIQ-10k	0.5608 \pm 0.0200	0.9146 \pm 0.0033	0.8620 \pm 0.0048	0.9136 \pm 0.0034	0.8223 \pm 0.0065	0.4613 \pm 0.0415
B5	KonIQ-10k	0.3316 \pm 0.3343	0.7390 \pm 0.4266	0.6894 \pm 0.4016	0.7167 \pm 0.4475	0.6802 \pm 0.3679	0.4629 \pm 0.2244
B0	SPAQ	0.5291 \pm 0.0428	0.8014 \pm 0.0072	0.9133 \pm 0.0022	0.8084 \pm 0.0208	0.7666 \pm 0.0037	0.6383 \pm 0.0461
B3	SPAQ	0.4715 \pm 0.0416	0.7821 \pm 0.0093	0.9095 \pm 0.0015	0.8302 \pm 0.0133	0.7661 \pm 0.0051	0.5819 \pm 0.0336
B5	SPAQ	0.4859 \pm 0.0463	0.8103 \pm 0.0052	0.9097 \pm 0.0009	0.7987 \pm 0.0144	0.7832 \pm 0.0301	0.6267 \pm 0.0786
B0	GFIQA-20k	0.4715 \pm 0.0374	0.6251 \pm 0.0420	0.7299 \pm 0.0278	0.9648 \pm 0.0010	0.4803 \pm 0.0225	0.3342 \pm 0.0900
B3	GFIQA-20k	0.3911 \pm 0.1333	0.4722 \pm 0.1929	0.6684 \pm 0.0370	0.9636 \pm 0.0031	0.4507 \pm 0.0349	0.2682 \pm 0.0983
B5	GFIQA-20k	0.4617 \pm 0.0432	0.6204 \pm 0.0211	0.7096 \pm 0.0383	0.9723 \pm 0.0005	0.4930 \pm 0.0588	0.3211 \pm 0.0536
B0	DIQA	0.3074 \pm 0.0632	0.2376 \pm 0.0459	0.4052 \pm 0.0437	0.4307 \pm 0.0684	0.2330 \pm 0.0322	0.7470 \pm 0.0525
B3	DIQA	0.2261 \pm 0.0287	0.0754 \pm 0.0452	0.1961 \pm 0.0778	0.2682 \pm 0.0489	0.1164 \pm 0.0581	0.7958 \pm 0.0393
B5	DIQA	0.2002 \pm 0.0747	0.1528 \pm 0.0700	0.2051 \pm 0.1348	0.4170 \pm 0.0983	0.1189 \pm 0.0745	0.8666 \pm 0.0400
B0	All	0.8597 \pm 0.0106	0.8759 \pm 0.0086	0.9052 \pm 0.0025	0.9610 \pm 0.0017	0.7773 \pm 0.0113	0.8021 \pm 0.0550
B3	All	0.8903 \pm 0.0096	0.8890 \pm 0.0040	0.9015 \pm 0.0029	0.9646 \pm 0.0009	0.7845 \pm 0.0044	0.8249 \pm 0.0237
B5	All	0.9102 \pm 0.0086	0.9056 \pm 0.0069	0.9064 \pm 0.0027	0.9619 \pm 0.0023	0.7700 \pm 0.0141	0.8363 \pm 0.0349

Table 4: Spearman’s rank correlation coefficients of the IQA models on all test sets (mean \pm standard deviation). Top-3 best models for each dataset are bolded.

