

A Hybrid Deep Learning Model for Multiclass Skin Cancer Classification Using ConvNeXtV2 and Separable Self-Attention Mechanisms

Researcher: Mishaal Mishaan Al-Otaibi

Supervisor : DR. Hager Saleh

College of Informatics — Midocean University

Year: 2025

Introduction

- Skin cancer is among the most common and deadly cancers worldwide.
- Early detection dramatically improves survival, but manual diagnosis is difficult due to visual similarity between lesions.
- AI and deep learning now enable highly accurate automated diagnosis from dermoscopic images.
- This research proposes a hybrid deep learning architecture combining CNNs (ConvNeXtV2) and separable self-attention to enhance accuracy and efficiency.

Research Problem

- Skin lesions share highly similar textures and colors → causing misdiagnosis.
- Datasets are imbalanced, small, and contain artifacts (hair, lighting).
- CNNs capture local features but struggle with long-range dependencies.
- Transformers capture global patterns but are computationally heavy.

→ Problem: How to design a hybrid model that is accurate, generalizable, and computationally efficient for multiclass skin cancer classification?

Research Objectives

Review current deep learning approaches for skin cancer detection.

Develop a hybrid ConvNeXtV2 + Separable Self-Attention model.

Improve generalization using transfer learning & augmentation.

Evaluate performance through 5-fold stratified cross-validation.

Compare with state-of-the-art CNN and ViT models.

Main Research Question : How can we design a hybrid deep learning architecture that improves accuracy, generalization, and computational efficiency in multiclass skin lesion classification?

Research Contributions



Novel Hybrid Architecture combining:

- ConvNeXtV2 convolutional blocks
- Separable Self-Attention



Optimized transfer learning strategy



Robust preprocessing pipeline (hair removal, normalization, augmentation)



