Automated Image Quality Assessment in Teledermatolgoy

Topic:	Teledermatology, Image Quality Assessment, Artificial Intelligence & Machine Learning
Student:	Choekyel Nyungmartsang
Advisor:	Dr. Amruthalingam Ludovic
Expert:	Andrin Bürli
Client:	Algorithmic Business Research Lab, University Hospital of Basel and derma2go
Keywords:	Teledermatology, Image Quality Assessment, AI in Healthcare, Machine Learning, Synthetic Distortion

1. Problem Definition

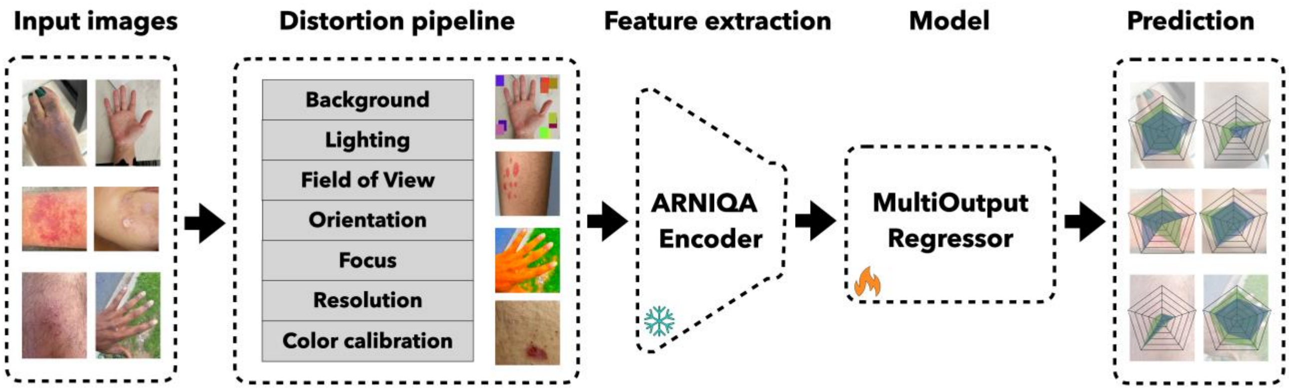
The goal of this thesis was to automate image quality assessment in the context of teledermatology. Teledermatology, a branch of telemedicine, has become increasingly popular, especially due to the COVID-19 pandemic. It allows patients to get dermatological advice remotely by taking pictures of their skin conditions with smartphones and sending them to dermatologists for analysis. However, the success of teledermatology heavily depends on the quality of these images. Many images do not meet the required standards due to poor lighting, blurriness, or insufficient clarity of the skin conditions.

This thesis focuses on assessing image quality so that the system can be used in real-time when patients are taking their images. This makes sure that only good quality images are sent to dermatologists, reducing the need for back-and-forth exchanges to retake better images. Additionally, this saves time, provides faster help, and reduces frustration for both patients and dermatologists, making teledermatology more efficient.

2. Solution Concept

To assess image quality, a distortion pipeline was developed that takes domain-specific good quality images from dermatology and distorts them based on seven criteria set by the International Skin Imaging Collaboration (ISIC): Background, Lighting, Focus, Orientation, Color Calibration, Resolution, and Field of View. Each dermatology quality criterion has different types of distortions and five levels of severity. This results in multiple possible distortion combinations, allowing the same image to be passed through the pipeline multiple times to generate diverse labeled images.

A state-of-the-art image quality assessment approach called ARNIQA was used to extract image features. The ARNIQA backbone captures distortion patterns without relying on the content, which is important for accurately assessing image quality. The extracted features were then used to train a model that can accurately evaluate each of the seven dermatology quality criteria. The final predictions are displayed in a radar chart where each dermatology criterion is around the outside with the severity levels ranging from 0 to 1, where 0 indicates no distortion and 1 indicates high distortion.



3. Special Challenges

This research faced several challenges, particularly in selecting the appropriate image quality assessment methods for teledermatology. Choosing the right method was very important as it determined how well the model would work. Understanding the field of dermatology was particularly helpful in this process. It allowed for a better connection between technical aspects and practical needs, making sure the focus remained on the main objective. Reviewing many relevant papers and receiving guidance from the advisor also played a key role in overcoming this challenge.