Model	Dataset	Kadid-10k	KonIQ-10k	SPAQ	GFIQA-20k	LIVE-ItW	DIQA
B0	Kadid-10k	0.8195 \pm 0.0202	0.3594 \pm 0.0581	0.6213 \pm 0.0422	0.4696 \pm 0.0564	0.4611 \pm 0.0175	0.3945 \pm 0.1072
B3	Kadid-10k	0.8444 \pm 0.0216	0.3338 \pm 0.0244	0.6037 \pm 0.0140	0.4881 \pm 0.0457	0.4068 \pm 0.0117	0.4721 \pm 0.0558
B5	Kadid-10k	0.8893 \pm 0.0156	0.3794 \pm 0.0623	0.6571 \pm 0.0292	0.6003 \pm 0.0432	0.4363 \pm 0.0285	0.4286 \pm 0.0726
B0	KonIQ-10k	0.5618 \pm 0.0222	0.8925 \pm 0.0067	0.8560 \pm 0.0048	0.9043 \pm 0.0052	0.8043 \pm 0.0066	0.5389 \pm 0.0455
B3	KonIQ-10k	0.5240 \pm 0.0230	0.8903 \pm 0.0063	0.8587 \pm 0.0046	0.9133 \pm 0.0037	0.7942 \pm 0.0070	0.5156 \pm 0.0352
B5	KonIQ-10k	0.4203 \pm 0.2628	0.7264 \pm 0.4205	0.6905 \pm 0.4022	0.7184 \pm 0.4445	0.6583 \pm 0.3562	0.5128 \pm 0.2363
B0	SPAQ	0.4855 \pm 0.0400	0.7710 \pm 0.0097	0.9107 \pm 0.0034	0.7976 \pm 0.0249	0.7586 \pm 0.0053	0.6412 \pm 0.0453
B3	SPAQ	0.4168 \pm 0.0412	0.7374 \pm 0.0231	0.9064 \pm 0.0029	0.8213 \pm 0.0158	0.7450 \pm 0.0058	0.6015 \pm 0.0375
B5	SPAQ	0.4223 \pm 0.0557	0.7837 \pm 0.0102	0.9075 \pm 0.0027	0.7928 \pm 0.0170	0.7672 \pm 0.0327	0.6335 \pm 0.0702
B0	GFIQA-20k	0.4827 \pm 0.0276	0.6031 \pm 0.0391	0.7296 \pm 0.0292	0.9643 \pm 0.0011	0.4469 \pm 0.0206	0.3657 \pm 0.0984
B3	GFIQA-20k	0.4727 \pm 0.0213	0.5579 \pm 0.0483	0.7024 \pm 0.0290	0.9643 \pm 0.0010	0.4140 \pm 0.0342	0.3112 \pm 0.1010
B5	GFIQA-20k	0.5054 \pm 0.0334	0.6145 \pm 0.0224	0.7318 \pm 0.0327	0.9715 \pm 0.0006	0.4640 \pm 0.0657	0.3716 \pm 0.0672
B0	DIQA	0.3162 \pm 0.0458	0.2266 \pm 0.0511	0.3807 \pm 0.0486	0.4065 \pm 0.0733	0.1912 \pm 0.0443	0.7019 \pm 0.0500
B3	DIQA	0.2149 \pm 0.0270	0.0701 \pm 0.0494	0.1589 \pm 0.0789	0.2582 \pm 0.0557	0.0892 \pm 0.0543	0.7657 \pm 0.0383
B5	DIQA	0.2024 \pm 0.0710	0.1456 \pm 0.0839	0.1605 \pm 0.1430	0.4151 \pm 0.1015	0.0958 \pm 0.0760	0.8466 \pm 0.0420
B0	All	0.8570 \pm 0.0108	0.8427 \pm 0.0064	0.9020 \pm 0.0037	0.9614 \pm 0.0017	0.7559 \pm 0.0152	0.7793 \pm 0.0600
B3	All	0.8891 \pm 0.0115	0.8584 \pm 0.0072	0.8975 \pm 0.0043	0.9639 \pm 0.0008	0.7578 \pm 0.0066	0.8153 \pm 0.0225
B5	All	0.9096 \pm 0.0107	0.8838 \pm 0.0109	0.9045 \pm 0.0049	0.9625 \pm 0.0019	0.7445 \pm 0.0138	0.8125 \pm 0.0364

We combined the existing natural IQA datasets with a novel artificially distorted dermatology image quality database, Legit.Health-DIQA-Artificial, annotated by non-expert human observers. In future works, we will explore the differences between the image quality ratings of non-expert human observers and dermatologists to discern whether it is mandatory to obtain quality ratings from expert observers or human crowds suffice, as well as more precise methods of MOS scaling between datasets. We will also leverage

dermatology IQA datasets with authentic distortions to account for real-life scenarios.

