2026

Research Document & Context Background: Data Challenge

DERMATRACE: ECZEMA SEVERITY CLASSIFICATION AND THE CHALLENGE OF DISTINGUISH ECZEMA FROM ATOPIC DERMATITIS

KAILEN ROA

Contents

1.Introduction

Eczema is a common inflammatory skin condition that presents with redness, dryness, and irritation. In both clinical practice and automated image-based diagnosis, a key challenge is that eczema is often used as an umbrella term, while atopic dermatitis (AD) represents its most common clinical subtype. Visually, these conditions can be highly similar, making them difficult to distinguish even for trained dermatology professionals.

This research project explores the use of a convolutional neural network (CNN) inspired by the EczemaNet methodology to analyse dermatological images and to investigate how well a model can distinguish between eczema and atopic dermatitis. Rather than directly reproducing the original EczemaNet system, an independently trained CNN is used to study model behaviour, misclassifications, and decision boundaries.

The goal of this project is not to develop a fully accurate diagnostic tool, but to **understand where and why the model fails**, how class imbalance and visual overlap affect predictions, and how lightweight improvements, such as class weighting and threshold tuning, can improve robustness and interpretability.

2. Background: Eczema vs Atopic Dermatitis

Eczema **is a general term used to describe a group of inflammatory skin conditions** that are characterized by symptoms such as itching, redness, dryness, and impaired skin barrier function. It encompasses several subtypes that differ in cause, progression, and clinical presentation.

Atopic dermatitis **is the most common subtype of eczema** and is a chronic, relapsing inflammatory skin disease. It is typically associated with a genetic predisposition, immune system dysregulation, and a higher prevalence of other atopic conditions such as asthma or allergic rhinitis. Atopic dermatitis often develops in early childhood and follows a long-term course with recurring flare-ups.

Although atopic dermatitis is a form of eczema, not all eczema is atopic dermatitis. Other eczema subtypes, such as contact dermatitis or dyshidrotic eczema, arise from different underlying mechanisms and may present distinct clinical features. This distinction is important in medical contexts, as diagnosis and treatment strategies can vary depending on the specific eczema subtype.

3. Dataset overview

The project uses an image-based dataset similar in structure and intent to **EczemaNet**, consisting of clinical photographs of skin conditions. The images are labeled into two primary classes: **eczema** and **dermatitis**, based on clinical annotation. No additional patient metadata or contextual information is provided alongside the images.

The dataset exhibits several characteristics that influence interpretation and analysis:

- Images vary considerably in resolution and aspect ratio.
- There is a clear variation in lighting conditions, contrast levels, and skin tones
- In addition, the visual differences between the eczema and dermatitis classes are often subtle, as both conditions can present with similar inflammatory features such as redness, scaling, and texture changes.

The dataset size is relatively limited, which increases the risk of overfitting and constrains the diversity of visual patterns available for learning.

4. Project Steps and Methodology

4.1. Data Preparation

All images were resized to a fixed input shape to ensure consistency during training. Minimal but meaningful preprocessing was applied to avoid destroying clinically relevant features. Over-aggressive augmentation was avoided due to the limited dataset size.

4.2. Model Training

The model was trained in multiple stages:

1. Transfer learning with feature extraction, using a MobileNetV2 backbone pretrained on ImageNet, with the backbone layers frozen

2. Applying class weights to address class imbalance

3. Threshold tuning to control prediction sensitivity

Each step was evaluated using accuracy, confusion matrices, and error analysis rather than accuracy alone.

4.3. Explainability

To understand model behaviour, Grad-CAM visualizations were generated. These heatmaps highlight which regions of the image most influenced the model’s prediction:

Fig.1. GradCam after baseline training. Model knows where the skin problem is but missclassified the label

