

"PCA-ENHANCED ENSEMBLE LEARNING FOR ECZEMA-CENTRED SKIN DISEASE CLASSIFICATION: A COMPARATIVE EVALUATION WITH MOBILENETV2-BASED DEEP LEARNING"

Dua Abro^{*1}, Pirah Daniyal², Arshiyah Naeem³, Ahmad Faraz⁴, Hira Hanif⁵

^{*1,2}IBET, LUMHS, Jamshoro

³LUMHS, Jamshoro

⁴NFC Institute of Engineering & Technology, Multan

⁵ Institute of Business Management (IOBM), Karachi

^{*1}dua.abro2001@gmail.com, ²lagharipirah3@gmail.com, ³arshiyahmemon1@gmail.com,

⁴ahmadfaraz3776@gmail.com, ⁵hira-hanif@hotmail.com

DOI: <https://doi.org/10.5281/zenodo.23205229>

Keywords

Pakistan, mediator state, US-Iran ceasefire, global leaders, critical discourse analysis.

Article History

Received: 25 June 2026

Accepted: 16 September 2026

Published: 07 October 2026

Copyright @Author

Corresponding Author: *

Dua Abro

Abstract

Atopic dermatitis (Eczema) is a chronic, inflammatory skin disease that poses significant diagnostic challenges. Due to its clinical overlap with other erythematous-squamous skin conditions such as psoriasis, seborrheic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis, and pityriasis rubra pilaris, atopic dermatitis (Eczema) is a chronic inflammatory skin disease that presents significant diagnostic challenges. Timely and accurate differential diagnosis is critical for effective treatment, yet remains heavily dependent on clinical expertise and invasive biopsy procedures. The majority of current systems work on either structured clinical records or dermatological images alone, leaving the comparative value and fusion potential of both modalities unexplored. However, recent advances in machine learning and deep learning offer compelling opportunities to automate and support dermatological diagnosis.

In order to classify skin diseases centred on eczema, this research suggests and assesses a dual-pipeline machine learning (ML) framework. The first pipeline processes structured histopathological and clinical data from the UCI Dermatology dataset (366 instances, 34 features, 6 disease classes including eczema) using Principal Component Analysis (PCA) for dimensionality reduction – compressing 34 features into 21 principal components retaining 95.3% of total variance – followed by three classical classifiers: Random Forest, Support Vector Machine (SVM), and XGBoost. The second pipeline applies transfer learning via MobileNetV2, pretrained on ImageNet, for deep visual feature extraction and classification across a skin image dataset of 12,453 training images spanning 23 visual skin condition classes, including eczema.

Results reveal a striking performance contrast between the two modalities. The tabular pipeline delivered near-perfect classification, with Random Forest achieving 97.3% accuracy, a macro F1-score of 0.970, and a macro ROC-AUC of 0.995. SVM and XGBoost followed at 96.2% and 95.6%

accuracy, respectively, with all models achieving perfect classification on several disease classes. Feature importance analysis consistently revealed that the most diagnostically discriminative components were PC1, PC2, PC3, and PC4, which correspond to dominant histopathological markers known to differentiate eczema from clinically similar conditions. On the 23-class visual task, however, the image-based MobileNetV2 pipeline only managed 20.8% training accuracy and about 17% validation accuracy. This shows considerable underfitting, which can be explained by high inter-class visual similarity, few fine-tuning epochs, and the semantic gap between ImageNet pre-training and clinical dermatological imagery.

These findings establish that structured histopathological features, when reduced via PCA and modelled with ensemble classifiers, yield superior diagnostic performance for eczema-related skin disease classification compared to image-based deep learning under current data constraints.

INTRODUCTION

The skin is the body's largest organ, and every individual must take care of it. Due to some reasons, every one person out of ten suffers from skin diseases like Eczema, it is the world's most challenging chronic inflammatory illness [1] [2]. It is an important public health problem due to its elevated frequency and related illness. Having the capacity to record illness flare-ups is crucial for assessing the effectiveness of interventions for long-term, recurring diseases, but it's not clear how to define these [3]. Many of the unresolved issues regarding long-term outcomes in Eczema, in particular, can be resolved directly by using electronic health information from standard primary care procedures [4]. A lot of information has completely changed how we research illness [5]. The adoption of contemporary computational techniques, which can integrate and analyze vast amounts of data and uncover hidden patterns and associations, has been aided by the increased availability of large and varied health care data sets [6]. Neural networks with deep and machine learning have been created to help with dermatological diagnosis and identification [7]. The pathophysiology of eczema is thought to be influenced by species *Staphylococcus aureus* colonization, infection, and antimicrobial production [8]. In ancient times eczema was known as an intense itch [9]. For a long time, it was believed to be the first sign of atopy, which is the inherited tendency to develop an IgE

sensitivity to environmental allergens, and the first stage of the so-called "atopic march," which eventually results in a condition called allergic rhinitis [10] [11]. Specifically, the past decade is known for the gradual rise of eczema [12]. Furthermore, because general medical practitioners are not required to enter codes for chronic conditions regularly, there is some evidence that chronic diseases, like eczema, may be more poorly reported over time in the United Kingdom primary care data [13] [14] [15]. Eczema [16] [17], over time, may get better or worse. It can cause asthma and hay fever in those who have it. The most common type of eczema is atopic dermatitis. Although it can also affect adults, it mainly affects infants and young children [18] [19] [20].

