

An Attention Fusion of Efficientnetv2 and Convnext Towards Effective Accuracy on Skin Cancer

T.Durga Prasad¹, D. Haritha²

^{1,2}Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur-522502, 2401050037@kluniversity.in, haritha_donavalli@kluniversity.in
Corresponding author: 2401050037@kluniversity.in

Abstract

Among malignant diseases, skin cancer is one kind that affect survival rate in the global health. Existing methods used such as machine learning (ML) approaches in which Support Vector Machines (SVM), Random Forest (RF), and shallow neural networks uses handcrafted features from dermo-scopic images, depends more on manual feature engineering of color, texture, and shape descriptors, that limits to capture complex lesion patterns, leads to reduced generalization across diverse datasets. The deep learning models in which Convolutional Neural Networks (CNNs) uses VGG, ResNet, and Inception although includes automatic feature extraction but still have issues like overfitting on limited datasets, sensitivity to imaging artifacts in terms of hair, and illumination variation, has trouble of both fine-grained textures and global contextual patterns. Other models like many single-backbone architectures suffer from more false negatives due to class imbalance and insufficient representation learning. Hence, this scenario demands a hybrid deep learning framework of EfficientNetV2 and ConvNeXt for fusion of multi-scale features, and transformer-inspired convolutional learning. This model uses data preprocessing, which includes color normalization, artifact removal, and lesion-focused region-of-interest (ROIs) to achieve quality. In addition, the techniques used such as imbalance-aware loss functions, data augmentation, and transfer learning would provide more robustness, improve accuracy, and reduce the error rate. Results observed state that hybrid deep learning model ensures better values than existing methods considered.

Keywords: EfficientNetV2, ConvNeXt, Skin cancer detection, Data Preprocessing, Dermoscopic images, Accuracy, and Robustness.

1. Introduction

Worldwide influencing deadly disease is skin cancer, in which severe form of cancer is melanoma, which quickly spreads to other body parts, resulting in mortality rates. Hence, early detection and accurate classification is required that ensures improving survival rates and treatment plans. There are many ways dermatologists follow such as biopsy mechanism, visual inspection, and dermatologist examination. These models were dependent on clinician expertise as well as consuming more time in diagnosing. There is significant progress in diagnostic tools, techniques development in domains like advancing of deep learning approaches as well as shift from traditional machine modeling, towards automation of prediction of skin cancer. Various ML methods rely on factors like shape, border, textures and color. Due to this, performance is not as expected and won't support capturing complex features and dependencies. To evaluate suspicious lesions, local properties Asymmetry, Border irregularity, Color variation, and Diameter were considered. Skin cell proliferation depends on irregular pigmentation, abrupt color transitions, and micro-textural variations. These attributes indicate changes in histopathology related to structures of melanocyte atypia, uneven melanin distribution, and abnormal keratinocyte. In addition, critical indicators of tumor progression include lesion asymmetry and structural deformation that define invasive patterns in melanoma. The relevant regions were focused by attention-based fusion, using prioritization of irregular borders or heterogeneous color zones. This process helps as a tool for clinicians to fix suspicious regions during examinations. In skin disease progress, advancements in molecular biomarkers extract genotype of the disease, in which phenotypic expression were captured by imaging whereas clinical attributes add a gain in classification accuracy.

Studies such as Nawaz et al. (2025) [6] demonstrate that FCDN-CNN Model, which learns features by layer-wise, improves accuracy but lacks global feature extraction and not able to handle complex lesion structure. Although the study Islam et al. (2026) [7] describes on Multimodal Image + Metadata Model, which supports decision making using attention mechanism, but lacks ensuring robust feature support. Other study such as Nathalie Nataren et al. (2023) [23] defines Molecular Biomarker-Based Diagnosis, which provides biological indicators for enhanced detection of disease

but lacks image-based category as well as limited scalability. Another study Al-Waisy et al. (2025) [10] provides Skin DeepNet Hybrid Model, which enhances accuracy using integrated approaches such as dual features learning, belief network, and segmentation but lacks adaptive attention fusion.

In DL mechanisms, CNN identified as successful approach in automatic learning of hierarchical features. DL methods still face challenges in class imbalance, sensitive to imaging artifacts like illumination and hair variations, and no single architecture in capturing global as well as local details of skin patterned lesions. Hence, a scenario demanded for integration of EfficientNet for local and ConvNeXt for global. When implemented individually, may not exploit complementary information in evaluating better accuracy. To ensure better accuracy, attention fusion mechanism is applied to highlight most significant features. The preprocessing components used such as normalization, ROI lesion segmentation, and artifact removal would improve image quality, then applies augmentation and balanced focal loss to improve robustness and handle class imbalance scenarios.

2. Literature review

The studies demonstrated were classified as per their significance, and its application. They are considered such as (i) Review and Survey studies (ii) CNN and Deep Learning-based detection on skin cancer, (iii) Hybrid architectures on skin cancer detection, (iv) Explainable AI and privacy preserving AI based, and (v) multi-modal and metadata-based detection.

