TECHNOLOGY FEATURE HealthTech Report / Research Findings

Deep Learning Classification of Skin Diseases with Explainable Predictions

Develop a deep learning model for automated classification of skin lesions using ISIC 2019 and HAM1000 datasets, integrate GRAD-CAM to think beyond the box, and highlight the relevant skin regions.

Hafiz Muhammad Usama | **University of Agriculture, Faisalabad** | **Published**: 14 April 2026
hafizmuhammadusama664@gmail.com | Student / Computer Sciences

Introduction

Early detection can help control disease progression, reduce symptoms, and improve the quality of life for various illnesses. With the technological development, artificial intelligence and machine learning approaches can intervene in the early stage of disease diagnosis.

The skin is the largest organ covering the entire human body [1]. Any disorder that affects the human skin is called a skin disease [2]. Skin disease is one of the most contagious diseases in the world. According to the World Health Organization (WHO) data published in 2020, skin disease deaths in Pakistan were recorded at 132 or 0.01% of total deaths [3]. Skin diseases are caused by viruses, bacteria, allergies, and fungal infections [4]. Skin disease typically affects the epidermis, the thin outer layer of skin, which means lesions are often visible to the naked eye.

There are different types of skin lesions: Actinic Keratosis (AK), Basal Cell carcinoma (BCC), Benign Keratosis (BKL), Dermatofibroma (DF), Melanoma (MEL), Melanocytic nevus (NV), Squamous cell carcinoma (SCC), Vascular Lesions (VASC), as shown in Fig 1. Out of these skin diseases, melanoma is the most deadly and dangerous type. Many people are unaware of the types, symptoms, and stages of skin disease, which creates a significant gap between patients and dermatologists. The automatic deep learning classification system has made it possible to detect skin diseases more accurately.

> ***"Accuracy alone does not earn clinical trust. GRAD-CAM shows exactly which pixels drove the prediction, and that changes everything."***

Technology Overview

Model selection started with U-Net and EfficientNet-B0. U-Net was tested for its segmentation capability, separating the lesions before classification [5]. EfficientNet-B0 ran alongside as a classifier model picked for its efficiency on suitable hardware [6]. Both produced remarkable results.

That gap pushed me towards Swin Transformer. Unlike CNNs, which process images through fixed local filters, Swim applies shifted window self-attention towards hierarchical structures [7]. Classification accuracy improved across all the disease categories. Training used a stratified split, AdamW optimization, and validation checkpoints to save the best-performing model weight rather than simply the final ones.

In this procedure, explainability was treated as a core requirement, not an afterthought. The system uses both GRAD-CAM and swin's internal attention map. GRAD-CAM traces gradient flow to identify affected input regions [8]. The visualization map reads the transformer's own self-attention weight to show how the model focuses during the interface [9]. These two methods complement each other; one emphasizes where gradients were strong, and the other shows where the model looked.

We use two well-known datasets, the "International Skin Imaging Collaboration (ISIC) 2019 "[10] and the second one is the HAM1000 "Human Against Machines with 10000 images" [11] dataset, to train, test, and validate the proposed model. The HAM1000 dataset consists of 10000 dermatoscopic images of pigmented skin lesions taken from patients in Australia and Austria.