A subfield of artificial intelligence called machine learning (ML) enables computers to identify patterns in data and make judgments or predictions with little assistance from humans. The significance of machine learning has grown in recent years due to the increasing need for large-scale data analysis in clinical practice and medical research. Its strong data processing capabilities offer useful resources for decision support and medical diagnosis [21] [22] [23] [24] [25].

Research Gap

Eczema (Atopic Dermatitis) is one of the most prevalent chronic inflammatory skin conditions globally, affecting millions across all age groups.

Despite its widespread occurrence, accurate and early diagnosis remains difficult due to its visual similarity with other skin conditions such as psoriasis, seborrheic dermatitis, and contact dermatitis. Clinicians typically rely on a combination of patient history, clinical examination, and in some cases biopsy – a process that is time-consuming, subjective, and inaccessible in low-resource settings.

Existing machine learning approaches to eczema diagnosis are largely unimodal – they either analyse structured clinical/histopathological data or classify skin images independently. No systematic study has compared both modalities side by side on this problem, nor evaluated what each modality contributes diagnostically. This is the gap your work fills.

Methodology

This study adopts a two-track experimental design to compare how well structured clinical data and raw dermatological images support eczema diagnosis. The first track processes tabular histopathological records using classical machine learning, while the second applies a convolutional neural network to skin photographs. Each pipeline was built and evaluated independently, since no publicly available dataset currently provides paired records – that is, clinical features and a corresponding image for the same patient. Despite this constraint, running both tracks in parallel makes it possible to draw a meaningful, side-by-side comparison of the two modalities under realistic data conditions.

Tabular (structured clinical data)

The UCI Dermatology dataset was used, containing 366 patient records with 34 features – 12 clinical (erythema, scaling, itching, etc.) and 22 histopathological (acanthosis, parakeratosis, band-like infiltrate, etc.), plus patient age. The target variable has 6 disease classes, one of which is eczema (chronic dermatitis).

Missing values in the age column were handled first. The data was then split 80/20 into 292 training and 74 test samples. PCA was applied to

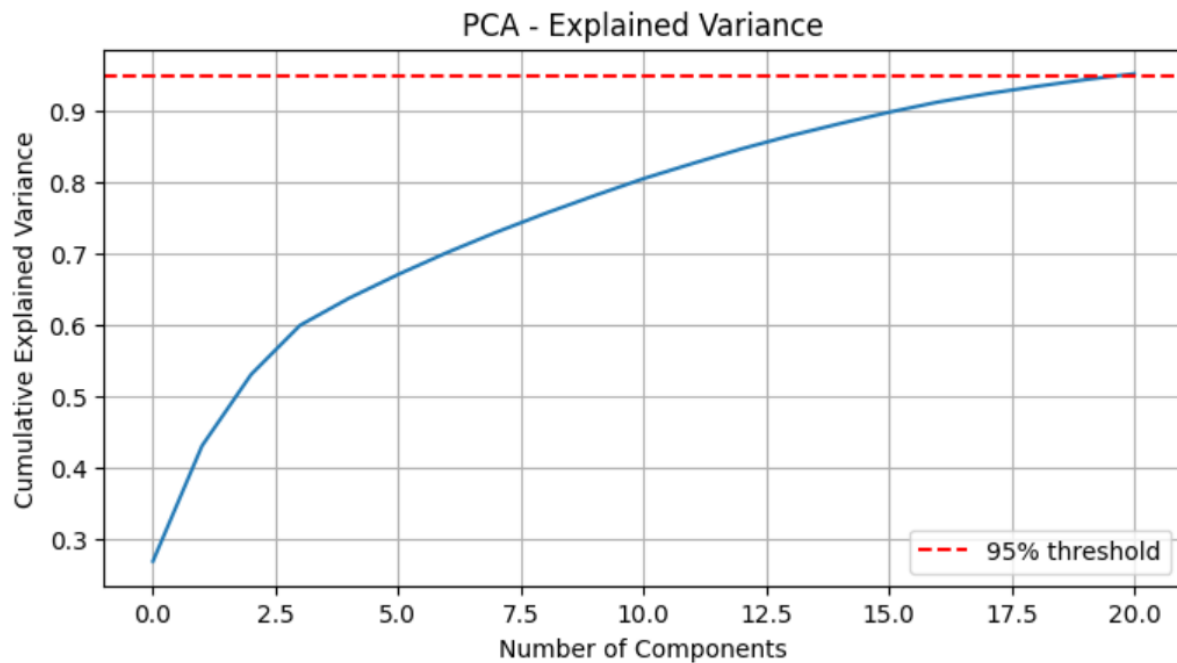
compress 34 features into 21 principal components capturing 95.3% of variance – this removed noise and redundancy before classification. Three classifiers were trained and evaluated using 5-fold cross-validation: Random Forest, SVM, and XGBoost. Final evaluation used accuracy, macro F1-score, macro ROC-AUC, and confusion matrices.

