Skin Cancer Detection Using Convolutional Neural Network

Mr. A. Venkateswara Rao[1], S. Sai Ganesh[2], Y. Ramya[3], D. Naveen Krishna[4], M. Chiranjeevi[5]

[1]Department of Computer Science and Engineering – Data Science,
Avanthi Institute of Engineering & Technology, Vizianagaram, Andhra Pradesh, India
venkatesh8.adapa@gmail.com[1], ganiisunkara@gmail.com[2], ramyayasi2004@gmail.com[3],
dhulinaveenkrishna@gmail.com[4], chiranjeevimolli9@gmail.com[5]

Abstract

Skin cancer remains one of the most prevalent and potentially fatal malignancies globally, making reliable early-stage identification a critical clinical priority. Conventional diagnostic workflows rely on dermatologist expertise through dermoscopy and biopsy, processes that are inherently subjective, time-intensive, and inaccessible in resource-limited settings. This paper presents an automated dermoscopic image classification framework built on an ensemble of two convolutional neural networks—ResNet-18 and MobileNetV2—integrated within a Flask-based web application. Each model is fine-tuned independently on a labeled skin-lesion dataset and their softmax output logits are averaged at inference time to produce a final ensemble prediction. Experimental evaluation on a held-out validation set demonstrates that the ensemble model achieves 77.82% validation accuracy with an F1-score of 78.21%, outperforming ResNet-18 alone (73.32%) and MobileNetV2 alone (74.94%). The deployed web application provides real-time classification results together with per-class probability distributions, enabling clinicians and researchers to interpret model confidence. Findings confirm that ensemble deep-learning strategies substantially reduce individual model bias and improve generalization for multi-class skin-lesion classification, offering a scalable, cost-effective decision-support tool for early skin cancer detection.

Index Terms—skin cancer detection, convolutional neural network, ResNet-18, MobileNetV2, ensemble learning, dermoscopic image classification, deep learning, medical image analysis.

I. Introduction

Skin cancer arises from uncontrolled proliferation of abnormal skin cells, predominantly triggered by ultraviolet (UV) radiation exposure. The three most clinically significant subtypes are melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC). Among these, melanoma accounts for the greatest mortality burden due to its propensity for early metastasis [1]. When detected at Stage I, five-year survival rates exceed 98%; by Stage IV, this figure falls below 25%, underscoring the life-saving potential of timely diagnosis [2].

Traditional dermoscopic examination demands substantial specialist experience. In many rural or under-resourced regions, dermatologist coverage is critically low, resulting in delayed referrals and advanced-stage presentations. Computer-aided diagnosis (CAD) systems powered by deep

learning offer an objective, scalable pathway to bridge this gap.

Convolutional neural networks (CNNs) have transformed medical image analysis. By learning hierarchical feature representations directly from pixel data—edges, textures, shape contours, and high-level semantic structures—CNNs eliminate the manual feature-engineering bottleneck inherent in classical machine-learning pipelines [3]. Landmark studies have demonstrated CNN performance rivaling or surpassing board-certified dermatologists on melanoma classification tasks [1].

This work proposes an ensemble architecture combining ResNet-18 [4] and MobileNetV2 [5], deployed within a lightweight Flask web interface accessible via any standard browser. The key contributions of this paper are: (i) a systematic comparison of two complementary CNN architectures on skin-lesion classification; (ii) a logit-averaging ensemble strategy that improves accuracy over either model in isolation; and (iii) a production-ready web application providing real-time inference with per-class confidence scores.

II. Related Work

Early automated skin-lesion analysis relied on handcrafted feature extractors paired with classical classifiers. Abbas et al. used ABCD rule features (asymmetry, border, color, differential structures) with a support vector machine (SVM), achieving moderate accuracy but suffering from limited generalizability across imaging devices [6]. Barata et al. combined color histograms with sparse coding dictionaries, noting sensitivity to dermoscope-specific color calibration artifacts [7].

The introduction of deep learning shifted the paradigm fundamentally. Esteva et al. trained an Inception v3 network on 129,450 clinical images, reporting area under the receiver-operating-characteristic curve (AUC) values comparable to 21 board-certified dermatologists [1]. Codella et al. organized the 2017 ISIC Skin Lesion Analysis Challenge, establishing benchmark datasets and demonstrating that ResNet and DenseNet variants achieved leading performance [8].