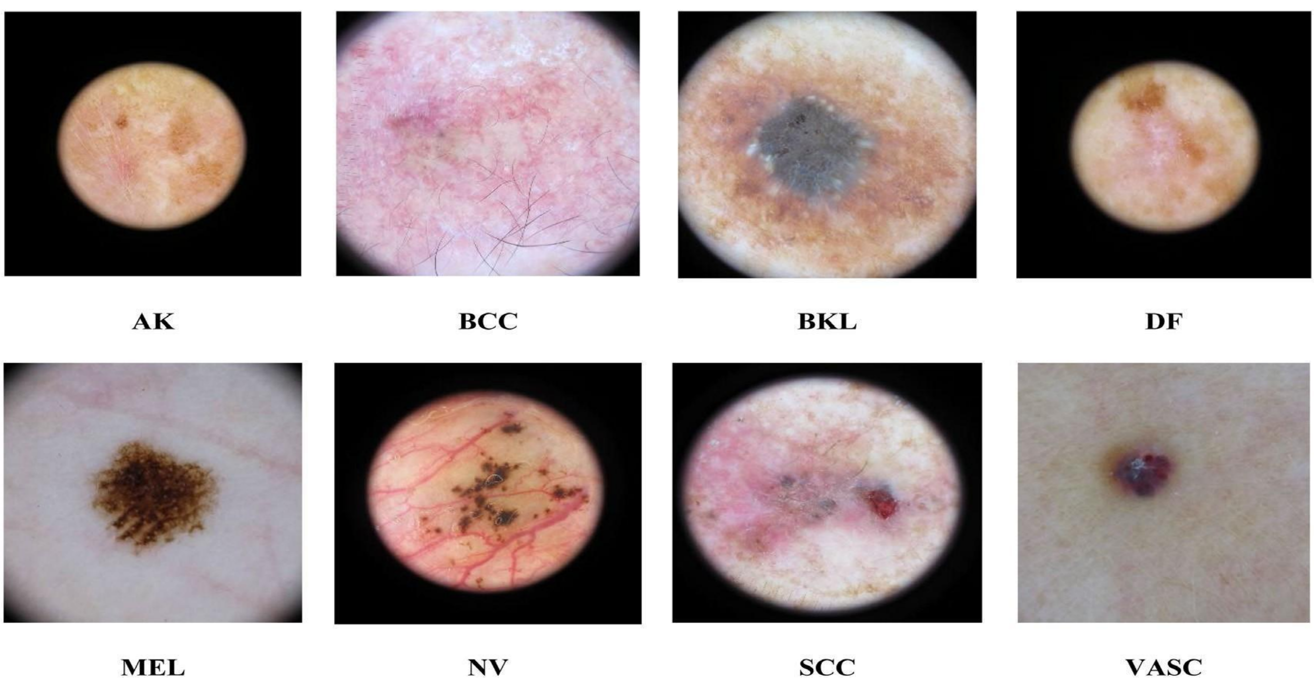

Figure 1: Representative Samples of Skin Lesion Categories.

For both the ISIC 2019 and HAM10000 datasets, we addressed class imbalance by randomly duplicating samples from smaller classes until all the classes matched the size of the largest. No other modifications were made for the data distribution. By doing this, it gave the model a fair chance to learn from the underrepresented classes without producing synthetic noise.

Key Findings & Impact

All the experiments were performed on the Python platform. We trained on the ISIC 2019 challenge dataset, which contains 25331 images, and tested on HAM10000 to check whether the approach holds up beyond a single dataset. We have completed the processing task before feature extraction.

Performance was measured using accuracy, precision, recall, and F1 score with direct comparisons to existing methods. Since both datasets are class-imbalanced, we evaluate the model twice, once on the raw distributions and once after oversampling. The two-pass evaluation makes it easier to isolate how much of the performance gain comes from the balanced data rather than the model itself.

	precision	recall	f1-score	support
0	0.64	0.80	0.71	223
1	0.95	0.94	0.94	1341
2	0.72	0.89	0.80	103
3	0.83	0.60	0.70	65
4	0.80	0.69	0.74	220
5	0.95	0.78	0.86	23
6	1.00	0.79	0.88	28
accuracy			0.88	2003
macro avg	0.84	0.78	0.80	2003
weighted avg	0.89	0.88	0.88	2003

Figure 2: Classification Performance metrics on the HAM10000 dataset

What makes this work more than a benchmark exercise is the explainability layer. A model that outputs a confidence score without showing its reason is a hard sell in clinical practice. Dermatologists need to know why something was flagged. Especially when the system is being deployed as the first pass screening tool in a region where there is no reliable specialist access, a wrong flag in any direction has resulted in real consequences.