2.1 Review and Survey studies:

This domain uses specific machine learning, deep learning, and imaging methods to automate detection of the disease. Haque S, Ahmad F, et al. (2025) [1] demonstrates reviewing of various approaches K nearest method, Convolutional network (CNNs), Generative network (GANs), etc. were applied on melanoma identification in which significant features influence the detection are lesion size, color, symmetry, and shape. In these methods, CNN experience better accuracy, GAN ensures strong augmentation via imaging simulation. The study focuses on noninvasive and early effective skin cancer detection for therapeutic aids. Zafar, M., Sharif, M. I., et al. (2023) [2] explores biggest organ in human body is the skin, which may lead to cancer due to abnormal growth. The approaches such as deep learning, ML methods, computer vision methods, and aided diagnostics on computers (CAD) are assessed in terms of cleaning the dataset, partitioning in terms of segmentation, engineering the features, and performing classifying the lesions. The limitations and challenges of these techniques are examined in this study. Naqvi, M., Gilani, S. Q., et al. (2023) [3] discuss on visual inspection for skin cancer identification is less accurate, hence the datasets, with exploration of deep learning to ensure accurate detection of the skin cancer, where severity level may spread the cancer to other body parts. This study focused on AI powered machine for skin cancer detection in real-time. Haniye Jafari, Mojtaba Esmailpour Roshan (2023) [21] demonstrate that diagnostic tools used for detecting the disease will improve survival rates when detected early. A method called Asymmetry, Border, Color, and Diameter (ABCD) uses advanced image processing methods for knowing the sample is either benign or malignant. In this, various databases, methods, and algorithms were discussed for automatic classification. Nazari, S., & Garcia, R. (2023) [22] explore severe disease namely melanoma, of skin cancer over patients' population. Machine learning is considered on single-lesion images in over 51 articles on skin cancer detection with existing dermatology datasets. In this study, temporal changes and lesion dynamic patterns are not considered during processing.

2.2 CNN and Deep Learning-based detection on skin cancer:

This category involves CNNs extracting hierarchical features such as lesion shape, texture, and color patterns, and involves DenseNet, MobileNet, and customized CNNs for improved performance and accuracy. Nawaz, K., Zanib, A., Shabir, I. et al. (2025) [6] demonstrates usage of machine learning which fails to extract spatial features, which creates to develop Fully Connected deep supervised (FCDN) - CNN which apply multiple convolution layers to extract hierarchical data, and deep supervision on intermediate layers to improve stability. This model, when compared to pre-trained models MobileNet, EfficientNet, ResNet, and InceptionNet, experiences 96% accuracy. Moturi, D., Surapaneni, R.K. & Avanigadda, V.S.G. (2024) [8] demonstrates that skin cancer is dangerous disease like melanoma and spreading aggressively on the globe increases mortality rate, which has resulted due to abnormalities in metabolic and genetic aspects. HAM10000 dataset is taken, and various deep learning models were performed on pretrained learning networks MobileNet and DenseNet. In those, custom CNN achieves 95% accuracy than MobileNet, and outputs benign or malignant. Esteva, A., Kuprel, B., Novoa, R. et al. (2017) [11] state that there are a variety of ways of examining skin

cancer, followed by a board of 21 dermatologists on biopsy-based examination. To determine disease, two datasets are used, one of which contains all disease images, and the other contains various levels of skin allergy images for verifying of cancer. Techniques like CNN, as well as AI were used to predict the disease more efficiently than dermatologists need and developing applications that run on mobile phones for ensuring user-friendliness. Al Mahmud, A., Azam, S., Khan, I.U. et al. (2024) [14] explore challenge in handling high-resolution images for disease prediction and experience time consuming, and less accuracy in detection rates, hence SkinNet-14 processes low resolution images, with image processing and augmentation techniques to enhance image quality, and address class imbalance using compact convolution transformer (CCT). This study applies to three datasets of HAM10000, ISIC, and PAD, with results in accuracies of 97.8, 96, and 98%, and experience less processing time. Ashfaq, N., Suhail, Z., Khalid, A. et al. (2025) [15] demonstrate that skin cancer is considered the deadliest disease nowadays, development has not progressed due to limited dataset sizes. Hence, CNN, as a deep learning technique applied on ISIC dataset with many classifications up to 8 categories, ensured an accuracy of 90%. This study has observed producing better results than other models in this study, along with diagnostic approaches. Shetty, B., Fernandes, R., Rodrigues, A.P. et al. (2022) [17] state that reviewing machine learning and CNN model, using HAM10000 obtained from diagnostic tools. The model is processed using data augmentation to ensure accuracy and k-fold cross-validation for ensuring robustness. Among the models reviewed, CNN outperforms other models. Irfan Ali Kandhro, Selvakumar Manickam, et al. (2023) [20] emphasizes that reviewing of different pretrained versions like dense connections network, InceptionNet, as well as residual network of connections, and different ML techniques used such as models, like SVMs, Decision trees, and random forests for skin cancer detection assessment. In this, features are extracted using pre-trained models, then fed into defined ML technique, then, were compared against other models for effectiveness.

2.3 Hybrid architecture on skin cancer detection:

Saeed, M.A., Afify, Y.M., et al. (2025) [5] explores various pre-trained models such as Inception, MobileNet, DenseNet, EfficientNet, and ResNet provides metadata, integrated with adaptive weighted ensemble strategy with weights dynamically assigned. SMOTE is used for class imbalance and experimented on ISIC datasets with notable accuracies and performances. Al-Waisy, A.S., Al-Fahdawi, S., et al. (2025) [10] explores existing methods like conventional baselines are consuming more time in skin cancer detection, hence Skin DeepNet is developed in which image contrast and lesion segmentation using GrabCut method and R-CNN approach, then performs dual feature extraction using attention block and pretrained HR model. Deep belief networks train on outputs of these models to capture higher level discriminative attributes. The fusion of these effectively decides skin cancer detection using stacking and boosting over HAM10000 and ISIC datasets. Kumar Lilhore, U., Simaiya, S., Sharma, Y.K. et al. (2024) [18] demonstrate that feature extraction, model design, and fine-tuning hyperparameters for achieving accuracy. Hence, U-Net and MobileNet were introduced with cross entropy loss function, squeeze and excitation, dilate convolutions were incorporated to enhance performance along with Bayesian optimization. The proposed model of this study involves MobileNet, CNN, and Bayesian, which ensures better results than other models. Jaber, N.J.F., Akbas, A. (2025) [19] describes that melanoma is a severe skin cancer raises due to abnormal cells, and spreads on the body leads to death. Early detection with CNN against other models like SVM, KNN, and Decision Trees (DT) using Harris Hawks (HHO) optimization. This model study [19] on evaluation performs superior to other models in skin cancer prediction.