In conclusion, cross-domain training of IQA models expands their capabilities and ensure proper performance across image domains, distortions, and observer groups. This innovation allows for seamless implementation of DIQA in teledermatology workflows, ensuring visual quality and maximizing the clinical utility of images. As a result, DIQA models not only have the potential to improve the reliability of daily clinical practice but also to strengthen the

quality and consistency of dermatological data used in clinical trials, ultimately benefiting patients and advancing dermatological research.

7 Data availability statement

The Legit.Health-DIQA-Artificial dataset image quality ratings and code are not publicly available due to intellectual property restrictions.

Online dermatology images were collected from Atlas Dermatológico <https://atlasdermatologico.com.br/>, Meddean Dermatology Atlas <https://www.meddean.luc.edu/lumen/meded/medicine/dermatology/melton/atlas.htm>, Black & Brown skin <https://www.blackandbrownskin.co.uk/campaigns>, and Hellenic Dermatological Atlas <https://www.hellenicdermatlas.com/en/search/browse>.

Acknowledgments

This project has been funded by the Department of Industry, Energy Transition and Sustainability of the Basque Government (BIKAIN-TEK Program). We also acknowledge the financial support provided by the European Union NextGenerationEU and the Centro para el Desarrollo Tecnológico y la Innovación (CDTI) through the support program PYME con Sello de Excelencia del Acelerador del EIC de Horizonte Europa (SoE-20221020), and the annotation services provided by LinkedAI.