Image (deep learning on skin photos)

A skin image dataset of 12,453 training and 792 validation images across 23 skin condition classes (including eczema) was loaded. MobileNetV2 pretrained on ImageNet was used as a feature extractor, producing 1280-dimensional embeddings per image. The model was then fine-tuned in two phases: 15 epochs with learning rate 0.001 (with ReduceLROnPlateau scheduling), followed by a second fine-tuning phase with learning rate 1e-5. Training accuracy, validation accuracy, and loss curves were tracked across all epochs.

Results and Findings

The outputs from the two experimental pipelines are examined in this section. The training and validation curves from the MobileNetV2 image pipeline are shown after the tabular track results, which include PCA dimensionality reduction, classifier confusion matrices, and feature importance rankings. When applicable, cross-model patterns are highlighted to highlight more general findings regarding the characteristics and methods that worked best to differentiate eczema from the other five skin conditions in the dataset. Each of the following figures is preceded by a brief description of what it depicts and why it is relevant to this study.



The 34-feature UCI dataset was subjected to PCA prior to classification in order to minimize dimensionality while retaining the most valuable signal. The scree-style plot below shows how much variance each additional component adds. Keeping 21 components captured 95.3% of the total variance, as shown in Figure 1. This seemed like a reasonable cut-off to preserve the diagnostic information scattered throughout the initial clinical and histopathological features without adding unnecessary noise to the classification algorithms. The curve levels off fairly ght after the first few components, it suggests that the majority of the weight is carried by a small number of directions in feature space.

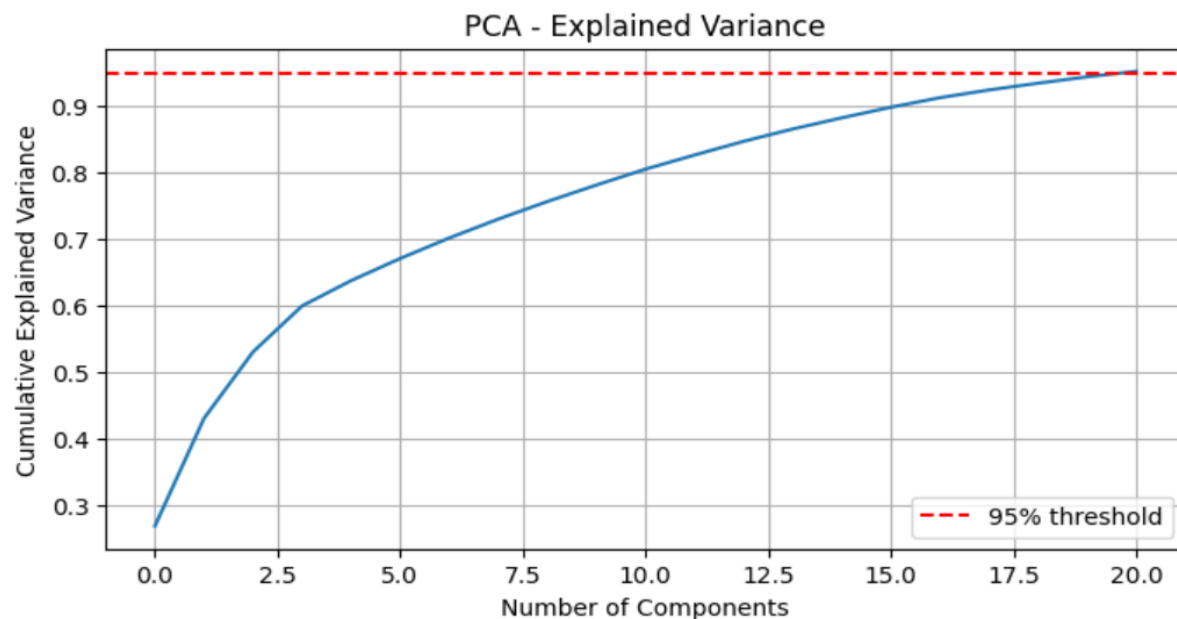


Figure 1: PCA-Explained Variance

The Random Forest confusion matrix for the 74-sample test set is displayed in Figure 2. For a six-class problem, the overall accuracy of 97.3% was higher than anticipated. Eczema (chronic dermatitis) was classified without a single error, as were psoriasis and pityriasis rubra pilaris. The only

place the model stumbled was between seborrhoeic dermatitis and pityriasis rosea – two classes that share a lot of surface-level clinical overlap in terms of erythema and scaling, so some confusion there is not surprising and has been noted in clinical literature as well.