The derived model outperforms the other models used in the detection of skin cancer.

2.4 Explainable AI and privacy preserving AI based:

This domain of methods includes multiple deep learning architectures for combining segmentation, feature extraction, and classification networks to improve robustness and performance. Mukherjee, S., Chandrakar, K., Chowdhury, S. et al. (2026) [4] demonstrates skin cancer is severe disease, detecting it early is challenging due to lack of diagnostic device, and complex visual inspection. The use of deep neural network CNN with swish function, improves accuracy to 98.3%, and achieves better performance measures as well as interpretability using global XAI. Ali, M. D., Iqbal, M. A., et al. (2025) [9] overviews on limitations such as lack of robustness, class imbalance leads to bias, insufficient feature denotation, and unable to capture complex dependencies, requires hybrid approach of denoising diffusion probability component and Augmented diffusion meta learning framework. This model also integrates with Vision Transformer, CNN, and CBAM for long range dependencies. The predictions are combined, and logistic regression learns from these outputs over HAM10000 dataset, ensures 98.6% accuracy. Naz, N.S., Mehmood, M.H., Ahmed, F. et al. (2025) [13] explore that detection of skin cancer is complicated due to factors like shape, color, and texture. Hence, privacy-preserving and explainable Deep learning is proposed in which federated learning is imposed. This model decomposes

the dataset into two classes and distributes over clients. The FL uses VGG16 as a pre-trained model, with CNN in 25 rounds, via finetuning and ImageNet. For improved decision making to the clinicians, uses Grad-CAM for transparency via Ur Rehman, M., Ahmed, F., et al. (2022) [16] demonstrate that manual dermatologist inspection on skin cancer is time-consuming, hence focused on computer diagnostic tools. The evolution of DenseNet and MobileNet under DL concept, and their modified versions by adding extra layers by stacking at the end. By observing, modified versions produce better results than original pre-trained models.

2.5 Multimodal, metadata-based, and biological analysis detection:

This domain provides trust and transparency via techniques Grad-CAM, attention mechanisms, and interpretable models for better decision making. Islam, S., Wishart, G.C., et al.(2026)[7] explores need of AI for skin lesion is infected or noninfected in three ways based on sensitivity such as AI for meta data only, AI for images only, and AI for both images and metadata, in which metadata and images ensure 99% sensitivity than other categories. Post processing in terms of attention module ensures explainability and supports experts to make decisions. Gabani V, Navamani TM, et al. (2026) [12] demonstrate that detection rates are costly with consulting dermatology experts, through equipment, and invasive biopsies. The data preprocessing is applied in terms of scaling, normalization, segmentation, and augmentation to eliminate noise and improve image quality. In this, two-phase tuning is applied in which the top layers of CNN are retrained first, then the rest layers and new layers are added for the custom CNN, then hyper parameter tuning take place, and enhances interpretability using Grad-CAM. Nathalie Nataren, Miko Yamada, et al. (2023) [23] demonstrates research on molecular biomarkers on early skin cancer detection, in which patient access is constrained by COVID and far geographical locations. Developing biomarkers for monitoring and detecting severity of disease such as malignant type. Various biomarkers are reviewed and which is best is identified during the review. Hasan, N., Nadaf, A., et al. (2023) [24] explore that skin cancer cases may be increased in coming years, which should be detected at early stages. Although it is prevented by appropriate drug invention, cell analysis is a challenge during this process. This study explores various therapies and treatment types used till 2023, as well as nano-based approaches and molecular analysis on affecting cancer growth. This study includes multi-drug trials, and their resistance.

3.Methodology

In this, the proposed approach is described in three aspects for better understanding in terms of Figure 1, that demonstrates on important modules involved in skin cancer detection, Figure 2 describes the order of activities involved in achieving better performance, and PS1 demonstrates the implementation of logic in classifying the image into benign or malignant. The objective of this model is to classify the image as benign or Malignant.

Fig.1 demonstrates integration of EfficientNetV2 and ConvNeXt architecture, that process given input sample towards prediction of skin cancer severity. Data acquisition is the first module in this process, where dermoscopic skin lesion images extracted from a dataset HAM10000. Then, data preprocessing is adapted, which have steps such as noise removal, normalization, resizing, and artifact elimination for image quality enhancement. Then, feature extraction using pre-trained models, where EfficientNetV2 and ConvNeXt independently capture complementary features such as EfficientNetV2 focuses on local texture patterns, while ConvNeXt extracts global contextual representations. These features are combined by an attention-based feature fusion, enabling important features from both networks. The fused representation is now processed for classification layers along with an appropriate loss function of focal loss, to assess severity level like either malignant or benign, along with assessing measures like error rate influence robustness and accuracy computing to make the model more effective.

