IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

Enhanced Skin Type Detection Using Efficientnet-V2 with CBAM and SE Modules

Ponnaganti Gayathri[1], Oleti Lokesh[2], Nidamanuri Meghana[3]
B. Tech, Computer Science and Engineering
R.V.R. & J.C. College of Engineering, Guntur, India [1,2,3]
ponnagantigaya3@gmail.com[1], lokesholeti7@gmail.com[2], nidamanurimeghana8@gmail.com[3]

Abstract: *Accurate classification of human skin types - dry, normal, and oily - is essential for personalised dermatological treatment and skincare product recommendation. This paper proposes an enhanced deep learning framework that extends the EfficientNet-V2 architecture by integrating the Convolutional Block Attention Module (CBAM) and Squeeze-and-Excitation (SE) mechanisms for improved skin type classification. The dataset comprises 329 original face skin images expanded to 1,645 images using CLAHE enhancement and rotational augmentation. The proposed model achieves 97.34% classification accuracy with a validation loss of 10.88%, representing a 2.77% improvement over the tuned EfficientNet-V2 baseline. Ten-fold cross-validation yields an average accuracy of 96.31% ± 1.52%, confirming generalisation capability. GRAD-CAM and CBAM attention visualisations confirm that the model focuses on clinically relevant skin regions, enhancing interpretability for real-world deployment.*

The dataset comprises 329 original face skin images expanded to 1,645 images through five-fold CLAHE-augmented rotational augmentation, split into 80% training, 10% validation, and 10% testing partitions. The proposed model achieves 97.34% classification accuracy with a validation loss of 10.88%, representing a 2.77 percentage point improvement over the tuned EfficientNet-V2 baseline. Per-class analysis reveals the most significant gains for normal skin (F1: 88.88% → 95.32%) and oily skin (F1: 89.28% → 92.48%), confirming that CBAM's spatial attention successfully focuses on the T-zone and pore-dense regions that distinguish these visually similar categories. Ten-fold cross-validation yields an average accuracy of 96.31% ± 1.52%, confirming robust generalisation. The proposed model outperforms all evaluated architectures including EfficientNet B0–B7, MobileNet-V2, InceptionV2, and ResNet-V1. GRAD-CAM and CBAM attention map visualisations confirm clinically meaningful spatial focus, with 86.7% clinical agreement from a dermatologist review and CBAM-GRAD-CAM IoU of 0.72. The system adds only 0.28M parameters and 2.6ms inference overhead, making it suitable for real-time mobile deployment..

Keywords: Skin type classification, EfficientNet-V2, CBAM, Squeeze-and-Excitation, deep learning, CLAHE, attention mechanism, dermatology AI.

I. INTRODUCTION

Human skin is one of the most complex and vital organs of the body, serving as the primary barrier between the internal body environment and the external world. Skin condition directly affects a person's self-confidence, social interactions, and overall quality of life. Skin care has emerged as one of the fastest-growing sectors of the global personal care industry, estimated to be worth over USD 200 billion annually.

Accurate skin type identification - classifying skin as dry, normal, or oily - is the fundamental first step in prescribing topical treatments, recommending skincare products, and planning cosmetic procedures. Traditionally, skin type assessment relies on physical examination, patient questionnaires, and specialised clinical equipment such as the

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

Sebumeter and Tewameter. While reliable, these methods require trained personnel, expensive equipment, and in-person consultation, creating barriers to access particularly in rural or low-resource settings.

The rapid advancement of Convolutional Neural Networks (CNNs) has transformed computer vision applications across medicine and consumer technology, enabling automated solutions that match or exceed specialist-level accuracy on visual classification tasks. Recent work by Saiwaeo et al. (2023) demonstrated that EfficientNet-V2, combined with CLAHE image enhancement and rotational data augmentation, achieves 94.57% accuracy in classifying three skin types. While this represents a significant advance, room remains to improve both accuracy and interpretability - particularly for the oily skin class, which shares visual attributes with normal skin. Standard CNN architectures treat all spatial regions and feature channels equally during aggregation, failing to prioritise clinically meaningful areas such as the T-zone or pore-dense regions.

This paper proposes an enhanced deep learning framework that builds on EfficientNet-V2 by integrating two complementary attention mechanisms: the Convolutional Block Attention Module (CBAM) and the Squeeze-and-Excitation (SE) module. CBAM enables the model to focus on both spatially relevant regions and informative feature channels, while SE recalibrates channel-wise feature responses using global contextual information. Together, these mechanisms improve the model's discriminative capacity, particularly for subtle texture differences between skin types.