References

- [1] Nisar Ahmed and Shahzad Asif. 2022. BIQ2021: a large-scale blind image quality assessment database. *Journal of Electronic Imaging* 31, 5 (2022), 053010–053010.
- [2] Marine Amouroux, Sébastien Le Cunff, Alexandre Haudrechy, and Walter Blondel. 2017. Image quality assessment for teledermatology: from consumer devices to a dedicated medical device. In *Design and Quality for Biomedical Technologies X*, Vol. 10056. SPIE, 105–112.
- [3] Atrousseau, Florent, and Stütz, Thomas, and Pankajakshan, Vinod. 2010. Subjective quality assessment of selective encryption techniques. <http://www.polytech.univ-nantes.fr/autrusseau-f/Databases/SelectiveEncryption/s/>.
- [4] Igor Barros Barbosa, Theoharis Theoharis, Christian Schellewald, and Cham Athwal. 2013. Transient biometrics using finger nails. In *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE, 1–6.
- [5] Igor Barros Barbosa, Theoharis Theoharis, and Ali E Abdallah. 2016. On the use of fingernail images as transient biometric identifiers: Biometric recognition using fingernail images. *Machine Vision and Applications* 27 (2016), 65–76.
- [6] Simone Bianco, Luigi Celona, Paolo Napolitano, and Raimondo Schettini. 2018. On the use of deep learning for blind image quality assessment. *Signal, Image and Video Processing* 12 (2018), 355–362.
- [7] Mollie R Cummins, Triton Ong, Julia Ivanova, Janelle F Barrera, Hattie Wilczewski, Hiral Soni, Brandon M Welch, and Brian E Bunnell. 2023. Consensus Guidelines for Teledermatology: Scoping Review. *JMIR dermatology* 6 (2023), e46121.
- [8] J Dahllén Gyllencreutz, E Johansson Backman, K Terstappen, and J Paoli. 2018. Teledermoscopy images acquired in primary health care and hospital settings—a comparative study of image quality. *Journal of the European Academy of Dermatology and Venerology* 32, 6 (2018), 1038–1043.
- [9] Kanjar De and V Masilamani. 2013. A new no-reference image quality measure for blurred images in spatial domain. *Journal of Image and Graphics* 1, 1 (2013), 39–42.
- [10] Yuming Fang, Hanwei Zhu, Yan Zeng, Kede Ma, and Zhou Wang. 2020. Perceptual quality assessment of smartphone photography. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 3677–3686.
- [11] Deepti Ghadiyaram and Alan C Bovik. 2015. Massive online crowdsourced study of subjective and objective picture quality. *IEEE Transactions on Image Processing* 25, 1 (2015), 372–387.
- [12] Mara Giavina Bianchi, Andre Santos, and Eduardo Cordioli. 2021. Dermatologists' perceptions on the utility and limitations of teledermatology after examining 55,000 lesions. *Journal of telemedicine and telecare* 27, 3 (2021), 166–173.
- [13] Gonzalez, Esther and Alvarez, Luis and Mazorra, Luis. 2024. AMI Ear Database. https://webctim.ulpgc.es/research_works/ami_ear_database/. Online; accessed 01-September-2024.
- [14] Ignacio Hernández Montilla, Alfonso Medela, Taig Mac Carthy, Andy Aguilar, Pedro Gómez Tejerina, Alejandro Vilas Sueiro, Ana María González Pérez, Laura Vergara de la Campa, Loreto Luna Bastante, Rubén García Castro, et al. 2023. Automatic International Hidradenitis Suppurativa Severity Score System (AIHS4): A novel tool to assess the severity of hidradenitis suppurativa using artificial intelligence. *Skin Research and Technology* 29, 6 (2023), e13357.
- [15] Daniel T Hogarty, John C Su, Kevin Phan, Mohamed Attia, Mohammed Hossny, Saeid Nahavandi, Patricia Lenane, Fergal J Moloney, and Anousha Yazdabadi. 2020. Artificial intelligence in dermatology—where we are and the way to the future: a review. *American journal of clinical dermatology* 21 (2020), 41–47.
- [16] Vlad Hosu, Hanhe Lin, Tamas Sziranyi, and Dietmar Saupe. 2020. KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment. *IEEE Transactions on Image Processing* 29 (2020), 4041–4056.
- [17] Esmee Irvine, Leela Sayed, Nick Johnson, and Joseph Dias. 2023. The ability of patients to provide standardized, patient-taken photographs for the remote assessment of Dupuytren disease. *HAND* 18, 1 (2023), 139–144.