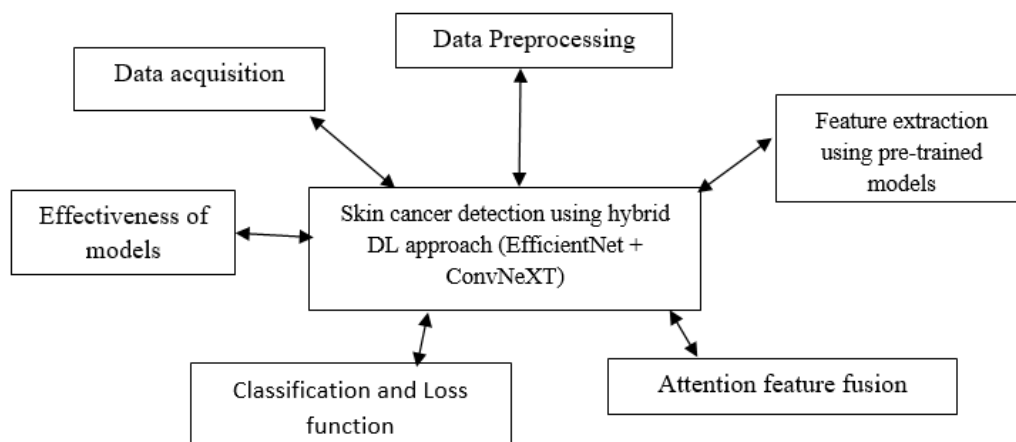


Fig. 1. Significant modules interaction of skin cancer detection using hybrid deep learning networks

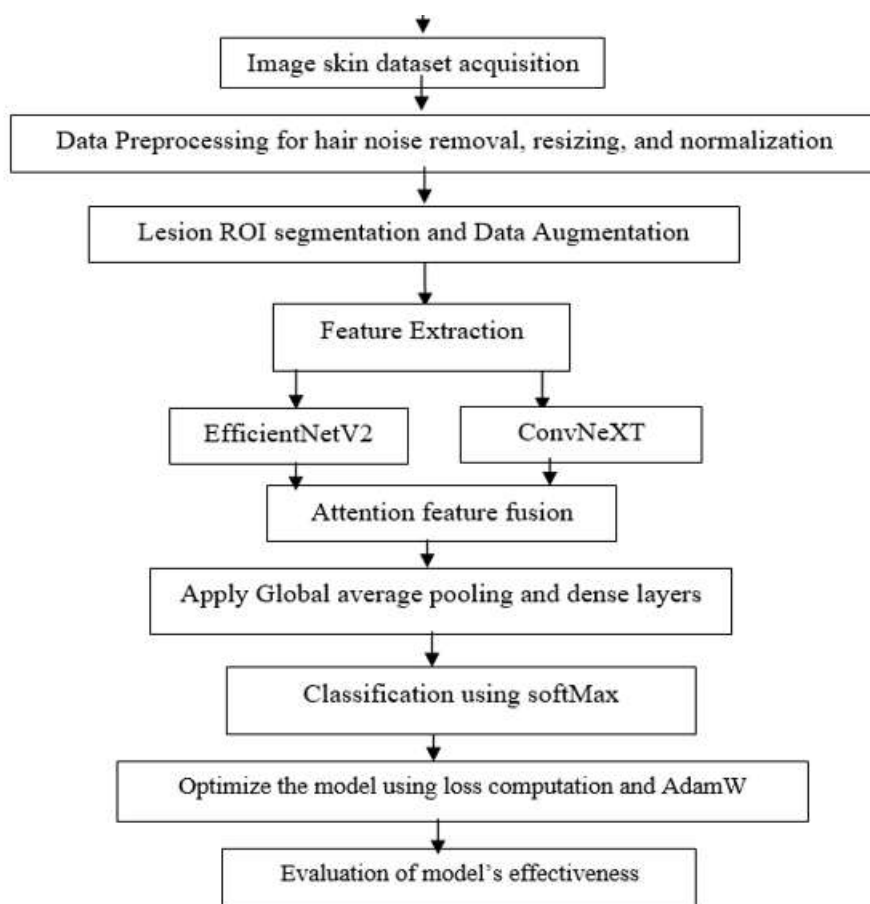


Fig. 2. Flowchart hybrid deep learning network model functionality

Fig. 2 denotes order of activities of the hybrid deep network to determine the skin lesion dataset possess benign or malignant. This flow starts with data collection, then images undergo data preprocessing, through components such as hair noise removal, normalization, and resizing. Next, lesion Region of Interest (ROI) segmentation and data augmentation are used to isolate the lesion area and expand transformation in terms of rotation and flipping. Then, feature extraction is carried out via two parallel deep learning models such as EfficientNetV2, which captures local texture features, and ConvNeXt, extracts high-level contextual representations. Both network's features are combined using an attention-based feature fusion to emphasize the significant informative patterns. Then, the global average

pooling and dense layers for processing fused representations to generate a compact feature notation. Classification is performed through a SoftMax layer, which predicts the skin lesion as benign or malignant. Then, focal loss function with the AdamW optimizer is applied, for improving the generalization. Finally, the effectiveness of the model is evaluated in terms of accuracy and performance.

PS1: Pseudo_Procedure hybrid deep learning network model_EfficientNetV2+ConvNeXT

(Dataset[[]]):

Input: Image dataset is represented by $D = \{(x_i, y_i)\}_{i=1}^N$ where x_i is input image, and predicted class label y_i

Output: Class label y_i , accuracy

Step1: Acquire images from database source HAM10000

Step2: For each image x_i , repeat the following sub-steps

2.1 Apply preprocessing to improve quality and reduce noise using operations such as resizing, color normalization, hair artifact removal, contrast enhancement, and Lesion focused ROI segmentation

$$x'_i = P(x_i) \quad \text{Eq. (1) where } P(\cdot) \text{ represents preprocessed function}$$

2.2 Apply Augmentation in terms of rotation, flipping, zooming, contrast perturbation, and random cropping on preprocessed image x'_i , and is denoted as $x_i^{aug} = A(x'_i)$ Eq.(2) where $(\cdot)$ represents Augmentation operator

2.3 Apply EfficientNetV2 for local feature extraction by taking augmented image as input, to assess lesion features such as texture, edge sharpness, color distribution, border irregularity, and pigment network.