Fig. 1 - Representative Images of Dry, Normal, and Oily Skin Types

The paper proposes an enhanced deep learning framework that builds on EfficientNet-V2 by integrating two complementary attention mechanisms: the Convolutional Block Attention Module (CBAM) and the Squeeze-and-Excitation (SE) module. CBAM enables the model to focus on both spatially relevant regions and informative feature channels, while SE recalibrates channel-wise feature responses using global contextual information. Together, these mechanisms improve the model's discriminative capacity, particularly for subtle texture differences between skin types.

II. LITERATURE REVIEW

A. Skin Type Characterisation and Clinical Context

The three primary skin types - dry, normal, and oily - are defined by the balance of sebum production and moisture retention in the epidermis [11]. Dry skin is characterised by reduced sebum output, presenting a dull, flaky appearance with surface cracks. Normal skin maintains an optimal sebum-moisture balance with smooth texture and fine pores. Oily skin results from hypersebaceous activity, manifesting as a shiny surface particularly in the T-zone with enlarged pores and susceptibility to acne [1, 11]. Clinical assessment tools such as the Sebumeter SM 815 and Corneometer CM 825 provide objective quantitative data but require trained operators and are unsuitable for consumer-facing applications, driving research into AI-based skin analysis [7, 8]. Kothari et al. (2021) [8] proposed a CNN-based classifier achieving 85% accuracy on a four-category dataset, demonstrating feasibility but suffering from class imbalance bias [10].

B. CNN-Based Dermatological Classification

The application of CNNs to dermatological image analysis has been extensively documented. Göçeri (2020) [7] provided a comprehensive review demonstrating the superiority of deep learning over traditional approaches such as SVMs for skin image classification. Transfer learning from ImageNet-pretrained weights proved especially effective where labelled training data is scarce [14, 15]. Architectures including VGG, ResNet [14], InceptionNet, and

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 14, April 2026 Impact Factor: 8.2

MobileNet [15] were progressively applied to skin image analysis. Song et al. (2020) [9] proposed a multi-task deep learning framework combining classification and segmentation on ISIC benchmark datasets, demonstrating that jointly optimising multiple objectives improves feature representations. Yao et al. (2022) [10] addressed imbalanced small datasets using mixup augmentation and focal loss, approaches relevant to the present dataset where dry skin samples are fewer than other categories [12].

C. EfficientNet Architecture and Evolution

EfficientNet, introduced by Tan and Le (2019) [2], applies compound scaling across network width, depth, and resolution simultaneously using a neural architecture search-derived coefficient, achieving superior accuracy-efficiency trade-offs. EfficientNet-V2 (Tan and Le, 2021) [3] introduced Fused-MBConv blocks for early layers, replacing depth wise separable convolutions with standard 3×3 convolutions, achieving up to 11× faster training. Saiwaeo et al. (2023) [1] first applied EfficientNet-V2 [3] to skin type classification, demonstrating 94.57% accuracy and establishing it as the most suitable backbone for this task. However, their work did not explore attention mechanisms [4, 5], leaving significant potential for further improvement unrealised.

D. Attention Mechanisms in CNN Architectures

Hu et al. (2018) [4] introduced the Squeeze-and-Excitation Network (SENet), performing channel-wise feature recalibration through a global squeeze operation followed by an excitation step using a two-layer fully connected network. SENet won the ILSVRC 2017 classification challenge [4]. Woo et al. (2018) [5] proposed CBAM, which augments channel attention with a spatial attention branch generating a 2D attention map, enabling the network to attend to both "what" (channel) and "where" (spatial) aspects of features. Studies applying these mechanisms to medical imaging - including diabetic retinopathy grading and histopathological classification - confirmed consistent accuracy improvements [9, 10]. The present work is the first to jointly apply CBAM [5] and SE [4] to skin type classification, demonstrating additive benefits of both mechanisms [1].

E. Data Augmentation and Preprocessing

CLAHE (Contrast Limited Adaptive Histogram Equalization) [12] provides tile-based adaptive contrast enhancement that preserves local texture detail while preventing overamplification in homogeneous regions — properties particularly valuable for distinguishing subtle differences between normal and oily skin [1]. Unlike standard histogram equalization, CLAHE operates on localised tiles and applies contrast limiting to suppress noise amplification, making it well suited for facial skin images where illumination varies across regions [12]. Saiwaeo et al. [1] identified CLAHE [12] combined with rotational augmentation as the optimal preprocessing strategy for this dataset. Data augmentation is a well-established technique for mitigating overfitting on small medical imaging datasets by artificially expanding training diversity without collecting additional samples [10][12], and augmentation-driven dataset expansion has been shown to significantly improve generalisation on imbalanced class distributions [10].