- [18] Raluca Jalaboi, Ole Winther, and Alfiia Galimzianova. 2023. Explainable image quality assessments in teledermatological photography. *Telemedicine and e-Health* (2023).
- [19] Eric C Larson and Damon M Chandler. 2010. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of electronic imaging* 19, 1 (2010), 011006–011006.
- [20] Jonathan J Lee and Joseph C English. 2018. Teledermatology: a review and update. *American journal of clinical dermatology* 19 (2018), 253–260.
- [21] Hanhe Lin, Vlad Hosu, and Dietmar Saupe. 2019. KADID-10k: A large-scale artificially distorted IQA database. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 1–3.
- [22] Hanhe Lin, Vlad Hosu, and Dietmar Saupe. 2020. DeepFL-IQA: Weak supervision for deep IQA feature learning. *arXiv preprint arXiv:2001.08113* (2020).
- [23] Xinwei Liu, Marius Pedersen, and Jon Yngve Hardeberg. 2014. CID: IQ—a new image quality database. In *Image and Signal Processing: 6th International Conference, ICISP 2014, Cherbourg, France, June 30–July 2, 2014. Proceedings 6*. Springer, 193–202.
- [24] Teresa Lopez de Coca, Lucrecia Moreno, Mónica Alacreu, and Maria Sebastian-Morello. 2022. Bridging the generational digital divide in the healthcare environment. *Journal of Personalized Medicine* 12, 8 (2022), 1214.
- [25] Remedios López-Liria, María Ángeles Valverde-Martínez, Antonio López-Villegas, Rafael Jesús Bautista-Mesa, Francisco Antonio Vega-Ramírez, Salvador Peiró, and Cesar Leal-Costa. 2022. Teledermatology versus face-to-face dermatology: An analysis of cost-effectiveness from eight studies from Europe and the United States. *International Journal of Environmental Research and Public Health* 19, 5 (2022), 2534.
- [26] Ilya Loshchilov, Frank Hutter, et al. 2017. Fixing weight decay regularization in adam. *arXiv preprint arXiv:1711.05101* 5 (2017).
- [27] Taig Mac Carthy, Ignacio Hernández Montilla, Andy Aguilar, Rubén García Castro, Ana María González Pérez, Alejandro Vilas Sueiro, Laura Vergara de la Campa, Fernando Alfageme, and Alfonso Medela. 2024. Automatic Urticaria Activity Score: Deep Learning–Based Automatic Hive Counting for Urticaria Severity Assessment. *JID Innovations* 4, 1 (2024), 100218.
- [28] Kate Maier, Luiz Zaniolo, and Oge Marques. 2022. Image quality issues in teledermatology: A comparative analysis of artificial intelligence solutions. *Journal of the American Academy of Dermatology* 87, 1 (2022), 240–242.
- [29] Alfonso Medela, Taig Mac Carthy, S Andy Aguilar Robles, Carlos M Chiesa-Estomba, and Ramon Grimalt. 2022. Automatic SCORing of atopic dermatitis using deep learning: a pilot study. *JID Innovations* 2, 3 (2022), 100107.
- [30] Teresa Mendonça, Pedro M Ferreira, Jorge S Marques, André RS Marcal, and Jorge Rozeira. 2013. PH 2-A dermoscopic image database for research and benchmarking. In *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 5437–5440.
- [31] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. 2012. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing* 21, 12 (2012), 4695–4708.
- [32] Ignacio Hernández Montilla, Taig Mac Carthy, Andy Aguilar, and Alfonso Medela. 2023. Dermatology Image Quality Assessment (DIQA): Artificial intelligence to ensure the clinical utility of images for remote consultations and clinical trials. *Journal of the American Academy of Dermatology* 88, 4 (2023), 927–928.
- [33] Krzysztof Okarma. 2019. Current trends and advances in image quality assessment. *Elektronika ir Elektrotechnika* 25, 3 (2019), 77–84.
- [34] Andre GC Pacheco, Gustavo R Lima, Amanda S Salomaa, Breno Krohling, Igor P Biral, Gabriel G de Angelo, Fábio CR Alves Jr, José GM Esgario, Alana C Simora, Pedro BC Castro, et al. 2020. PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. *Data in brief* 32 (2020), 106221.
- [35] Sabela Paradelo-De-La-Morena, Rosa Fernandez-Torres, Walter Martínez-Gómez, and Eduardo Fonseca-Capdevila. 2015. Teledermatology: diagnostic reliability in 383 children. *European Journal of Dermatology* 25 (2015), 563–569.