$$F_E = f_E(x_i^{aug}; \theta_E) \quad \text{Eq.(3) where } F_E \text{ is EfficientNet feature map, } f_E(\cdot) \text{ is EfficientNet backbone, and } \theta_E \text{ denotes trainable arguments}$$

2.4 Same image is processed simultaneously by ConvNeXT for global context data such as lesion symmetry, surrounding symmetry, semantic data, and Lesion geometry.

$$F_C = f_C(x_i^{aug}; \theta_C) \quad \text{Eq. (4) where } F_C \text{ denotes ConvNeXT feature map, } f_C(\cdot) \text{ denotes backbone, and } \theta_C \text{ shows trainable parameters.}$$

2.4 Perform attention feature fusion by combining EfficientNetV2 and ConvNeXT for higher focus on channels and lesion regions using

$$F_{cat} = [F_E \oplus F_C] \quad \text{Eq.(5) where } \oplus \text{ denotes concatenation}$$

2.4.1 Denoting attention fusion is $F_{att} = \mathcal{A}(F_{cat})$ Eq.(6) where $\mathcal{A}(\cdot)$ denotes attention fusion operation and F_{att} denotes attention fusion feature.

2.4.2 For channel attention, refinement is done by

$$F_{att} = M_s(M_c(F_{cat}) \odot F_{cat}) \quad \text{Eq.(7) where } \odot \text{ denotes element-wise multiplication, } M_s \text{ denotes spatial fusion map, and } M_c \text{ denotes channel fusion map.}$$

2.5 Apply Global average pooling for reducing dimensionality, and perform on attention fused feature map using $F_{gap} = GAP(F_{att})$ Eq.(8) where F_{gap} denotes a potential feature vector, and $GAP(\cdot)$ denotes global average pooling operation.

2.6 Pooling feature map is applied onto fully connected layers to represent nonlinear learning as $h = \sigma(W_1 F_{gap} + b_1)$ Eq.(9) where h is a hidden feature representation, σ denotes ReLu Activated function, W_1 denotes weighted matrix, and b_1 denotes bias.

2.6.1 Apply dropout to reduce overfitting using $h' = Dropout(h)$ Eq.(10)

2.7 Classify lesions using SoftMax, add another dense layer for classification score using

$$z = W_2 h' + b_2 \quad \text{Eq.(11)}$$

$$2.7.1 \text{ To compute class probability using SoftMax as } \hat{y}_k = \frac{e^{z_k}}{\sum_{j=1}^C e^{z_j}} \quad \text{Eq.(12)}$$

where C is number of lesion classes, and $\hat{y}_k$ is predicted probability for class k . In our case, $C=2$, denotes benign or Malignant

Step3: Class balanced focal loss is defined for the mode using

$$L = - \sum_{i=1}^N \sum_{k=1}^C \alpha_k (1 - \hat{y}_{ik})^{\gamma} y_{ik} \log(\hat{y}_{ik}) \quad \text{Eq.(13)}$$
 where the class weight for class k is denoted as α_k , γ denotes focusing parameter, ground truth indicator is denoted by y_{ik} , and $\hat{y}_{ik}$ denotes predicted probability.

Step4: Updating the model parameters using AdamW optimizer as defined $\Theta^{t+1} = \Theta^t - \eta \nabla_{\Theta} L$ Eq.(14)

Where η denotes learning parameter, and Θ denotes all trainable parameters.

Step5: Evaluate effectiveness of the models using measures such as accuracy, and Error rate

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{Eq.(15)}$$

$$\text{Error Rate} = 1 - \text{Accuracy} \quad \text{Eq.(16)}$$

PS1 demonstrates that skin cancer classification performed on dermatology dataset HAM10000. Initially, the dataset is denoted as $D = \{(x_i, y_i)\}_{i=1}^N$, where x_i denotes the input skin lesion image and y_i denotes the predicted class label. Equations used such as Eq. (1) to Eq. (16) denote transforming mathematical formula into intended functionality. Each image passes through preprocessing operations such as resizing, color normalization, hair artifact removal, contrast enhancement, and lesion-focused ROI segmentation, followed by augmentation using rotation, flipping, zooming, contrast variation, and random cropping to improve image quality and dataset diversity. The augmented images passed over two parallel feature extraction methods such as EfficientNetV2, for local lesion characteristics extraction such as texture, border irregularity, and pigmentation patterns, and other is ConvNeXt, for global contextual details extraction on lesion symmetry, geometry, and surrounding tissue patterns. Then, attention fusion mechanism is performed by combining resulting feature maps of these two backbones with channel and spatial attention. The fused features are reduced in dimensionality using global average pooling and processes by fully connected layers with ReLU activation and dropout to learn nonlinear feature representations while preventing overfitting. A SoftMax classifier predicts the probability of each class as benign or malignant, and the class imbalance is handled by a class-balanced focal loss function. Network parameters are optimized using the AdamW optimizer, and the model effectiveness is evaluated in terms of metrics such as accuracy and error rate, using true and false prediction outcomes.

4. Results

In this section, the dataset used along with its features, and the derived model processing environment in terms of training, testing, and validation. Then, the models used in the assessment of skin melanoma are compared using their evaluated measures such as accuracy, precision, recall, and F1 score, as well as robustness by evaluating error rate. This demonstration is represented in Table 1. Then, the methods that were contributed in full hybrid approach are compared by assessing in absence of contributions and is denoted in Table 2.