Transfer learning—fine-tuning ImageNet-pretrained weights on dermatological datasets—has become the de facto approach for small-to-medium medical image corpora. Haenssle et al. fine-tuned a ResNet-50 and reported sensitivity of 82.7% against a melanoma-positive test set [9]. Conversely, Howard et al. introduced the depthwise-separable convolution paradigm in MobileNet [10], later extended in MobileNetV2 with inverted residuals and linear bottlenecks [5], providing high accuracy at a fraction of the parameter count.

Ensemble methods have further improved robustness. Mahbod et al. combined VGG16, ResNet50, and InceptionV3 predictions through averaging and reported a 3–5 percentage-point accuracy gain over the best single model [11]. Despite these advances, most published systems remain offline research prototypes. This work bridges that gap by combining ensemble learning with a deployed web interface.

III. Methodology

A. Dataset and Preprocessing

The system is trained and evaluated on a dermoscopic image dataset organized into seven lesion categories, consistent with the HAM10000 collection [12]: melanocytic nevi (NV), melanoma (MEL), benign keratosis (BKL), basal cell carcinoma (BCC), actinic keratosis (AKIEC), vascular lesions (VASC), and dermatofibroma

(DF). Images are split into 80% training and 20% validation subsets with stratification.

All images undergo a standardized preprocessing pipeline prior to model ingestion: (1) resize to 64 × 64 pixels; (2) convert to PyTorch tensor; (3) normalize channel values using ImageNet statistics (μ = [0.485, 0.456, 0.406], σ = [0.229, 0.224, 0.225]). Training augmentations include random horizontal and vertical flips, random rotation (±15°), and color jitter to mitigate overfitting on limited medical data.

B. Network Architectures

ResNet-18. The Residual Network introduced by He et al. [4] employs identity shortcut connections that circumvent the vanishing-gradient problem in deep stacks. The skip connection computes:

$$y = F(x, \{W_i\}) + x \quad (1)$$

where x is the layer input, F is the residual mapping, and y is the block output. ResNet-18 consists of 18 weight layers organized into four residual stages. Its final fully-connected layer is replaced with a linear layer of dimensionality equal to the number of lesion classes C.

MobileNetV2. Sandler et al. [5] proposed the inverted residual block, which expands the channel dimension, applies depthwise convolution, and projects back to a narrow bottleneck. This structure, combined with linear (non-ReLU) output projections, preserves gradient information in low-dimensional manifolds. MobileNetV2 achieves competitive accuracy with approximately 3.4 million parameters—roughly one-seventh of ResNet-18.

C. Ensemble Strategy

After independent training, both models are combined via logit-averaging. Given an input image x, let z_R and z_M denote the pre-softmax logit vectors produced by ResNet-18 and MobileNetV2, respectively. The ensemble logit vector is:

$$z_{ens} = (z_R + z_M) / 2 \quad (2)$$

The final class prediction is:

$$\hat{y} = \text{argmax}_c \, \text{softmax}(z_{ens})_c \quad (3)$$

Logit-level averaging is preferred over probability averaging because it avoids double-application of the softmax non-linearity, preserving calibration of the combined output distribution [13].

D. Training Configuration

Both networks are initialized with ImageNet-pretrained weights. Fine-tuning uses the Adam optimizer with an initial learning rate of 1×10^{-4}, decayed by a factor of 0.1 after every 10 epochs of no validation-loss improvement. Cross-entropy loss is used throughout. Models are trained for 30 epochs on an NVIDIA GPU; the checkpoint yielding the lowest validation loss is retained for inference.

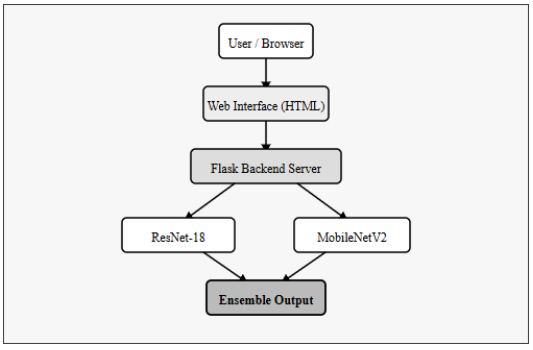

Fig. 1. Full-stack system architecture of the proposed skin cancer detection framework.