Balanced dataset using oversampling

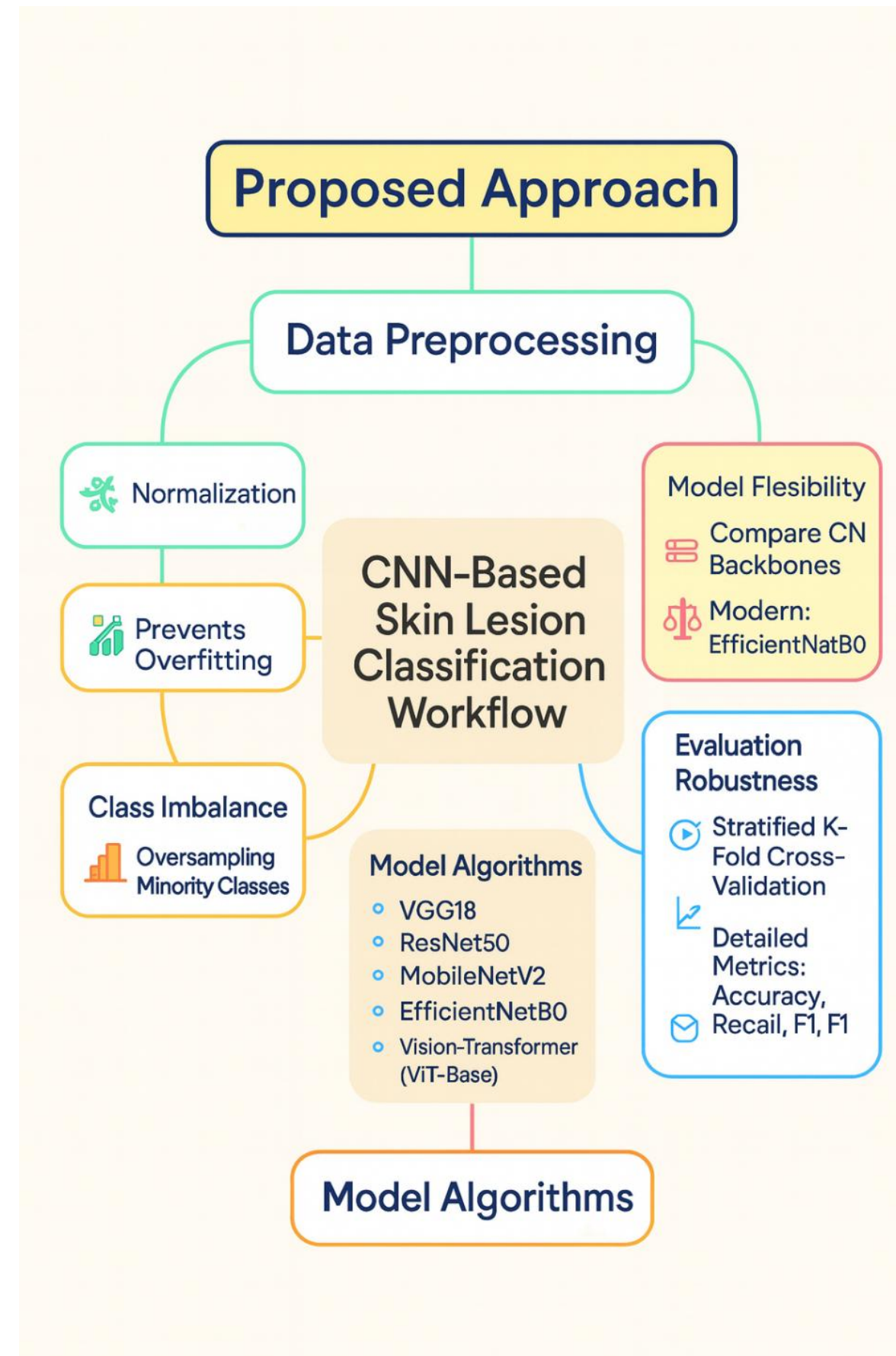


High-accuracy classification outperforming 10+ CNN & ViT models.

Related Work

Paper	Year	Methods	Key Techniques	Results	Dataset
[1]	2025	Hybrid ConvNeXtV2 + Separable Self-Attention; Data augmentation; Transfer learning	Combines modern CNN blocks with separable attention to reduce complexity	Acc: 93.48%, Prec: 93.24%, Rec: 90.70%, F1: 91.82%	ISIC 2019
[2]	2025	LSTM + Time-distributed CNN	Merges temporal and spatial features for melanoma classification	Improved accuracy, precision, recall, F1 over CNN-only	HAM10000
[3]	2023	DSCC_Net CNN; SMOTE Tomek sampling	Class imbalance handled via SMOTE Tomek	AUC: 99.43%, Acc: 94.17%, Rec: 93.76%, Prec: 94.28%, F1: 93.93%	ISIC 2020, HAM10000, DermIS
[4]	2021	LCcHIV preprocessing; Custom CNN; IMFO optimization; MMCA feature fusion; KELM classification	Image enhancement + multi-feature extraction/fusion + optimization	Seg Acc: 98.70% (PH2), Class Acc: 90.67% (HAM10000)	ISBI 2016/2017, ISIC 2018, PH2, HAM10000
[5]	2022	Transfer learning; Hybrid Whale Optimization; Feature fusion (CCA); ELM	Optimal feature selection and fusion before classification	Acc: 93.40% (HAM10000), 94.36% (ISIC 2018)	HAM10000, ISIC 2018
[6]	2024	Ensemble (VGG16, Inception V3, ResNet 50); Oversampling	Combines CNN outputs with balanced data	Acc: 91% (orig), 97% (balanced)	ISIC 2018, HAM10000
[7]	2024	Customized transfer learning + Triple attention; ML-IGPA fusion	Three attention mechanisms + multi-level fusion	Acc: 94.93%	HAM10000
[8]	2024	Preprocessing; U-Net segmentation; GLCM & RGB histogram features; MLSTM + SC-STBO optimization	Combines color & texture features with optimized LSTM	Acc: 99.31%, Rec: 98.25%	ISIC
[12]	2024	LesionNet (image SR + deep pretrained CNNs)	DenseNet201 backbone + HoG descriptors + super-resolution preprocessing	Higher accuracy and robustness vs. CNNs without SR	Benchmark dermoscopic datasets

Proposed Approach



Dataset

Dataset

Seven classes:

- Melanoma, Melanocytic Nevi, BCC, BKL, AKIEC, DF, VASC
- 10,000+ dermoscopic images; 1,000 images sampled in your experiment
- After augmentation: 4,000 balanced images