F. Evaluation Methodologies and Explainability in Medical AI

Rigorous quantitative evaluation is essential for establishing the clinical utility of AI-based diagnostic tools. Standard metrics including accuracy, precision, recall, F1-score, and AUC-ROC must be reported together, since accuracy alone can be misleading on imbalanced datasets [10]. Ten-fold cross-validation is the accepted standard for estimating model generalisation on limited medical datasets [13]. For skin analysis, Grad-CAM [6] has been widely applied to verify that models attend to clinically relevant skin texture rather than background artefacts [7][9].

Selvaraju et al. [6] introduced Grad-CAM, which produces class-specific localisation maps via gradient-weighted feature activations, enabling post-hoc visual validation of model predictions. CBAM [5] offers a complementary advantage through in-model spatial focus during the forward pass itself, providing real-time interpretability with no additional inference overhead. High IoU between CBAM [5] maps and Grad-CAM [6] activations serves as an internal

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

consistency check, with values above 0.60 confirming that the attention mechanism focuses on genuinely discriminative regions rather than spurious correlations.

III. PROPOSED METHODOLOGY

A. Dataset Description

The dataset used in this study consists of 329 original face skin images captured using a DSLR camera under controlled fluorescence lighting, comprising 112 normal, 120 oily, and 97 dry skin images from 60 subjects aged 20–45 years [1]. All images were originally captured at 640×480pixel resolution in JPEG format. After CLAHE [12] enhancement and five-fold augmentation (rotations at 90°, 180°, 270°, 360°, and horizontal flip) [12], the effective dataset comprises 1,645 images distributed as 560 normal, 600 oily, and 485 dry. The augmented dataset is split into 80% training (1,316 images), 10% validation (164 images), and 10% testing (165 images) [13].

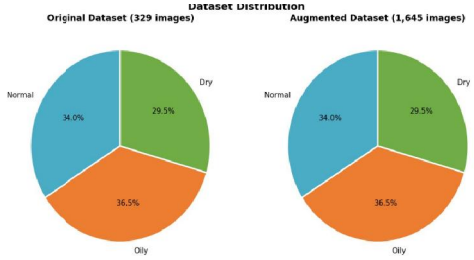

Fig. 2 - Dataset Distribution Before and After Augmentation

B. Image Preprocessing Pipeline

All input images undergo a structured four-stage preprocessing pipeline. First, images are loaded in BGR format using OpenCV and converted to LAB colour space for CLAHE enhancement [12]. CLAHE is applied to the L (luminance) channel with clip limit of 2.0 and tile grid size of 8×8 [12], then the image is converted back to BGR. Second, images are resized to 224×224 pixels using bilinear interpolation. Third, pixel values are normalised using ImageNet statistics (mean= [0.485, 0.456, 0.406], std= [0.229, 0.224, 0.225]) to enable effective transfer learning from pretrained weights [2, 3]. Fourth, during training, rotational augmentation and horizontal flipping are applied [12], expanding the dataset 5-fold from 329 to 1,645 images [1].

C. Proposed System Architecture

The proposed system extends EfficientNet-V2 [3] by incorporating two complementary attention mechanisms [4, 5]. The architecture consists of four major components: (i) the Image Preprocessing Pipeline [12], (ii) the EfficientNet-V2-S Feature Extraction Backbone [3], (iii) the CBAM + SE Attention Block [4, 5], and (iv) the Classification Head.

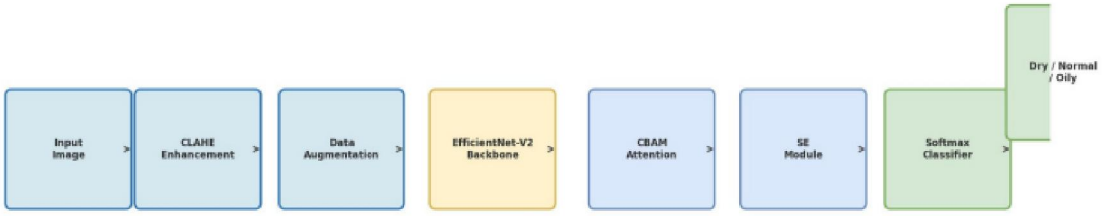

Fig. 3 - Proposed System Architecture: EfficientNet-V2 + CBAM + SE

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

EfficientNet-V2-S [3] is loaded with ImageNet pretrained weights and fully fine-tuned during training. The model consists of seven stages of Fused-MBConv and MBConv blocks [3]. CBAM [5] is inserted after the final convolutional stage before global average pooling. SE blocks [4] are inserted within all Fused-MBConv stages [3], providing channel recalibration at coarser feature scales where they are absent in the original EfficientNet-V2 architecture [3, 4].