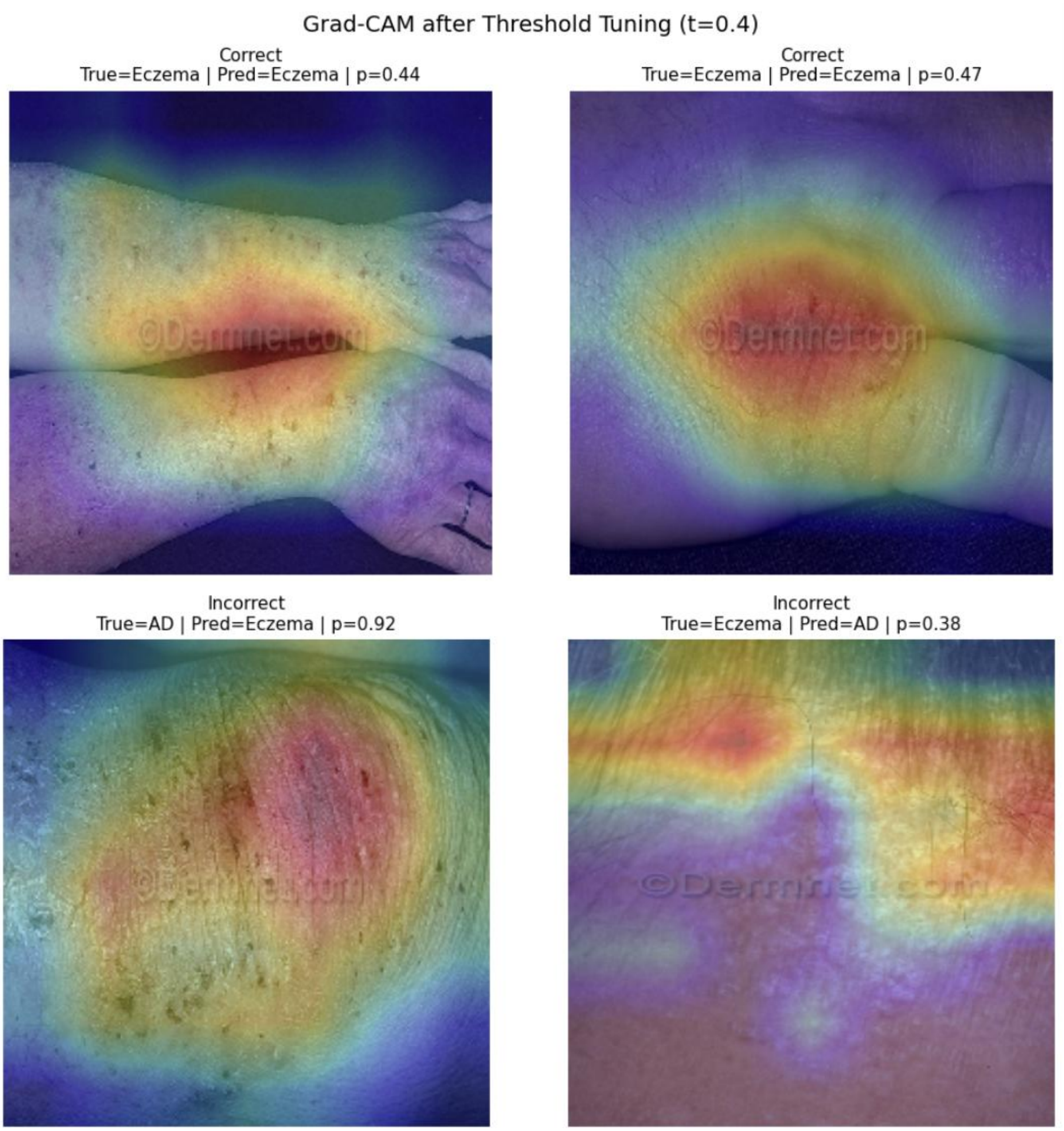


Fig.2. After threshold tuning we can see the difference. The class imbalances are balanced now, and the model still recognizes where the skin problem is.

5. Observations and Results

One of the most important findings is that the model **systematically over-predicts eczema** when presented with atopic dermatitis images. This confirms that visual similarity is the dominant source of error rather than random noise or poor training.

False negatives (eczema predicted as atopic dermatitis) tend to occur in:

- Mild cases
- Low-contrast images
- Subtle redness or texture changes

Grad-CAM results show that the model often focuses on general redness areas rather than lesion borders or texture irregularities. This suggests that the model learns **coarse visual cues** instead of clinically nuanced features.

Importantly, improving class weights or thresholds sometimes reduced one type of error while increasing another, highlighting the trade-off between sensitivity and specificity.