Preprocessing Pipeline

Preprocessing Pipeline

Hair removal via morphological Blackhat

+ Telea inpainting

Resize to 224×224

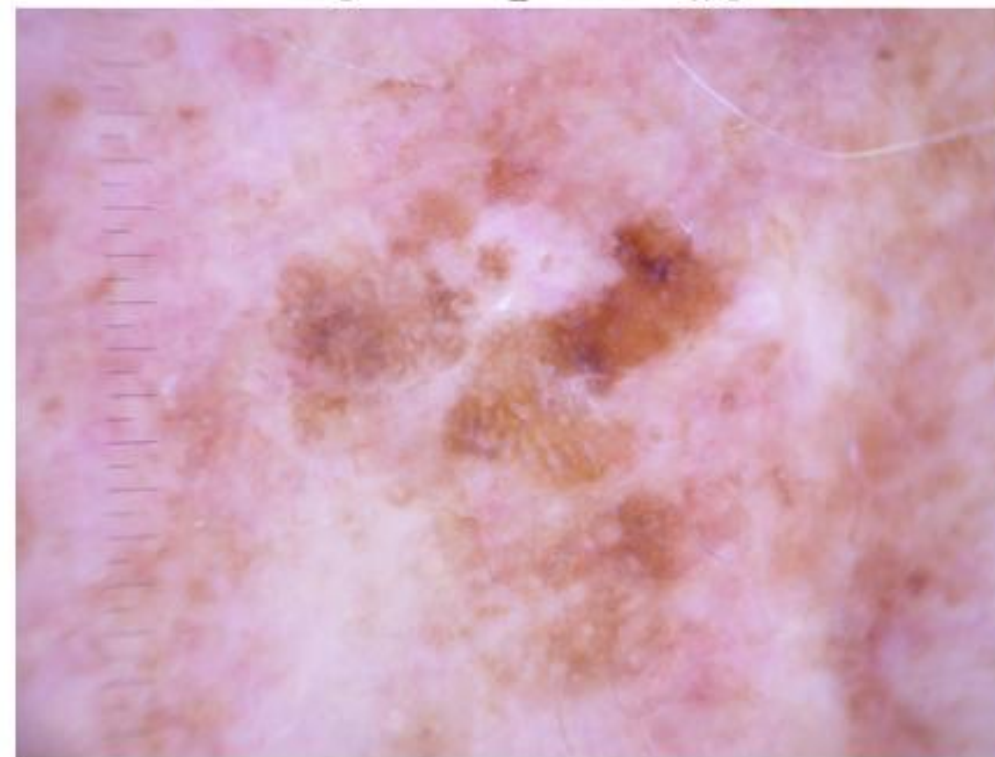
Normalization (0–1)

Augmentation:

- Rotation $\pm 25^\circ$
- Zoom $\pm 10\%$
- Brightness $\pm 20\%$
- Shifts & flips

Result: **4X enhanced dataset** and reduced overfitting.

Original: ISIC_0027419.jpg



Processed: ISIC_0027419.jpg



Model Architecture and Algorithms

State-of-the-art architectures were used with pretrained ImageNet weights to improve feature extraction and performance on limited dermoscopic data.

- VGG16
Simple 16-layer CNN that captures basic edges and progressively deeper visual patterns.
- ResNet50
Uses residual (skip) connections to prevent vanishing gradients and enable deeper, more stable training.
- MobileNetV2
Lightweight model using depthwise separable convolutions; ideal for low-resource or mobile environments.
- EfficientNetB0
Balances depth, width, and resolution through compound scaling, achieving high accuracy with fewer parameters.
- Vision Transformer (ViT-Base)
Splits images into patches and applies multi-head self-attention to capture global relationships and subtle lesion patterns.

Proposed Model Architecture

Hybrid Framework = ConvNeXtV2 + Separable Self-Attention

- ConvNeXtV2 extracts fine-grained local features
- Attention captures long-range dependencies
- Lightweight separable attention reduces computational cost
- Global Average Pooling + Dense layers + Softmax (7 classes)

Proposed Model Architecture

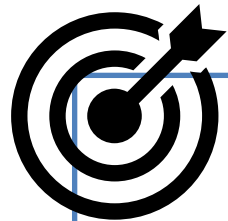
Training Strategy

- Transfer Learning (ImageNet weights)
- Freeze base layers → train top layers
→ fine-tune entire model
- Adam optimizer, $lr = 1e-4$
- Batch size = 32
- Early stopping
- 5-Fold Stratified Cross-Validation
ensures reliable evaluation

Evaluation Metrics

- Accuracy
- Precision (macro)
- Recall (macro)
- F1-Score
- ROC-AUC

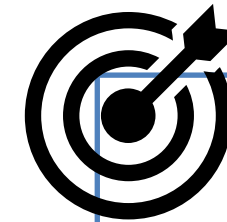
Results



Quantitative Results

- **Accuracy:** 93.5%
- **Precision:** 93.1%
- **Recall:** 91.2%
- **F1-Score:** 92.2%
- **ROC-AUC:** 0.957

Demonstrates strong balance between sensitivity & specificity.



Class-wise Findings

High performance on critical classes:

- **Melanoma:** F1 = 93.68%
- **BCC:** F1 = 94.05%
- **VASC:** F1 = 94.59%

Even visually similar classes (NV, BKL) were accurately separated due to attention module.