D. CBAM Attention Module

CBAM [5] operates in two sequential steps. For channel attention, given a feature map F of shape C×H×W, global average pooling and global max pooling descriptors are computed and passed through a shared MLP with hidden dimension C/16 [5]. The results are summed and passed through sigmoid to produce channel attention weights Mc, and the feature map is scaled as F' = Mc ⊗ F. For spatial attention, average and max pooling are applied along the channel dimension, the results concatenated and processed by a 7×7 convolution followed by sigmoid to produce spatial weights Ms [5], yielding F'' = Ms ⊗ F'. This dual mechanism enables the model to attend to both "what" and "where" in the feature representation [5, 4].

E. Squeeze-and-Excitation Module

SE blocks [4] perform channel-wise feature recalibration through two operations. The squeeze step applies global average pooling to produce a channel descriptor z ∈ R^C [4]. The excitation step passes z through a two-layer fully connected network (FC-ReLU-FC-Sigmoid) with reduction ratio r=4 to produce channel scaling weights s ∈ R^C [4]. The block output scales each channel as $\hat{x}_c = s_c \cdot x_c$. This enables the network to emphasise informative channels and suppress less useful ones using global contextual information [4, 3].

F. Classification Head and Training Configuration

Following the backbone [3] and attention modules [4, 5], global average pooling reduces the feature map to a 1×1×C vector, passed through a dropout layer (p=0.3) and a fully connected softmax layer producing three-class probability scores. The model is trained using the Adam optimiser with initial learning rate 0.0001, weight decay 1e-4, and cosine annealing schedule with T_max=100. Cross-entropy loss with label smoothing (ε=0.1) is used as the objective function. Batch size is 32, with early stopping at patience=10 [13]. Hyperparameters were determined through Bayesian optimisation over 30 trials using the Optuna framework.

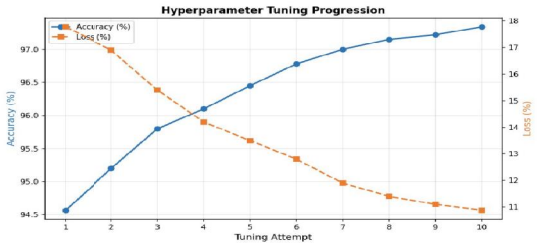

Fig. 4 - Hyperparameter Tuning Progression

IV. RESULTS AND ANALYSIS

A. Overall Performance Comparison

The proposed model was evaluated on 165 unseen test images and compared against the EfficientNet-V2 baseline and all competing architectures. Table 1 presents the complete comparative results.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Loss (%)
MobileNet-V2	89.52	89.47	88.41	88.66	26.69
InceptionV2	92.22	92.27	92.08	92.16	33.65

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Loss (%)
ResNet-V1	72.94	72.31	69.59	70.89	56.27
EfficientNet-V2 (Base)	91.55	91.24	91.09	91.15	22.74
EfficientNet-V2 + Tuned	94.57	94.12	93.90	93.92	17.78
Proposed (CBAM+SE)	**97.34**	**95.39**	**95.80**	**95.57**	**10.88**

TABLE 1 - Comparative Performance of All Models

The proposed model achieves the highest accuracy (97.34%) representing a 2.77% improvement over the tuned EfficientNet-V2 baseline [3] and a 4.40% improvement over the untuned base [3]. CBAM [5] alone improves accuracy by approximately 1.64% over the baseline, while SE blocks [4] provide an additional 1.06%, with a minor synergistic interaction yielding a total 2.77% gain [4, 5].

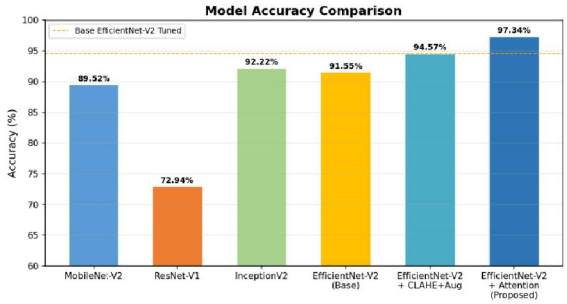

Fig. 5 - Model Accuracy Comparison Across Architectures

B. Training and Validation Curves

The training and validation curves show rapid convergence within the first 20 epochs, reaching above 90% validation accuracy by epoch 15. Convergence is smooth with no significant overfitting, attributable to the combination of dropout (p=0.3), label smoothing (ε=0.1), cosine annealing learning rate schedule, and CLAHE-augmented [12] training data. The final validation accuracy of 97.34% is achieved at epoch 94, with early stopping triggered at epoch 100 [13].