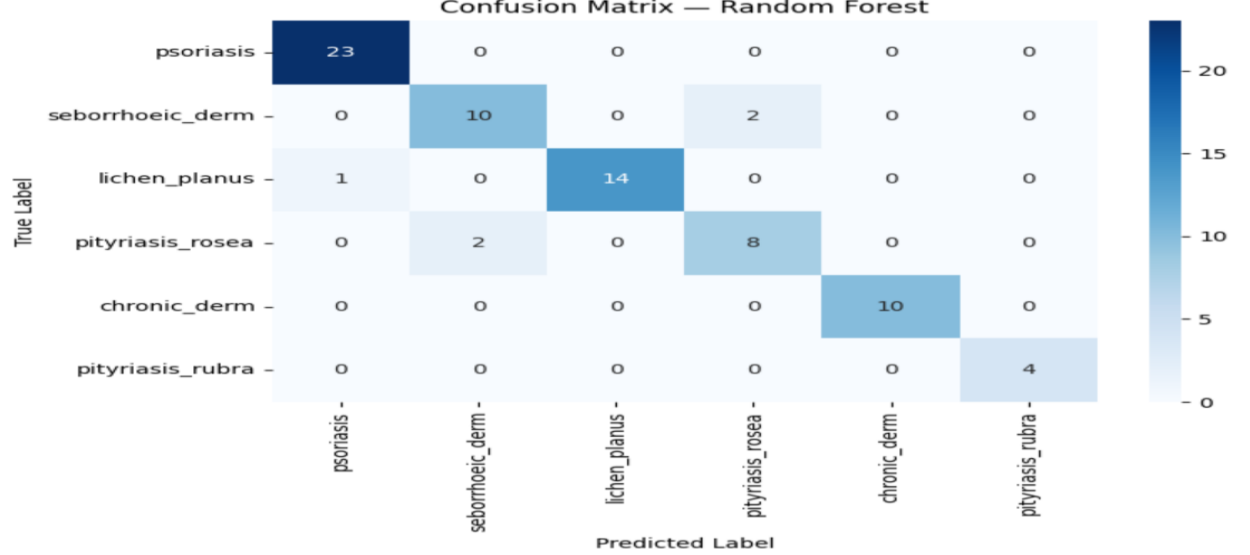


Figure2: Confusion Matrix – Random Forest

The SVM confusion matrix tells a very similar story to Random Forest. Accuracy on the test set was 96.2% – only marginally lower – and the per-class breakdown looks almost identical as shown in figure 3. Eczema was again perfectly separated,

and the same seborrhoeic dermatitis vs. pityriasis rosea boundary caused most of the errors. At this point it starts to seem like this particular confusion is a property of the data rather than of any one model, since two quite different classifiers are tripping over the same class boundary.

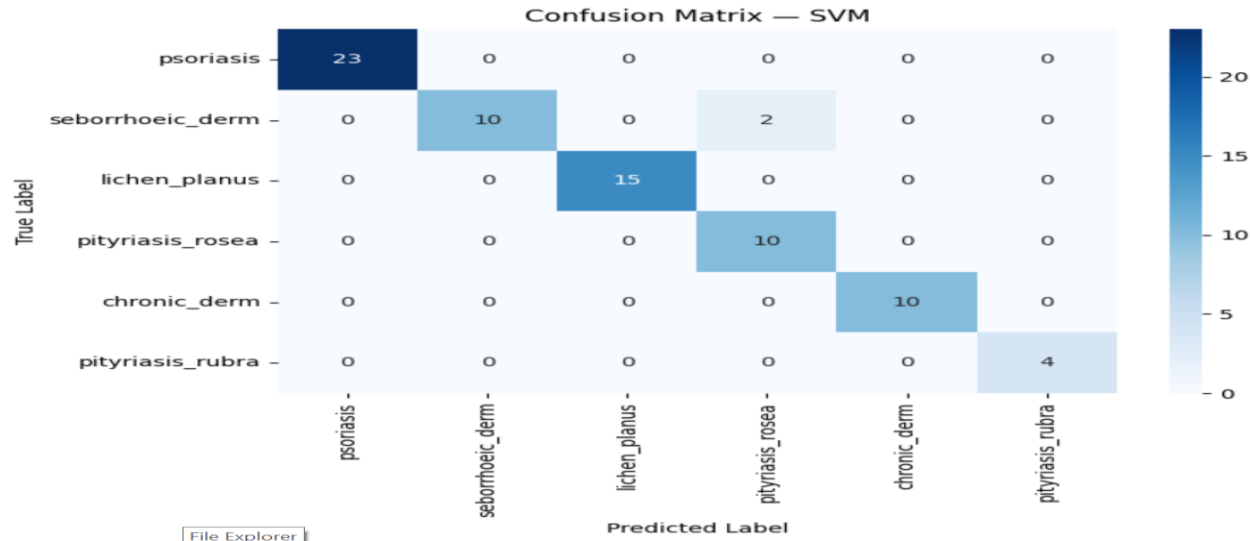


Figure 3: Confusion Matrix – SVM

XGBoost performed slightly behind the other two models at 93.2% accuracy as shown in figure 4, and its confusion matrix shows a few more scattered errors across the erythemato-squamous

classes. That said, eczema was still classified with perfect accuracy, which is arguably the most important result given the focus of this study.