6. Discussion

These results demonstrate that the limitations of the model are closely tied to the limitations of the data. Without contextual or metadata information, the model can only rely on visual patterns, which are often insufficient for fine-grained dermatological distinctions.

Rather than indicating failure, these misclassifications validate the complexity of the problem. They also emphasize the importance of explainability tools like Grad-CAM, which make model behaviour transparent and interpretable.

This aligns with the project's objective: **understanding model behaviour**, not merely maximizing performance metrics.

7. Conclusion

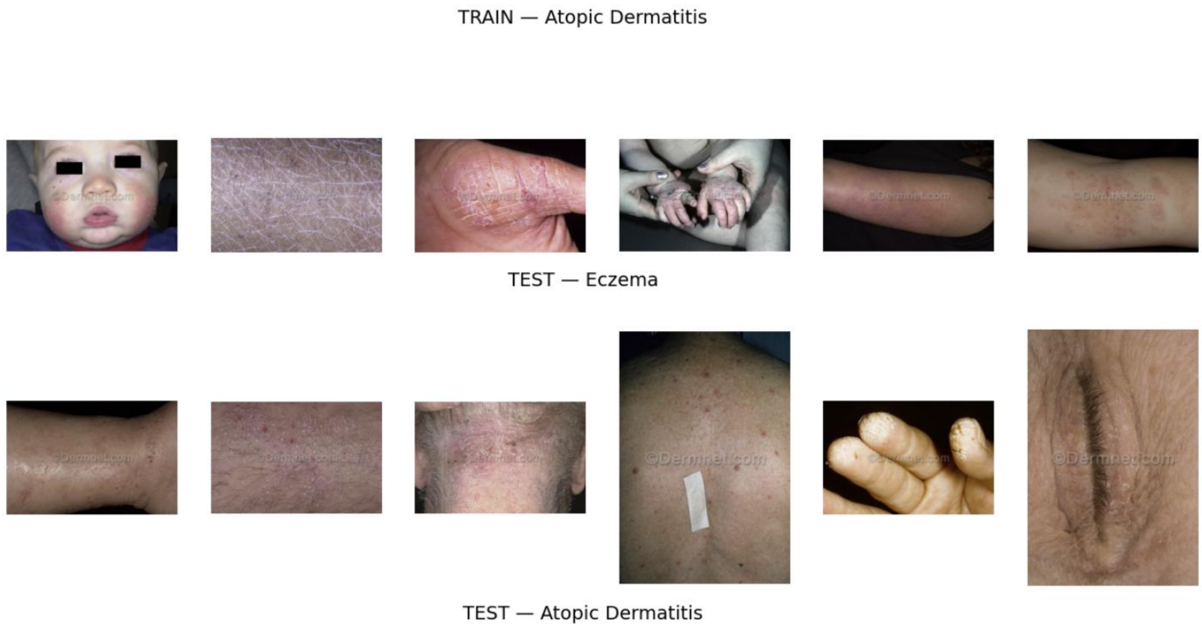
This project shows that CNN-based models can learn meaningful visual patterns related to eczema severity, but they struggle to reliably distinguish eczema from atopic dermatitis due to overlapping visual features and dataset ambiguity.

The model’s errors are systematic and explainable, not random. This confirms that automated skin condition classification is feasible but must be approached carefully, especially in medical or health-related contexts.

Overall, the project successfully demonstrates:

- The feasibility of image-based eczema severity classification
- The importance of explainability in medical AI
- The limitations of relying on images alone

8. Future Improvements



There’s a lot of improvements that can be done in this model to improve model’s and enhanced the performance of the system:

- Collecting more diverse images across skin tones, lighting conditions, and severity stages
- Integrating metadata such as age, skin type, body location, and lighting conditions
- Using a multi-stage classification pipeline (eczema vs non-eczema → subtype classification)
- Combining image features with environmental or patient-reported data
- Training with clinician-verified labels to reduce ambiguity

These improvements would move the model closer to a clinically meaningful decision-support tool rather than a purely visual classifier.

Final Note

This research should be seen **as an exploratory and explanatory study, not a diagnostic system**. Its value lies in understanding how AI behaves in complex medical domains and how it can be responsibly improved.