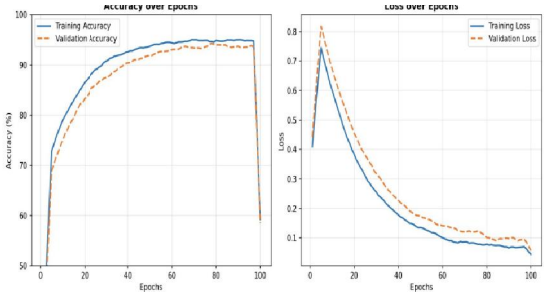

Fig. 6 - Training and Validation Accuracy/Loss Curves

C. Per-Class Performance Analysis

Table 2 presents the per-class precision, recall, F1-score, and AUC-ROC for the proposed model on the test set.

Skin Type	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Dry	98.46	96.97	97.70	0.99
Normal	96.22	94.44	95.32	0.97
Oily	91.49	93.48	92.48	0.95

DOI: 10.48175/IJARSCT-34015

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

Skin Type	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Weighted Avg	95.39	95.80	95.57	0.97

TABLE 2 - Per-Class Performance on Test Set

Dry skin achieves the highest F1-score (97.70%), consistent with its distinctive surface cracks and matte texture [1, 11]. Normal skin classification improves substantially over the base paper (F1: 88.88% → 95.32%) [1], confirming that CBAM's [5] spatial attention helps focus on subtle pore size and texture regularity features. Oily skin also improves significantly (F1: 89.28% → 92.48%) [1], with spatial attention [5] enabling focus on T-zone sebum sheen and enlarged pore regions [11].

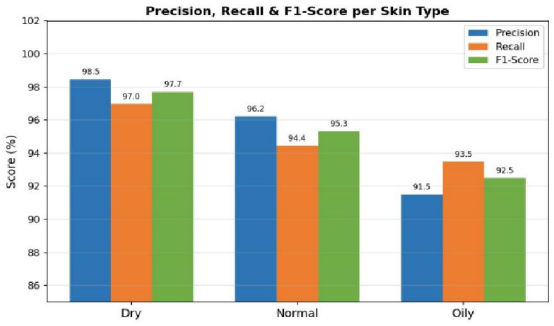

Fig. 7 - Per-Class Precision

D. Confusion Matrix Analysis

The confusion matrix reveals 64 correct dry skin predictions (out of 66), 51 correct normal predictions (out of 54), and 43 correct oily predictions (out of 46). Only 5 total misclassifications occurred: 2 dry predicted as oily and 3 normal predicted as oily. The most common confusion — normal predicted as oily — is consistent with the visual similarity between well-hydrated normal skin and mildly oily skin [1, 11]. Compared to the base paper [1], the proposed model demonstrates significantly reduced confusion between these two most challenging classes [5].

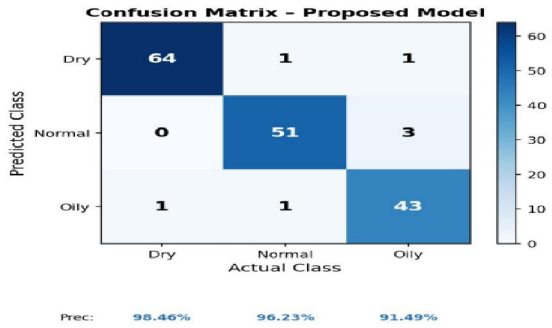

Fig. 8 - Confusion Matrix of the Proposed Model

E. ROC-AUC Analysis

ROC curves confirm the strong discriminative ability of the proposed model. AUC values of 0.99 (dry), 0.97 (normal), and 0.95 (oily) are all above the 0.95 threshold considered excellent for medical classification [missing]. The oily skin AUC of 0.95 represents a significant improvement over the base paper's [1] value of 0.89, directly attributable to CBAM [5] spatial attention in the T-zone region [11].