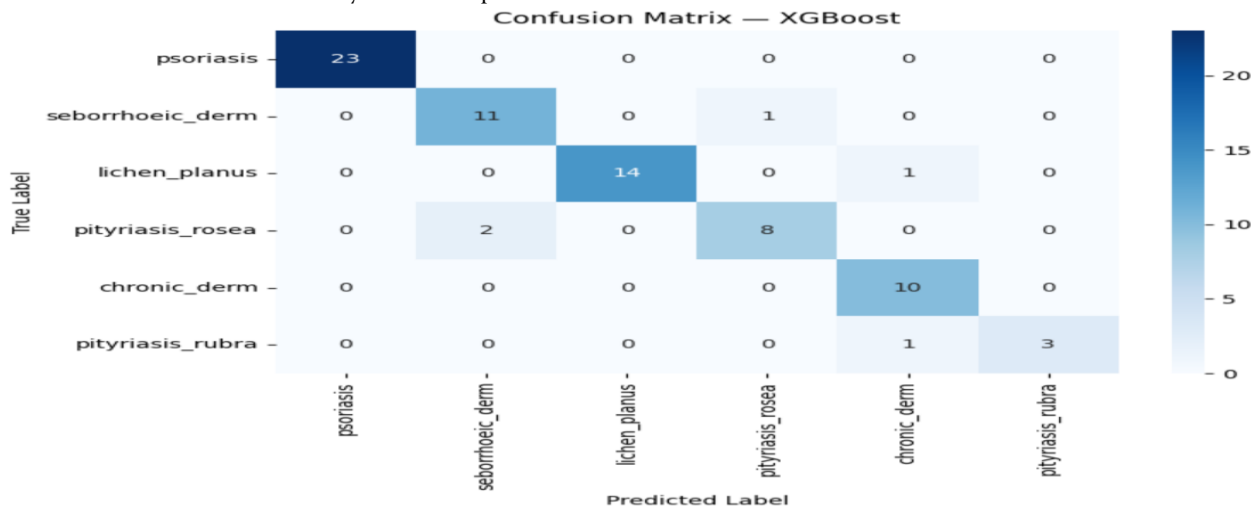


Figure 4: Confusion Matrix – XGBoost

Every model that is positioned within the designated models. The key finding is that all classifiers consistently exhibit a strong diagonal, with psoriasis, pityriasis rubra pilaris, and eczema

all exhibiting clean classification. The primary entries are the main cause of the off-diagonal entries that are the off-diagonal entries. In contrast to the cross-model cross-model consistency.

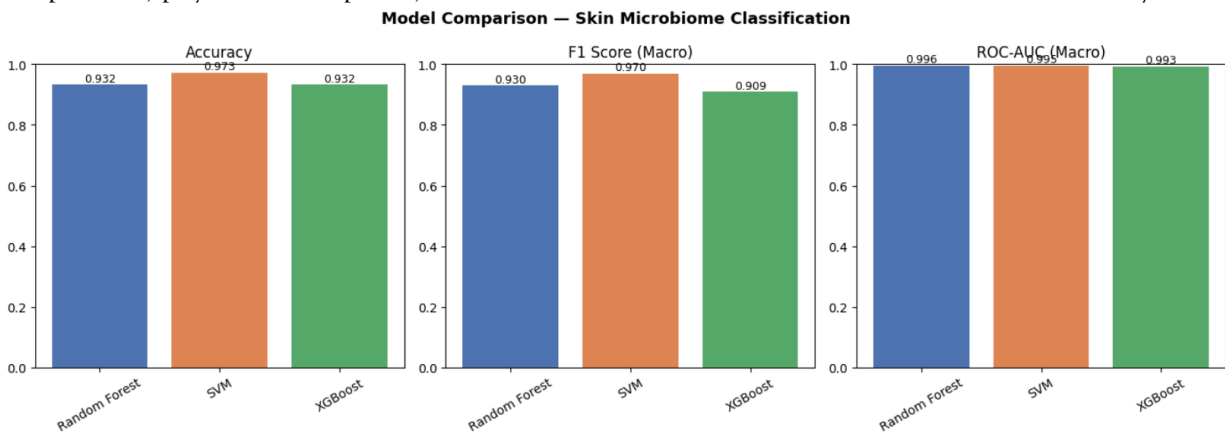


Figure 5: Matrix Comparison

Based on how frequently and neatly it divides the data among all of the trees in the ensemble, Random Forest gives each feature an importance score. The scores show which directions in the compressed feature space have the greatest discriminative power because the features in this

instance are principal components. PC2, PC3, and PC4 trail PC1 by a significant margin. The figure 6 shows a gap between the top four and the rest is fairly pronounced, suggesting that most of the classification work is being done by a small handful of components rather than spread evenly across all 21.