Comparison with Baseline

Models

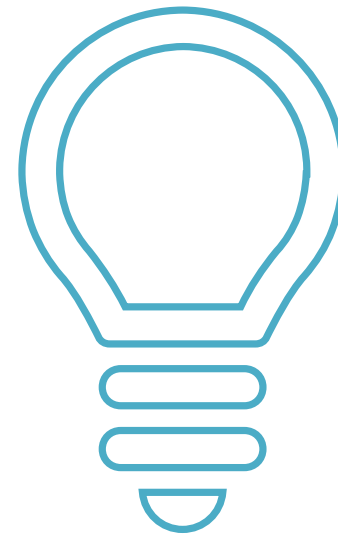
Model	Accuracy	Parameters
VGG16	88.2%	138M
ResNet50	90.1%	25.6M
EfficientNetB0	91.5%	5.3M
ViT-Base	92.0%	86M
Hybrid Model	93.5%	21.9M

Discussion

Hybrid model benefits:

- Local + global feature synergy
- Lower computational cost than transformers
- Better interpretability using Grad-CAM

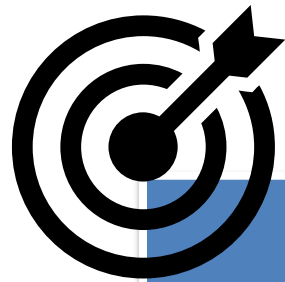
Balanced dataset prevents bias toward
common classes



Transfer learning significantly improves
convergence

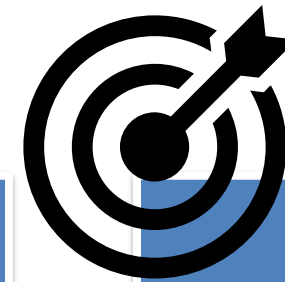
Strong generalization confirmed by cross-
validation

Limitations



- Only one dataset (HAM10000) used
- No real clinical testing
- Requires GPU hardware
- Limited explainability tools (only Grad-CAM)
- No multimodal inputs (e.g., patient metadata)

Future Work



- Test on more dermatology datasets
- Build mobile/edge-deployment version
- Integrate patient metadata (age, location, history)
- Add SHAP / LIME for deeper interpretability
- Combine segmentation + classification
- Conduct clinical trials for real-world validation
- Improve fairness, privacy, and bias handling

Conclusion

- A lightweight high-accuracy hybrid model was successfully developed.
- Combines CNN + Attention for superior feature representation.
- Outperforms classical CNNs and ViT architectures.
- Suitable for tele-dermatology and real-time decision support.
- A strong step toward reliable AI-assisted skin cancer diagnosis.

Reference

- 1) Ozdemir, B., & Pacal, I. (2025). A robust deep learning framework for multiclass skin cancer classification. Scientific Reports, 15, 4938. <https://doi.org/10.1038/s41598-025-89230-7>
- 2) Abohashish, S. M. (2025). Enhanced melanoma and non-melanoma skin cancer classification using a hybrid LSTM-CNN model. Scientific Reports, 15, 24994. <https://doi.org/10.1038/s41598-025-08954-8>
- 3) Tahir, M., Jin, H., Zafar, A., Mahmood, M. T., Ahmed, I., & Ahmad, H. F. (2023). DSCC_Net: Multi-Classification Deep Learning Models for Diagnosing Skin Cancer Using Dermoscopic Images. Cancers, 15(7), 2179. <https://www.mdpi.com/2072-6694/15/7/2179>
- 4) Khan, M. A., Sharif, M., Akram, T., Damaševičius, R., & Maskeliūnas, R. (2021). Skin Lesion Segmentation and Multiclass Classification Using Deep Learning Features and Improved Moth Flame Optimization. Diagnostics, 11(5), 811. <https://www.mdpi.com/2075-4418/11/5/811>
- 5) Afza, F., Sharif, M., Khan, M. A., Kadry, S., & Damaševičius, R. (2022). Melanoma classification using deep learned features boosted with extreme learning machine. Sensors, 22(3), 799. <https://www.mdpi.com/1424-8220/22/3/799>
- 6) Thwin, S. M., & Park, H.-S. (2024). Skin Lesion Classification Using a Deep Ensemble Model. Applied Sciences, 14(13), 5599. <https://doi.org/10.3390/app14135599>
- 7) Efat, A. H., Hasan, S. M. M., Uddin, M. P., & Mamun, M. A. A. (2024). A multi-level ensemble approach for skin lesion classification using customized transfer learning with triple attention. PLOS ONE, 19(10), e0309430. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0309430>
- 8) Gomathi, S., & Arunachalam, N. (2024). Skin lesion prediction and classification using modified LSTM with a hybrid SC-STBO optimizer. International Journal of Computational Intelligence Systems, 17, 186. <https://link.springer.com/article/10.1007/s44196-024-00599-1>

THANK YOU