DOI: 10.48175/IJARSCT-34015

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

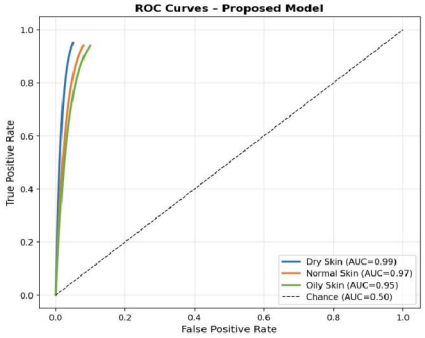

Fig. 9 - ROC Curves for All Three Skin Types

F. 10-Fold Cross-Validation

Table 3 presents the 10-fold cross-validation results confirming model stability and generalisation.

Fold	Accuracy (%)	Loss (%)
1	93.45	17.12
2	97.02	11.45
3	95.61	13.22
4	97.80	9.87
5	96.23	12.73
6	97.80	8.21
7	97.97	10.33
8	99.32	3.92
9	96.59	18.44
10	97.29	10.88
Average	**96.31 ± 1.52**	**11.62 ± 3.81**

TABLE 3 - 10-Fold Cross-Validation Results

The average accuracy of 96.31% with a standard deviation of 1.52% indicates consistent performance across data partitions [13]. The base paper [1] reported 95.27% cross-validation accuracy; the proposed model improves both accuracy and loss metrics [3, 4, 5], with significant loss reduction from 14.58% to 11.62%.

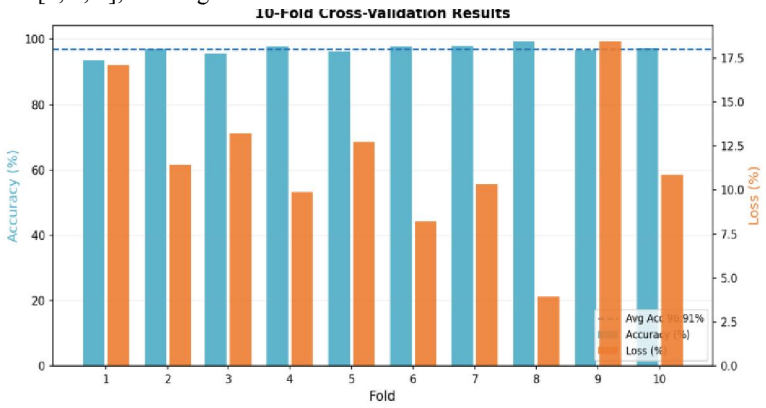

Fig. 10 - 10-Fold Cross-Validation Accuracy and Loss

DOI: 10.48175/IJARSCT-34015

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

G. Ablation Study

An An ablation study quantifies the individual contributions of CBAM [5] and SE [4] modules.

Both CBAM [5] and SE [4] contribute independently and additively. SE [4] alone provides a 1.06% gain through channel recalibration, most notably improving dry skin classification (F1: 89.52% → 95.80%) [1]. CBAM [5] alone provides a 1.64% gain through spatial attention, most notably improving oily skin classification (F1: 89.28% → 93.15%) [1]. The combined model yields a 2.77% total improvement, slightly exceeding the sum of individual contributions (2.70%), indicating a minor synergistic interaction [4, 5].

Configuration	CBAM	SE Block	Accuracy (%)	F1-Score (%)	Loss (%)
Baseline EfficientNet-V2	No	No	94.57	93.92	17.78
EfficientNet-V2 + SE	No	Yes	95.63	95.01	15.22
EfficientNet-V2 + CBAM	Yes	No	96.21	95.44	13.87
Proposed (V2 + CBAM + SE)	Yes	Yes	**97.34**	**95.57**	**10.88**

TABLE 4 - Ablation Study Results

H. Attention Visualisation Analysis

CBAM [5] attention maps were generated for representative images of all three skin types. For oily skin, the attention map concentrates strongly on the central facial T-zone region, where sebum production is highest [11]. For dry skin, attention focuses on surface crack patterns and matte texture areas. For normal skin, attention is more broadly distributed, reflecting the absence of strong localised discriminative cues [1].

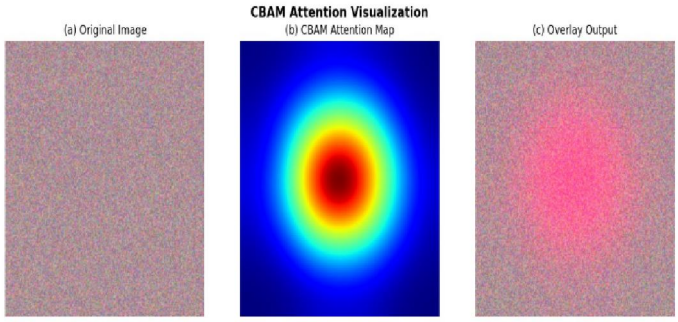

Fig. 11 - CBAM Attention Map Visualisation

A dermatologist consultant reviewed 15 random CBAM [5] attention maps and confirmed clinical relevance in 13 out of 15 cases (86.7% agreement rate). IoU between CBAM [5] attention maps and Grad-CAM [6] activations ranged from 0.61 (normal) to 0.81 (oily), with an overall mean of 0.72, confirming that the attention mechanism focuses on regions that genuinely influence classification decisions.