4.1 Dataset description and experimental setup:

The dataset named Human against machine with 10,000 samples is used for automatic skin melanoma detection. The dataset is formed by collecting images from different imaging devices and from diverse populations. The dataset contains seven types of images such as basal cell carcinoma, vascular lesions, melanocytic nevi, melanoma, benign keratosis-like lesions, actinic keratoses, and dermatofibroma. Each image is associated with ground truth label via expert assessments or other examinations. The dataset possesses vascular patterns, globules, pigmentation, dots, etc. requires imaging techniques for better visualization of structures. The dataset possesses class imbalance which pose challenge for proposed model to handle. The images possess variations in border, color, hair artifacts, size, and illumination, would require data preprocessing methods. The images are in RGB format, with meta data. For experimental setup, the dataset is split into training data 70%, testing 15%, and rest 15% for validation in order to do hyper parameter tuning. To handle imbalance, and improve model generalization, must require augmentation which involve operations such as flipping, rotation, cropping, contrast adjustment and translation. When fused feature vector is produced after integration of EfficientNet, ConvNeXt, and attention fusion mechanism. The biological related parameters were used in training the model are 224X224 as image size, 32 as batch size, AdamW as optimizer, 1×10^{-4} as learning rate, 50 epochs, class balanced focal loss as loss function, 0.5 as drop rate, and ReLU as activation function. The executable platform considered is colab environment (python), NVIDIA as GPU, 16GB as RAM, and Windows as operating system.

4.2 Comparison of methods:

From Table 1, methods such as Support Vector Machine (SVM) with handcrafted features ensure 84.3% of accuracy with

an error rate of 15.7%, while Random Forest improves performance to 86.7% accuracy and error rate 13.3%, but both approaches are limited on manually extracted features. Deep learning models have ability to automatically learn hierarchical image features, like for instance, VGG16 achieves 91.2% accuracy and ResNet50 reaches 93.4% accuracy, showing better results compared to conventional methods. Advanced architectures such as EfficientNetV2 and ConvNeXt further enhance performance, achieving 96.3% and 97.1% accuracy, respectively, with reduced error rates, due to improved feature extraction and optimized network design. Finally, the proposed hybrid model integrating EfficientNetV2 and ConvNeXt with attention-based feature fusion achieves the best performance, obtaining 99.02% accuracy, 98.9% recall, 98.8% precision, and 98.8% F1-score, while the error rate reduced to 0.98%. This scenario is depicted visually in Fig. 3 for end user easy understanding.

Table 1. Comparison of methods

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Error Rate (%)
SVM + Handcrafted Features	84.3	83.5	81.2	82.3	15.7
Random Forest	86.7	85.4	83.9	84.6	13.3
VGG16	91.2	90.1	89.3	89.7	8.8
ResNet50	93.4	92.6	91.9	92.2	6.6
EfficientNetV2	96.3	95.8	95.4	95.6	3.7
ConvNeXt	97.1	96.6	96.1	96.3	2.9
Proposed Hybrid Model	99.02	98.8	98.9	98.8	0.98

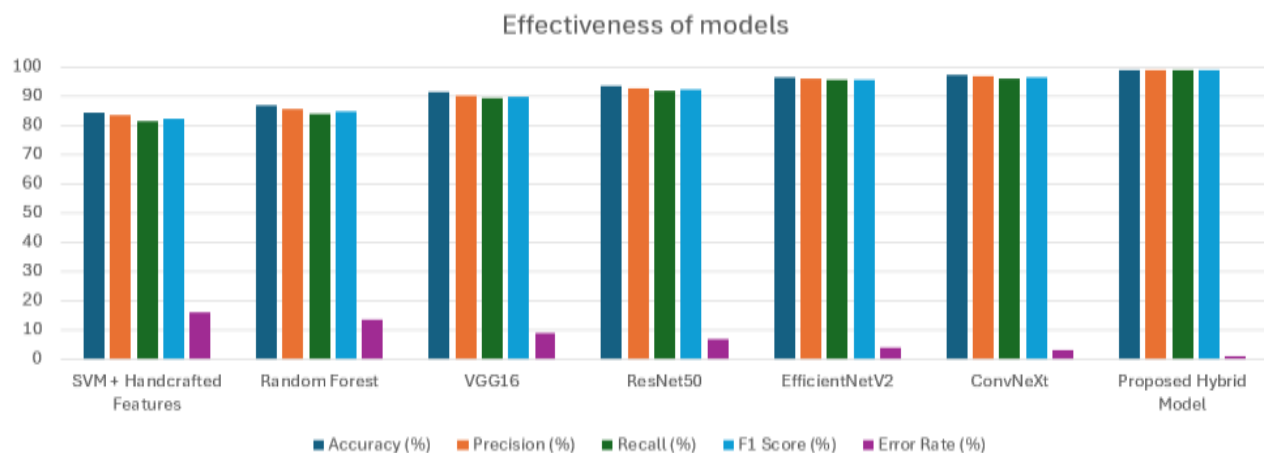


Fig. 3. Effectiveness of the models

4.3 Ablation Study:

In Table 2, where each component contributes to overall system in terms of improving performance. When EfficientNetV2 is used, achieves an accuracy of 96.3% with a 3.7% error rate, that captures fine-grained local lesion textures and color patterns. Simultaneously, the ConvNeXt model ensures 97.1% accuracy with a 2.9% error rate, benefiting the features such as global contextual representations and structural lesion patterns. When both architectures in a hybrid EfficientNetV2–ConvNeXt model, the accuracy improves to 98.2%, shows integration of local and global features enables hyper feature extraction. When attention-based feature fusion is introduced, increasing accuracy to 98.7% and reducing the error rate to 1.3%, that emphasizes clinically lesion regions. Incorporating focal loss enhances managing of class imbalance and reduces misclassification of minority classes, achieving 98.9% accuracy and a 1.1% error rate. Finally, the full model, hybrid feature extraction, attention fusion, optimized loss functions, and improved training, achieves the highest performance with 99.02% accuracy and only 0.98% error rate, shows

effectiveness of the derived model.