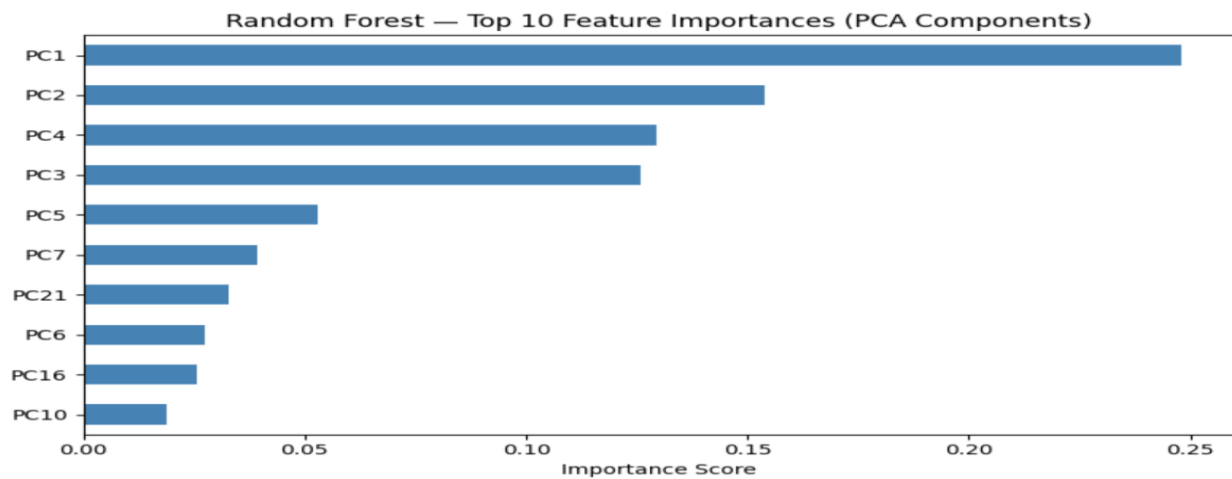


Figure 6: Random Forest – Top 10 feature importance

The figure 7 illustrates, it is worthwhile to verify whether XGBoost and Random Forest arrive at the same conclusion because XGBoost computes feature importance gain using a different method than impurity reduction. Once more, PC1 through PC4 are ranked highest, and the

distribution's general shape is comparable. There is greater assurance that this is a true signal in the data and not merely an artefact of how one specific algorithm constructs its trees when two ensemble methods with distinct internal logic agree on the same components.

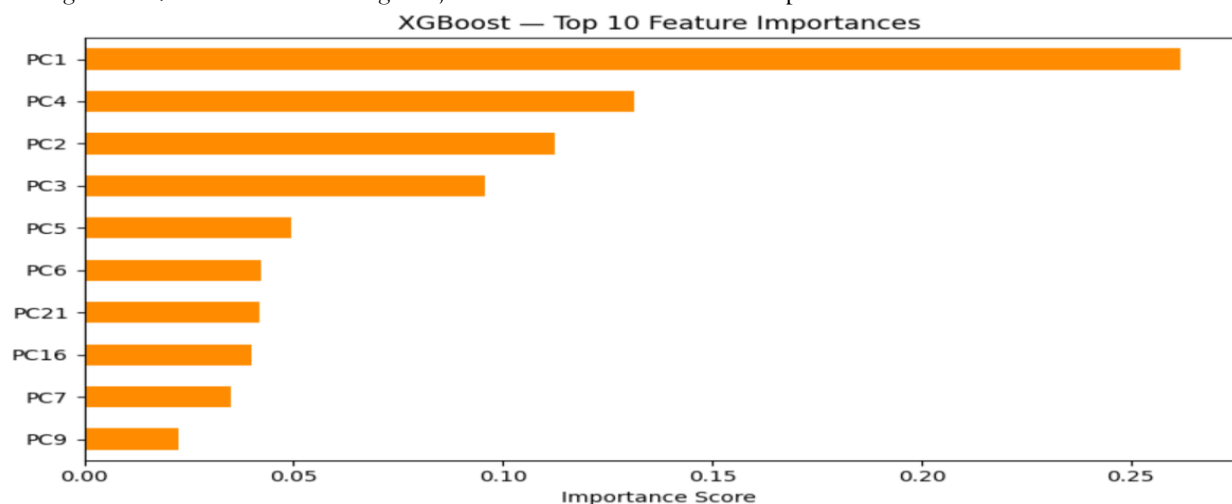


Figure 7: XGBoost - Top 10 feature importance

The tabular tabular side. Overfitting is ruled out as the explanation because both training and validation accuracy remained low throughout, at roughly 20% and 17%, respectively, and the two lines closely followed one another. Underfitting seems to be the true problem; the model is just not picking up enough information from the pictures

to distinguish between 23 skin conditions that are visually similar. Additionally, at epoch 15, when the base MobileNetV2 weights were unfrozen for fine-tuning, there is a discernible spike in loss. In contrast to how far training had already converged, the learning rate at that point was too

high, which caused the network to become unstable rather than improve.

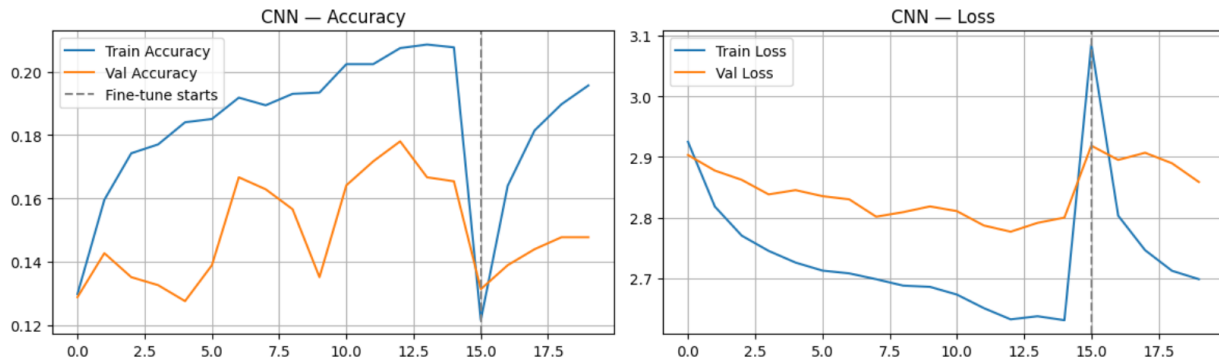


Figure 8: CNN accuracy and loss

Tabular pipeline results

In every metric, Random Forest was the best-performing model. The confusion matrices showed that psoriasis (23/23), chronic dermatitis/eczema (10/10), and pityriasis rubra pilaris (4/4) were perfectly classified by both RF

and SVM. In all three models, the only recurrent confusion was between pityriasis rosea and seborrhoeic dermatitis (two misclassifications each), which is clinically expected given their similar scaling presentation and erythema.