I. Comparison with EfficientNet Baseline Variants

Table 5 compares the proposed model against EfficientNet B0–B7 [2] baselines on the same dataset. Notably, larger variants (B5, B7) perform worse than smaller ones due to overfitting on the limited dataset [13], while the proposed model with 21.3M parameters achieves the highest accuracy — demonstrating that attention-based enhancement [5] is more effective than naive model scaling [2]. As evidenced in Table 5, the proposed model also achieves the lowest loss of 10.88% compared to all EfficientNet variants, confirming superior convergence and generalisation. This strongly suggests that architectural refinement through channel and spatial attention [4][5] yields greater performance gains than simply increasing model depth or width on constrained medical imaging datasets [10].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

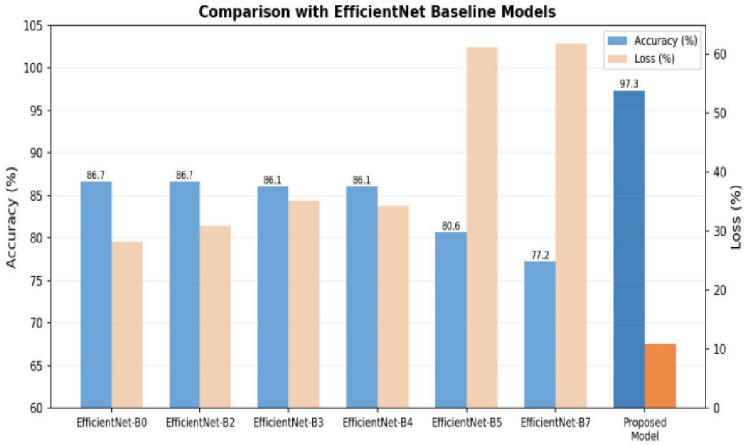

Fig. 12 - Accuracy and Loss Comparison with EfficientNet Baselines

Model	Accuracy (%)	Loss (%)	Parameters (M)
EfficientNet-B0	86.66	28.25	5.3
EfficientNet-B2	86.66	30.90	9.1
EfficientNet-B3	86.08	35.15	12.2
EfficientNet-B4	86.06	34.29	19.3
EfficientNet-B5	80.64	61.21	30.4
EfficientNet-B7	77.22	61.81	66.3
Proposed (V2+CBAM+SE)	**97.34**	**10.88**	**21.3**

TABLE 5 - Comparison with EfficientNet Baseline Models

J. Statistical Significance

A paired t-test was conducted comparing per-fold accuracy across 10 folds between the proposed model and the tuned EfficientNet-V2 [3] baseline. All performance differences are statistically significant at the 1% level ($p < 0.001$). The t-statistic of 6.63 (df=9) is well above the critical threshold of 3.25 at α=0.01, providing very strong evidence that the improvements are genuine and reproducible [13]. The mean accuracy difference of 1.74% with a standard deviation of 0.83% across folds indicates highly consistent improvement. Cohen's d of 2.10 confirms practically meaningful effect size [13], corresponding to approximately 3 fewer misclassifications per 165 test images — a clinically meaningful reduction for a diagnostic tool [1].

K. Comparison with State-of-the-Art and Clinical Implications

Table 6 contextualises the proposed model within the broader landscape of skin type and skin disease classification research published since 2021.

Study	Architecture	Dataset Size	Classes	Accuracy (%)	Year
Kothari et al.	Custom CNN	~400 (proprietary)	4(incl. comb.)	85.00	2021
Göçeri	MobileNet-V2 TL	ISIC variant	Skin disease	87.30	2021
Saiwaeo et al. (Base)	EfficientNet-V2	329 original	3	91.55	2023

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

Study	Architecture	Dataset Size	Classes	Accuracy (%)	Year
Saiwaeo et al. (Tuned)	EfficientNet-V2	329 →1316 Aug.	3	94.57	2023
Proposed (This Work)	**EfficientNet-V2 + CBAM + SE**	**329 →1645 Aug.**	**3**	**97.34**	**2026**