Table 2. Comparison of contributed methods as per ablation study

Approach	Accuracy (%)	Error Rate (%)	Key identification
EfficientNetV2 only	96.3	3.7	Processing of local lesion textures
ConvNeXt only	97.1	2.9	Supports global contextual learning
EfficientNetV2 + ConvNeXt (Hybrid)	98.2	1.8	Hyper feature extraction
Hybrid + Attention Fusion	98.7	1.3	Capturing important features
Hybrid + Focal Loss	98.9	1.1	Handles class imbalance
Full Proposed Model	99.02	0.98	Ensures better Performance

4.4 Statistical performance:

The statistical validation is performed in five independent experimental runs, where each run uses the same train-validation-test split strategy with different random seeds. Evaluated measure values are denoted in mean $\pm$ standard deviation ($\mu \pm \sigma$) form, where μ is average performance and σ indicates stability. A paired t-test compares the proposed model against each baseline. The mathematical formulae used for mean, standard deviation, pair p-test are demonstrated as Mean performance is computed using $\mu = \frac{1}{n} \sum_{i=1}^n x_i$ Eq.(17) where x_i is the performance value in the i^{th} run and n is the number of runs.

Standard deviation is denoted using $\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2}$ Eq.(18) where σ represents performance stability. A lower σ indicates more stable model behavior.

Paired t-test significance is computed using $t = \frac{\bar{d}}{s_d/\sqrt{n}}$ Eq.(19) where $\bar{d}$ is two models across runs difference, s_d is the standard deviation, and the number of paired runs is denoted by n .

A p-value less than 0.05 indicates statistically significant improvement, while $p < 0.01$ indicates strong significance.

Table 3. Comparison of statistical performance of models using paired t-tests

Method	Accuracy (%) $\mu \pm \sigma$	p-value vs Proposed
SVM + Handcrafted Features	84.30 $\pm$ 1.12	<0.001
Random Forest	86.70 $\pm$ 1.05	<0.001
VGG16	91.20 $\pm$ 0.84	<0.001
ResNet50	93.40 $\pm$ 0.71	<0.001
EfficientNetV2	96.30 $\pm$ 0.52	0.004
ConvNeXt	97.10 $\pm$ 0.45	0.006
EfficientNetV2 + ConvNeXt	98.20 $\pm$ 0.38	0.018
Hybrid + Attention Fusion	98.70 $\pm$ 0.33	0.031
Hybrid + Focal Loss	98.90 $\pm$ 0.29	0.042

Full Proposed Model	99.02 ± 0.24	—
---------------------	-----------------	---

Table 3 demonstrates the full proposed model performance as an integrated of EfficientNetV2, ConvNeXt, attention fusion, which results focal loss improves both performance and stability. In this process, EfficientNetV2 extracts texture, pigmentation, and border irregularity, while ConvNeXt captures global attributes such as lesion symmetry and contextual patterns. Then, attention fusion emphasizes significant regions, and focal loss reduces the effect of class imbalance.

5. Conclusion

The hybrid deep learning network of EfficientNetV2 and ConvNeXt with attention feature fusion is used for efficiently classifying given sample as benign or malignant based on dermatology dataset. This derived approach involves preprocessing, lesion-focused ROI segmentation, and data augmentation technique, for improving image quality and diversity. EfficientNetV2 as first approach extracts fine-grained local lesion characteristics, while ConvNeXt as second step captures broader contextual information, and applied attention fusion mechanism, which highlights significant spatial and channel features for fixing the cancer regions. The integration of global average pooling, fully connected layers, SoftMax for classification, and class-balanced focal loss addresses class imbalance, and improves classification performance. Evaluated accuracy and error rate of full proposed model demonstrates distinguishing the sample is benign or malignant lesion. In future, incorporation of vision transformers, incorporation of explainable AI for ensuring interpretability along with multi modal data and including of federated learning for privacy-preserving distributed training and would support of scalability on large real world clinical datasets.