- [36] Paola Pasquali. 2014. Photography in Dermatology. In *Skin Cancer*. <https://api.semanticscholar.org/CorpusID:74385942>
- [37] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. 2015. Image database TID2013: Peculiarities, results and perspectives. *Signal processing: Image communication* 30 (2015), 57–77.
- [38] Nikolay Ponomarenko, Vladimir Lukin, Alexander Zelensky, Karen Egiazarian, Marco Carli, and Federica Battisti. 2009. TID2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of modern radioelectronics* 10, 4 (2009), 30–45.
- [39] G Romero, D De Argila, L Ferrandiz, MP Sánchez, S Vañó, R Taberner, P Pasquali, C de la Torre, F Alfageme, J Malveyh, et al. 2018. Modelos de práctica de la teledermatología en España. Estudio longitudinal 2009-2014. *Actas Dermo-Sifiliográficas* 109, 7 (2018), 624–630.
- [40] Jayesh Ruikar and Saurabh Chaudhury. 2023. NITS-IQA Database: A New Image Quality Assessment Database. *Sensors* 23, 4 (2023), 2279.
- [41] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik. 2006. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing* 15, 11 (2006), 3440–3451.
- [42] Ernestasia Siahaan, Alan Hanjalic, and Judith A Redi. 2016. Does visual quality depend on semantics? A study on the relationship between impairment annoyance and image semantics at early attentive stages. *Electronic Imaging* 28 (2016), 1–9.
- [43] Leslie N Smith. 2017. Cyclical learning rates for training neural networks. In *2017 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 464–472.
- [44] Leslie N Smith and Nicholay Topin. 2019. Super-convergence: Very fast training of neural networks using large learning rates. In *Artificial intelligence and machine learning for multi-domain operations applications*, Vol. 11006. SPIE, 369–386.
- [45] Shaolin Su, Hanhe Lin, Vlad Hosu, Oliver Wiedemann, Jinqiu Sun, Yu Zhu, Hantao Liu, Yanning Zhang, and Dietmar Saupé. 2023. Going the extra mile in face image quality assessment: A novel database and model. *IEEE Transactions on Multimedia* (2023).
- [46] Hossein Talebi and Peyman Milanfar. 2018. NIMA: Neural image assessment. *IEEE transactions on image processing* 27, 8 (2018), 3998–4011.
- [47] Mingxing Tan. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946* (2019).
- [48] Nello Tommasino, Matteo Megna, Sara Cacciapuoti, Alessia Villani, Fabrizio Martora, Angelo Ruggiero, Lucia Genco, and Luca Potestio. 2024. The Past, the Present and the Future of Teledermatology: A Narrative Review. *Clinical, Cosmetic and Investigational Dermatology* Volume 17 (3 2024), 717–723. <https://doi.org/10.2147/CCID.S462799>
- [49] S Vañó-Galván, A Hidalgo, I Aguayo-Leiva, M Gil-Mosquera, L Rios-Buceta, MN Plana, J Zamora, A Martorell-Calatayud, and P Jaén. 2011. Store-and-forward teledermatology: assessment of validity in a series of 2000 observations. *Actas Dermo-Sifiliográficas (English Edition)* 102, 4 (2011), 277–283.
- [50] Maria João M Vasconcelos, Luís Rosado, and Márcia Ferreira. 2014. Principal axes-based asymmetry assessment methodology for skin lesion image analysis. In *International symposium on visual computing*. Springer, 21–31.
- [51] Toni Virtanen, Mikko Nuutinen, Mikko Vaaherankoska, Pirkko Oittinen, and Jukka Häkkinen. 2014. CID2013: A database for evaluating no-reference image quality assessment algorithms. *IEEE Transactions on Image Processing* 24, 1 (2014), 390–402.
- [52] Kailas Vodrahalli, Justin Ko, Albert S Chiou, Roberto Novoa, Abubakar Abid, Michelle Phung, Kiana Yekrang, Paige Petrone, James Zou, and Roxana Daneshjou. 2022. Development and Clinical Evaluation of an AI Support Tool for Improving Telemedicine Photo Quality. *arXiv preprint arXiv:2209.09105* (2022).
- [53] Xiaoping Wu, Ni Wen, Jie Liang, Yu-Kun Lai, Dongyu She, Ming-Ming Cheng, and Jufeng Yang. 2019. Joint acne image grading and counting via label distribution learning. In *Proceedings of the IEEE/CVF international conference on computer vision*. 10642–10651.
- [54] Jingtao Xu, Peng Ye, Qiaohong Li, Haiqing Du, Yong Liu, and David Doermann. 2016. Blind Image Quality Assessment Based on High Order Statistics Aggregation. *IEEE Transactions on Image Processing* 25 (9 2016), 4444–4457. Issue 9. <https://doi.org/10.1109/TIP.2016.2585880>
- [55] Dan Yang, Veli-Tapani Peltoketo, and Joni-Kristian Kamarainen. 2019. CNN-based cross-dataset no-reference image quality assessment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. 0–0.
- [56] Weixia Zhang, Kede Ma, Guangtao Zhai, and Xiaokang Yang. 2021. Uncertainty-aware blind image quality assessment in the laboratory and wild. *IEEE Transactions on Image Processing* 30 (2021), 3474–3486.
- [57] Timothy Zoltie, Sigrid Blome-Eberwein, Sarah Forbes, Mike Theaker, and Walayat Hussain. 2022. Medical photography using mobile devices. *bmj* 378 (2022).