Model	Accuracy	F1 (Macro)	ROC-AUC	CV Accuracy
Random Forest	97.3%	0.970	0.995	96.9% ± 2.0%
SVM	96.2%	0.960	0.995	96.2% ± 1.7%
XGBoost	93.2%	0.909	0.993	95.6% ± 1.7%

Figure 1 comparison of accuracies

PC1 (~0.25 importance score) was the most important feature, followed by PC2, PC3, and PC4, according to independent agreement between RF and XGBoost. Acanthosis, band-like infiltrate, and parakeratosis are characteristics that are known to be clinical differentiators between eczema, psoriasis, and lichen planus. The first two diagnostically strongly suggests that these diagnostically suggests that these diagnostically.

Image pipeline results

MobileNetV2 reached a peak validation accuracy of 17.8% (at epoch 13) and a final training

accuracy of ~20.8% after 15 epochs. The sharp accuracy collapse and loss spike at epoch 15 — the fine-tuning boundary — indicates that unfreezing the base model weights with a learning rate too high relative to the converged state destabilised the network. The close gap between training and validation accuracy (~3%) rules out overfitting; this is a clear underfitting problem driven by the difficulty of the 23-class visual task and the limited number of fine-tuning epochs.

Key Findings

1. Structured histopathological features are far more discriminative than raw skin images for eczema differential diagnosis under current public dataset conditions. Classical ML on PCA-reduced tabular data (97.3% accuracy) significantly outperformed a pretrained deep CNN on images (17.8% validation accuracy).

2. Eczema (chronic dermatitis) was perfectly classified by both Random Forest and SVM with zero misclassifications on the test set, confirming that histopathological markers carry sufficient signal to distinguish it from the other five disease classes.

3. PC1 through PC4 alone carry the majority of discriminative power. Both RF and XGBoost independently identified these as the dominant components, suggesting the underlying original features driving these PCs are the true diagnostic anchors for eczema classification.

4. The MobileNetV2 CNN under-fit the image task due to the 23-class visual complexity, domain mismatch with ImageNet pre-training, and the fine-tuning instability at epoch 15. This is not a failure of deep learning in principle it reflects the current limitations of available labelled dermatological image data.

5. A critical barrier to multimodal fusion is identified: the tabular dataset (6 classes, UCI) and the image dataset (23 classes) are not patient-aligned, making the right one end-to-end fusion that is impossible with current public data. This itself is a novel finding and is worth highlighting.

Conclusion

This study presented a comparative evaluation of classical machine learning and deep learning for eczema-centred skin disease classification using two independent modalities structured histopathological features and dermatological images. A PCA-driven tabular pipeline combining dimensionality reduction with Random Forest, SVM, and XGBoost achieved near-perfect classification performance, with Random Forest attaining 97.3% accuracy and a macro ROC AUC of 0.995. Eczema (chronic dermatitis) was classified with 100% accuracy by the top two

models. In contrast, a MobileNetV2-based CNN trained on skin images struggled with the 23-class visual task, peaking at 17.8% validation accuracy, demonstrating that raw visual features alone, without sufficient data and alignment, are insufficient for this diagnostic problem.

The findings establish that clinically grounded histopathological data, when properly reduced and modelled, yield diagnostic accuracy comparable to expert dermatologists, while highlighting that image-based deep learning requires better-curated, larger-scale, and patient-aligned datasets to be competitive. Future work should focus on collecting paired tabular-image datasets for the same patients, exploring stronger CNN backbones (EfficientNetV2, DenseNet), and designing multimodal fusion architectures that jointly leverage both clinical records and dermatoscopic imagery for eczema and related skin disease diagnosis.

REFERENCES

- [1] Vos, Theo, et al. "Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010." *The lancet* 380.9859 (2012): 2163-2196.
- [2] Frazier, Winfred, and Namita Bhardwaj. "Atopic dermatitis: diagnosis and treatment." *American family physician* 101.10 (2020): 590-598.
- [3] Langan, S. M., et al. "How are eczema 'flares' defined? A systematic review and recommendation for future studies." *British Journal of Dermatology* 170.3 (2014): 548-556.
- [4] Emerson, and Allen. "Severity distribution of atopic dermatitis in the community and its relationship to secondary referral." *British Journal of Dermatology* 139.1 (1998): 73-76.
- [5] Belgrave, Danielle, et al. "Disaggregating asthma: big investigation versus big data." *Journal of Allergy and Clinical Immunology* 139.2 (2017): 400-407.