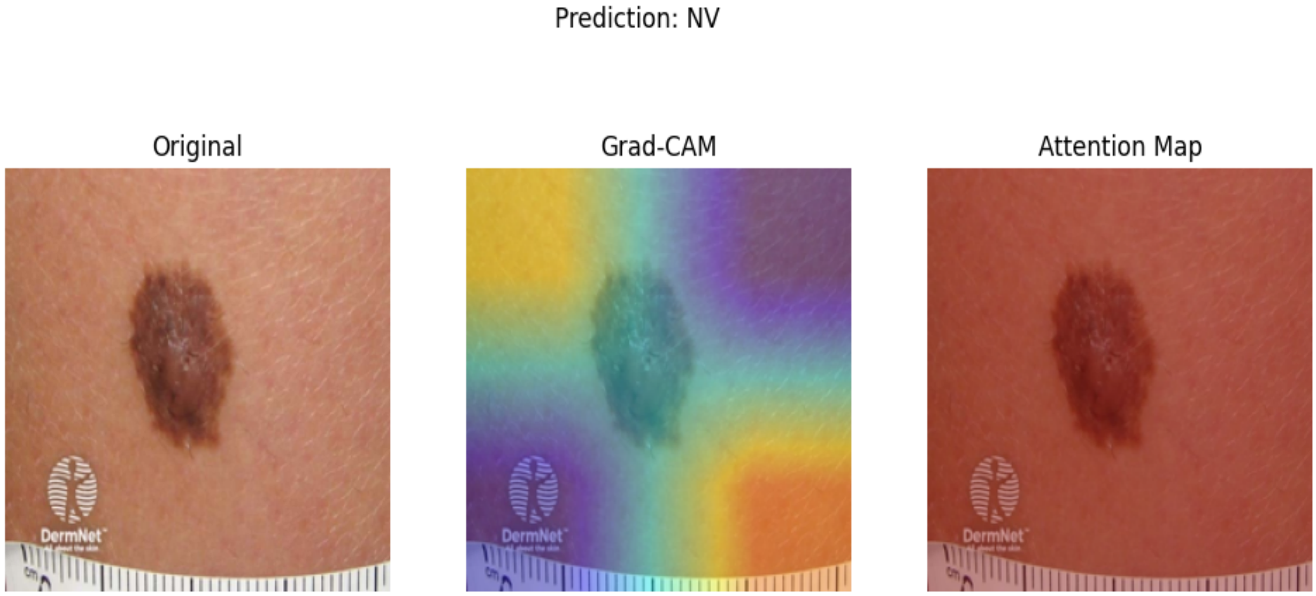

Figure 3: Model interpretability on the HAM10000 dataset

What's Next?

Skin disease is a widespread problem. People of many countries or regions suffer from different types of skin diseases. We can fight against these diseases by developing various techniques and methods. In this research, we have performed the work in several phases. We pre-processed two public datasets: ISIC 2019 and HAM10000, resized the dataset images to a uniform format, trained and compared several classification models, kept the best one, and lastly used GRAD-CAM to make the model's decision interpretable.

The model works well on balanced data and could extend to other skin disease classification tasks. Classification accuracy is still imperfect because the automatic segmentation occasionally fails to isolate the lesions. As a result, it leads to misclassification, which is the limitation of our case study. Stronger segmentation methods, or ensemble and deep learning approaches, could close that gap. Ultimately, the aim is earlier, more reliable detection for patients.

About the Author

Hafiz Muhammad Usama is a student of Big Data Analytics at the University of Agriculture, Faisalabad, with an interest in the field of deep learning and data science. He can be reached at hafizmuhammadusama664@email.com.

Keywords: Deep Learning; Machine Learning; Data Science; Big Data Analytics; ISIC 2019; HAM10000; GRAD-CAM; Pakistan Tech

References

[1] Anatomy of the skin, Stanford Children's Health, 2021, [Online]. Available: https://www.stanfordchildrens.org/

[2] M.W. Greaves, Skin disease, Britannica, 29, 2020, [Online]. Available: https://www.britannica.com/human-skin-disease

[3] Pakistan: Skin disease, 2020, [Online]. Available: https:worldlifeexpectancy.

[4] Cancer, World Health Organization, 21, 2021, [Online]. Available: https://www. who.int/news-room/detail/.

[5] Ranneberger, O., Fischer, P., & Brox, T. (2015). Available: U-Net: for Biomedical Image Segmentation. In MICCAI.

[6] Tan, M., & Le, Q. (2019). Available: *Efficient Net: Rethinking Model Scaling for Convolutional Neural Networks*. ICML.

[7] Liu, Z., Lin, Y., Cao, Y., et al. (2021). Available: Swin Transformer: Hierarchical Vision Transformer using Shifted Windows.

[8] Selvaraju, R. R., Cogswell, M., Das, A., et al. (2017). Available: Grad-CAM: Visual Explanations from Deep Networks.

[9] *Dudovskiy, A., Beyer, L., Kolesnikov, A., et al. (2020). Available: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. ICLR.*

[10] *N. Codella, D. Gutman, B. Helba, M.E. Celebi, M. Combalia, J.M.H. Kittler, V. Rotenberg, P. Tschandl, Skin lesion analysis towards melanoma detection, in: ISIC, 2019, [Online]. Available:* https://challenge.isic-archive.com/landing/2019/

[11] Tschandl, P., Rosendahl, C., & Kittler, H. (2018).Available: The HAM10000 dataset: A large collection of pigmented skin lesions.

[12] Noise filtering in digital image processing, medium, 3, 2018, [Online]. Available: https://medium.com/noise-filtering-in-digital-image-processing

[13] N. Codella, D. Gutman, B. Helba, M.E. Celebi, M. Combalia, J.M.H. Kittler, V. Rotemberg, P. Tschandl, Skin lesion analysis towards melanoma detection, in: ISIC, 2019, [Online]. Available: https://challenge2019.isic-archive.com.

[14] Automatic diagnosis of skin diseases using a CNN. Available: https://www.sciencedirect.com/science/article/abs/pii/S0141933120300223

[15] Multiclass classification using support vector machines, Baeldung, 25, 2021, [Online]. Available: https://www.baeldung.com/cs/svm-multiclass-classification.