REFERENCES

1. Haque S, Ahmad F, et al. (2025), Skin Cancer Detection Using Deep Learning Approaches. *Cancer Biotherapy and Radiopharmaceuticals*. 2025;40(5):301-312. doi:10.1089/cbr.2024.0161.
2. Zafar, M., Sharif, M. I., et al. (2023). Skin Lesion Analysis and Cancer Detection Based on Machine/Deep Learning Techniques: A Comprehensive Survey. *Life*, 13(1), 146. <https://doi.org/10.3390/life13010146>.
3. Naqvi, M., Gilani, S. Q., et al. (2023). Skin Cancer Detection Using Deep Learning—A Review. *Diagnostics*, 13(11), 1911. <https://doi.org/10.3390/diagnostics13111911>.
4. Mukherjee, S., Chandrakar, K., Chowdhury, S. et al.(2026), Explainable deep learning for skin cancer detection using swish-activated convolutional networks. *Discov Onc* 17, 256 (2026). <https://doi.org/10.1007/s12672-026-04386-6>.
5. Saeed, M.A., Afify, Y.M., et al.(2025), Multimodal deep learning ensemble framework for skin cancer detection. *Sci Rep* 15, 45660 (2025). <https://doi.org/10.1038/s41598-025-30534-z>.
6. Nawaz, K., Zanib, A., Shabir, I. et al.(2025), Skin cancer detection using dermoscopic images with convolutional neural network. *Sci Rep* 15, 7252 (2025). <https://doi.org/10.1038/s41598-025-91446-6>.
7. Islam, S., Wishart, G.C., Walls, J. et al.(2026), Advancing skin cancer detection through deep learning and fusion of patient metadata and skin lesion images. *Sci Rep* 16, 1968 (2026). <https://doi.org/10.1038/s41598-025-26392-4>.
8. Moturi, D., Surapaneni, R.K. & Avanigadda, V.S.G. (2024), Developing an efficient method for melanoma detection using CNN techniques. *J Egypt Natl Canc Inst* 36, 6 (2024). <https://doi.org/10.1186/s43046-024-00210-w>.
9. Ali, M. D., Iqbal, M. A., Lee, S., Duan, X., & Kim, S. K. (2025). Explainable AI Based Multi Class Skin Cancer Detection Enhanced by Meta Learning with Generative DDPM Data Augmentation. *Applied Sciences*, 15(21), 11689. <https://doi.org/10.3390/app152111689>.
10. Al-Waisy, A.S., Al-Fahdawi, S., Khalaf, M.I. et al.(2025), A deep learning framework for automated early diagnosis and classification of skin cancer lesions in dermoscopy images. *Sci Rep* 15, 31234 (2025). <https://doi.org/10.1038/s41598-025-15655-9>.
11. Esteva, A., Kuprel, B., Novoa, R. et al.(2017), Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118 (2017). <https://doi.org/10.1038/nature21056>.
12. Gabani V, Navamani TM, Shyamala K and Vaswani Rajpal VK (2026), Multimodal skin lesion classification for early cancer diagnosis using deep learning. *Front. Physiol.* 17:1717517. doi: 10.3389/fphys.2026.1717517.
13. Naz, N.S., Mehmood, M.H., Ahmed, F. et al.(2025), Privacy preserving skin cancer diagnosis through federated deep learning and explainable AI. *Sci Rep* 15, 36094 (2025). <https://doi.org/10.1038/s41598-025-19905-8>.
14. Al Mahmud, A., Azam, S., Khan, I.U. et al.(2024), SkinNet-14: a deep learning framework for accurate skin cancer classification using low-resolution dermoscopy images with optimized training time. *Neural Comput & Applic* 36, 18935–18959 (2024). <https://doi.org/10.1007/s00521-024-10225-y>.

15. Ashfaq, N., Suhail, Z., Khalid, A. et al.(2025), SkinSight: advancing deep learning for skin cancer diagnosis and classification. *Discov Computing* 28, 63 (2025). <https://doi.org/10.1007/s10791-025-09541-1>.
16. Zia Ur Rehman, M., Ahmed, F., et al. (2022). Classification of Skin Cancer Lesions Using Explainable Deep Learning. *Sensors*, 22(18), 6915. <https://doi.org/10.3390/s22186915>.
17. Shetty, B., Fernandes, R., Rodrigues, A.P. et al.(2022), Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. *Sci Rep* 12, 18134 (2022). <https://doi.org/10.1038/s41598-022-22644-9>.
18. Kumar Lilhore, U., Simaiya, S., Sharma, Y.K. et al. (2024), A precise model for skin cancer diagnosis using hybrid U-Net and improved MobileNet-V3 with hyperparameters optimization. *Sci Rep* 14, 4299 (2024). <https://doi.org/10.1038/s41598-024-54212-8>.
19. Jaber, N.J.F., Akbas, A.(2025), Melanoma skin cancer detection based on deep learning methods and binary Harris Hawk optimization. *Multimed Tools Appl* 84, 25709–25722 (2025). <https://doi.org/10.1007/s11042-024-19864-8>.
20. Irfan Ali Kandhro, Selvakumar Manickam, et al. (2023), Performance evaluation of E-VGG19 model: Enhancing real-time skin cancer detection and classification, *Heliyon*, A cell press Journal, DOI: 10.1016/j.heliyon.2024.e31488.
21. Haniye Jafari , Mojtaba Esmailpour Roshan(2023), Review on Automated Skin Cancer Detection Using Image Processing Techniques, *Asian Pacific Journal of Cancer Biology*, Vol 8 No 4 (2023), DOI 10.31557/apjcb.2023.8.4.387-393.
22. Nazari, S., & Garcia, R. (2023). Automatic Skin Cancer Detection Using Clinical Images: A Comprehensive Review. *Life*, 13(11), 2123. <https://doi.org/10.3390/life13112123>.
23. Nathalie Nataren, Miko Yamada, et al. (2023), Molecular Skin Cancer Diagnosis: Promise and Limitations, *The Journal of Molecular Diagnostics*, Volume 25, Issue 1, January 2023, Pages 17-35, <https://doi.org/10.1016/j.jmoldx.2022.09.008>.
24. Hasan, N., Nadaf, A., et al(2023), Skin cancer: understanding the journey of transformation from conventional to advanced treatment approaches, *Mol Cancer* 22, 168 (2023), <https://doi.org/10.1186/s12943-023-01854-3>.
25. Dey NS, Sangaraju HKR. A particle swarm optimization-inspired global and local stability-driven predictive load balancing strategy. *Indones J Electr Eng Comput Sci*. 2024;35(3). doi:10.11591/ijeecs.v35.i3.pp1688-1701.
26. Dey NS, Sangaraju HKR. Hybrid load balancing strategy for cloud data centers with novel performance evaluation strategy. *Int J Intell Syst Appl Eng*. 2023;11(3):883-908.
27. Allamudi AK, Sangaraju HR. Hybrid machine learning for fraud detection: balancing accuracy and security in digital transactions. *Int J Saf Secur Eng*. 2025;15(2):331-337. doi:10.18280/ijssse.150214.
28. Raju SH, Adinarayana S, et al. An optimized approach for defect detection in flyovers using Faster R-CNN, LSTM, and transfer learning. *Math Model Eng Probl*. 2025;12(8). doi:10.18280/mmep.120828.
29. Allamudi AK, Raju SH. An accurate fraud source path identification using integration of graph neural networks, long short-term memory, and XGBoost. *Int J Saf Secur Eng*. 2025;15(11):2343-2352. doi:10.18280/ijssse.151114.