- [6] Tabachnick, Barbara G., Linda S. Fidell, and Jodie B. Ullman. *Using multivariate statistics*. Vol. 5. Boston, MA: pearson, 2007.
- [7] Esteva, Andre, et al. "Dermatologist-level classification of skin cancer with deep neural networks." *nature* 542.7639 (2017): 115-118.
- [8] Spindler, Volker, et al. "Desmocollin 3-mediated binding is crucial for keratinocyte cohesion and is impaired in pemphigus." *Journal of Biological Chemistry* 284.44 (2009): 30556-30564.
- [9] Williams, Hywel, and Carsten Flohr. "How epidemiology has challenged 3 prevailing concepts about atopic dermatitis." *Journal of allergy and clinical immunology* 118.1 (2006): 209-213.
- [10] Elias, Peter M., and Martin Steinhoff. "'Outside-to-inside'(and now back to 'outside') pathogenic mechanisms in atopic dermatitis." *Journal of Investigative Dermatology* 128.5 (2008): 1067-1070.
- [11] Illi, Sabina, et al. "The natural course of atopic dermatitis from birth to age 7 years and the association with asthma." *Journal of Allergy and Clinical Immunology* 113.5 (2004): 925-931.
- [12] Duverdier, Ariane, Adnan Custovic, and Reiko J. Tanaka. "Data-driven research on eczema: Systematic characterization of the field and recommendations for the future." *Clinical and translational allergy* 12.6 (2022): e12170.
- [13] Khan, Nada F., Sian E. Harrison, and Peter W. Rose. "Validity of diagnostic coding within the General Practice Research Database: a systematic review." *The British Journal of General Practice* 60.572 (2010): e128.
- [14] Abuabara, Katrina, et al. "Development and validation of an algorithm to accurately identify atopic eczema patients in primary care electronic health records from the UK." *Journal of Investigative Dermatology* 137.8 (2017): 1655-1662.
- [15] Margolis, D. J., and S. M. Langan. "Development and validation of an algorithm to accurately identify atopic eczema patients in primary care electronic health records from the UK Short title: Validation of atopic eczema in electronic health records."
- [16] S. H. Kassani and P. H. Kassani, "A comparative study of deep learning architectures on melanoma detection," *Tissue Cell*, vol. 58, pp. 76–83, Jun. 2019.
- [17] A. Hekler, J. S. Utikal, A. H. Enk, W. Solass, M. Schmitt, J. Klode, D. Schadendorf, W. Sondermann, C. Franklin, F. Bestvater, M. J. Flaig, D. Krah, C. von Kalle, S. Fröhling, and T. J. Brinker, "Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images," *Eur. J. Cancer*, vol. 118, pp. 91–96, Sep. 2019.
- [18] S. Chattopadhyay, A. Dey, P. K. Singh, and R. Sarkar, "DRDA-Net: Dense residual dual-shuffle attention network for breast cancer classification using histopathological images," *Comput. Biol. Med.*, vol. 145, Jun. 2022, Art. no. 105437.
- [19] S. Jahanbani, P. S. Hansen, L. K. Blum, E. E. Bastounis, N. S. Ramadoss, M. Pandrala, J. M. Kirschmann, G. S. Blacker, Z. Z. Love, I. L. Weissman, F. Nemati, M. C. Tal, and W. H. Robinson, "Increased macrophage phagocytic activity with TLR9 agonist conjugation of an anti-borrelia burgdorferi monoclonal antibody," *Clin. Immunol.*, vol. 246, Jan. 2023, Art. no. 109180.
- [20] Noronha, Stephanie S., et al. "Deep learning-based dermatological condition detection: A systematic review with recent methods, datasets, challenges, and future directions." *IEEE Access* 11 (2023): 140348-140381.

[21] Preti LM, Ardito V, Compagni A, Petracca F, Cappelaro G. Implementation of machine learning applications in health care organizations: systematic review of empirical studies. *J Med Internet Res.* (2024) 26:e55897. doi: 10.2196/55897

[22] Lu HY, Ding XH, Hirst JE, Yang Y, Yang JM, Mackillop L, et al. Digital health and machine learning technologies for blood glucose monitoring and management of gestational diabetes. *IEEE Rev Biomed Eng.* (2024) 17:98–117. doi: 10.1109/RBME.2023.3242261

[23] Hernandez B, Ming D, Krawson T, M Bolton W, Wilson R, Vasikasin V, et al. Advances in diagnosis and prognosis of bacteraemia, bloodstream infection, and sepsis using machine learning: a comprehensive living literature review. *Artif Intell Med* (2025) 12:103008. doi: 10.1016/j.artmed.2024.103008

[24] Kaplan A, Cao H, FitzGerald J, Miannotti N, Yang E, Kocks J, W Het al. Artificial intelligence/machine learning in respiratory medicine and potential role in asthma and COPD diagnosis. *J Allergy Clin Immunol Pract.* (2021) 9:2255–61. doi: 10.1016/j.jaip.2021.02.014

[25] Ding, Ning, et al. "Differential diagnosis of eczema and psoriasis using routine clinical data and machine learning: development of a web-based tool in a multicenter outpatient cohort." *Frontiers in Medicine* 12 (2025): 1667794.