Creating realistic distortions for the seven dermatology quality criteria was another major challenge. It involved a lot of trial and error to find the right types and ranges of distortions, as these distortions simulate how patient-taken images could be. This was critical because the accuracy of the assessment depends on it. Selecting good quality images for training and testing was also a time-consuming and repetitive part of the research. This task was done manually, requiring focused attention, and had to be done in intervals to avoid rushing. Another challenge was labeling images with combinations of authentic distortions. For example, if an image was too dark, it became difficult to judge other criteria like resolution or color calibration, making the labeling process more complex. This difficulty in assessing each quality criterion accurately made the task more challenging.

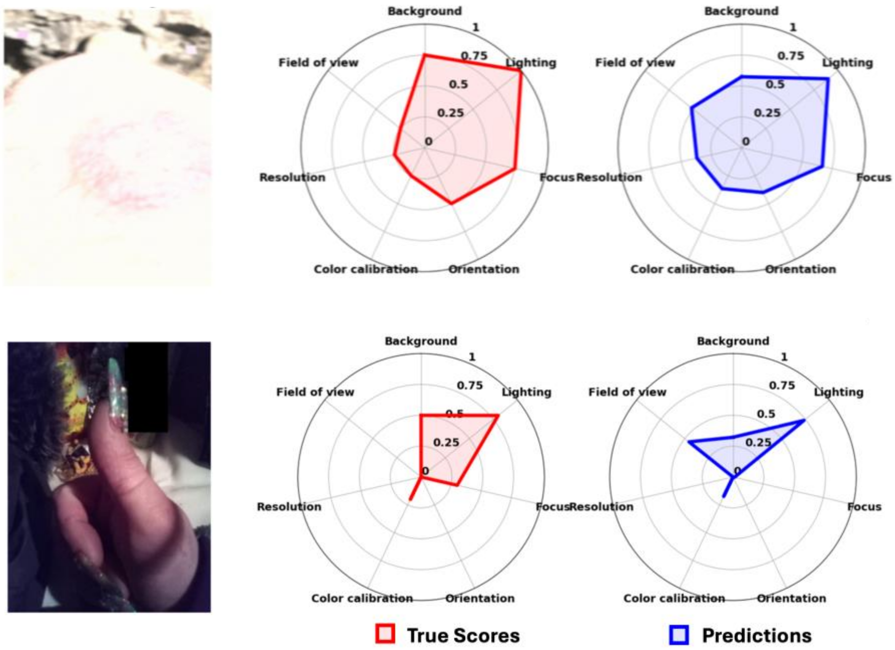
In total, 1'400 labels were assigned and 1'175 images were filtered for having good quality, which included different body parts and various skin issues. Despite these difficulties, these tasks were essential to achieve the research goals.

4. Results

The experiments were conducted using four different multi-output models: Extreme Gradient Boosting (XGBoost) regressor, XGBoost classifier, Multi-Layer Perceptron (MLP) regressor, and MLP classifier. The MLP regressor emerged as the best-performing model, consistently achieving higher Spearman's Rank Order Correlation Coefficient (SRCC) scores than the other models on the combined dataset across all seven dermatology quality criteria. However, the model faced challenges with certain criteria, such as background, orientation, and field of view distortions, which resulted in higher errors and less accurate predictions.

When combining the seven quality scores into a single final quality score using the weighted average, the proposed approach achieved a higher SRCC score than ARNIQA. This demonstrates that using synthetic distortion on domain-specific images improves performance.

Additionally, out-of-distribution testing with images like vehicles, animals, and nature showed generally accurate predictions for most distortions. However, the model struggled with orientation and field of view distortions in these cases as well. This shows that while the model is not overfitted to teledermatology images and is adjusted to handle important distortions, there is still room for improvement in these specific areas. This balance makes sure that the model is flexible and effective in different situations without losing its focus on teledermatology.



5. Outlook

Future research should focus on expanding the dataset with more diverse and representative teledermatology images, particularly addressing the challenging quality criteria like background, orientation, and field of view. Collaborating with dermatologists to help filter and label images will improve accuracy and reduce human error. Adjusting the distortion ranges and types with expert input can further improve performance.

Making the model more understandable is another key area for future work. Techniques like GradCam can highlight which parts of an image the model focuses on when making predictions. This will help in understanding the model's decision-making process and spotting potential errors.

The reproducible repository developed in this research provides a solid foundation for ongoing exploration and development. It allows future researchers to build upon and improve the methods and tools created in this thesis. Improving image quality assessment in teledermatology will significantly improve remote consultations, making sure that patients and dermatologists can rely on good quality images for accurate diagnoses and effective treatment.