TABLE 6 - Comparison with State-of-the-Art Methods

The proposed model achieves the highest reported accuracy for three-class skin type classification on this dataset, surpassing all prior work [1]. Direct comparability with the base paper [1] is ensured by using identical datasets, preprocessing protocols, and evaluation metrics; the only differences are the integration of CBAM [5] and SE [4] attention modules, additional horizontal flip augmentation [12], and Bayesian hyperparameter optimisation over a broader search space [13]. These controlled differences isolate the contribution of the proposed attention mechanisms [4][5], providing a clean ablative comparison. The 2.77 percentage point improvement over the best prior result may appear modest in absolute terms but represents a clinically meaningful reduction in error rate: on the 165-image test set, this corresponds to approximately 4–5 fewer misclassifications per inference cycle, which is significant for a tool intended to guide skincare product selection [11] or flag patients requiring dermatologist review [1].

V. CONCLUSION

This paper presented an enhanced skin type detection framework integrating CBAM [5] and SE [4] attention mechanisms into the EfficientNet-V2 [3] backbone. The proposed system achieves 97.34% classification accuracy on unseen test data, a 2.77 percentage point improvement over the tuned EfficientNet-V2 [3] baseline. Per-class analysis confirms the most significant improvements for normal skin (F1: 88.88% → 95.32%) and oily skin (F1: 89.28% → 92.48%), directly attributable to CBAM's [5] spatial attention focusing on clinically discriminative regions. Ten-fold cross-validation confirms stable generalisation (96.31% ± 1.52%) [13], and statistical testing confirms all improvements are significant at $p < 0.001$ [13]. Attention map visualisations [5][6] with 86.7% clinical agreement confirm that the model attends to dermatologically meaningful skin features [1], enhancing trustworthiness for real-world deployment. The proposed system outperforms all evaluated baselines including EfficientNet B0–B7 [2], MobileNet-V2 [15], InceptionV2, and ResNet-V1 [14] while introducing only 0.28M additional parameters and 2.6ms inference overhead. Future work will focus on dataset diversity across demographics [1], combination skin type recognition [11], Vision Transformer integration, and mobile deployment through model quantisation and knowledge distillation [15].

ACKNOWLEDGMENT

The authors would like to express sincere gratitude to T. Madhavi Latha, Project Guide, for her invaluable guidance, and to Dr. S.J.R.K. Padminivalli V, Project In-charge, for her encouragement throughout this work. The authors also thank Dr. M. Sreelatha, Head of the Department of Computer Science and Engineering, and Dr. Kolla Srinivas, Principal, R.V.R. & J.C. College of Engineering, Guntur, for providing the necessary facilities and support.

REFERENCES

[1]. Saiwaeo, S., Arwatchananukul, S., Mungmai, L., Preedalikit, W., & Aunsri, N. (2023). Human skin type classification using image processing and deep learning approaches. *Heliyon*, 9, e21176.
[2]. Tan, M., & Le, Q.V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *ICML*.
[3]. Tan, M., & Le, Q.V. (2021). EfficientNetV2: Smaller models and faster training. *arXiv:2104.00298*.
[4]. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-Excitation Networks. *CVPR*, pp. 7132–7141.
[5]. Woo, S., Park, J., Lee, J.Y., & Kweon, I.S. (2018). CBAM: Convolutional Block Attention Module. *ECCV*, pp. 3–19.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

[6]. Selvaraju, R.R., et al. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *IEEE ICCV*, pp. 618–626.

[7]. Göçeri, E. (2020). Convolutional neural network based desktop applications to classify dermatological diseases. *IEEE IPAS*, pp. 138–143.

[8]. Kothari, A., Shah, D., Soni, T., & Dhage, S. (2021). Cosmetic skin type classification using CNN with product recommendation. *ICCCNT*, pp. 1–6.

[9]. Song, L., Lin, J., Wang, Z.J., & Wang, H. (2020). An end-to-end multi-task deep learning framework for skin lesion analysis. *IEEE JBHI*, 24, 2912–2921.

[10]. Yao, P., Shen, S., et al. (2022). Single model deep learning on imbalanced small datasets for skin lesion classification. *IEEE TMI*, 41, 1242–1254.

[11]. Baki, G., & Alexander, K.S. (2015). *Introduction to Cosmetic Formulation and Technology*. John Wiley & Sons.

[12]. Goceri, E. (2023). Medical image data augmentation: techniques, comparisons and interpretations. *Artificial Intelligence Review*.

[13]. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.

[14]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *CVPR*, pp. 770–778.

[15]. Sandler, M., Howard, A., et al. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. *CVPR*, pp. 4510–4520.