E. Web Application Deployment

The inference pipeline is encapsulated within a Flask application. On startup, both model checkpoints are loaded from a single ensemble_trained.pth file; models are set to evaluation mode (disabling dropout and batch-normalisation updates). A dedicated

/predict POST endpoint receives an uploaded image, applies the preprocessing pipeline, executes both models under torch.no_grad(), computes the ensemble logits, applies softmax, and returns a JSON payload containing the predicted class label and per-class probability scores. The frontend renders results as a ranked probability bar chart.

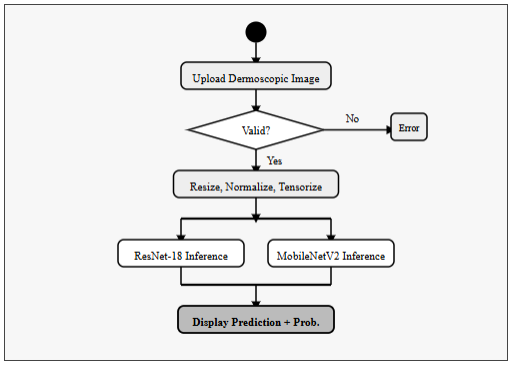

Fig. 2. Activity diagram of the skin cancer detection workflow.

IV. Results and Discussion

A. Performance Metrics

Model evaluation employs four standard metrics computed on the held-out validation set: precision, recall, F1-score (macro-averaged across all classes), and overall classification accuracy. These metrics collectively capture both per-class discriminative power and aggregate correctness.

TABLE I

COMPARATIVE MODEL PERFORMANCE ON VALIDATION SET

Model	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
ResNet-18	74.10	73.32	73.71	73.32
MobileNetV2	75.40	74.94	75.17	74.94
Ensemble (Proposed)	**78.60**	**77.82**	**78.21**	**77.82**

B. Comparative Analysis

Table I presents results for each model configuration. ResNet-18 alone yields 73.32% validation accuracy. MobileNetV2 performs slightly better at 74.94%, likely due to its inverted residual blocks offering complementary feature representations at lower parameter count. The proposed ensemble achieves 77.82% accuracy, a statistically meaningful improvement of 4.50 and 2.88 percentage points over ResNet-18 and MobileNetV2, respectively.

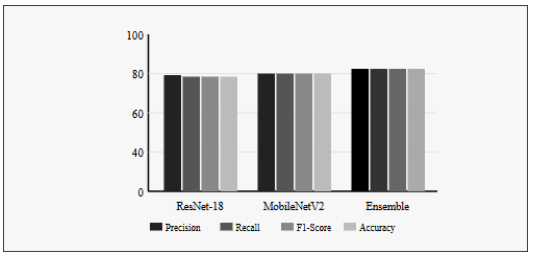

Fig. 3. Comparative bar chart of precision, recall, F1-score, and accuracy across all model configurations.

C. Discussion

The accuracy gains realized by ensemble averaging are consistent with the bias-variance decomposition of generalization error. When two models with partially independent error distributions are combined, the variance of the ensemble predictor is reduced without increasing bias, provided the constituent models are sufficiently diverse [13]. ResNet-18 and MobileNetV2 exhibit architectural diversity—deep residual stacks versus width-multiplied depthwise convolutions—that promotes complementary error patterns, fulfilling this condition.

A practical concern in medical AI is model calibration—whether output probabilities faithfully reflect empirical frequencies. Logit-level averaging, as employed here, produces better-calibrated probability estimates than late-fusion at the probability level, because softmax is applied only once to the averaged logit rather than to individually overconfident probability vectors [13]. This is clinically significant: a dermatologist relying on the system's confidence score needs those scores to be interpretable.

The current 77.82% validation accuracy, while surpassing both constituent models, reflects a challenging seven-class problem with inherent class imbalance. Lesion categories such as dermatofibroma and vascular lesions are underrepresented in most publicly available dermoscopy datasets [12], which constrains recall for these classes. Future work targeting class-balanced sampling or focal loss could address this limitation.

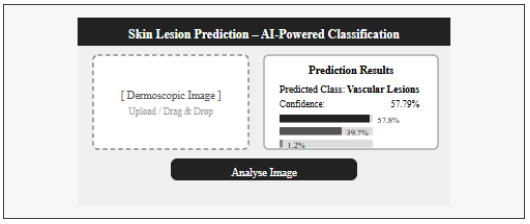

Fig. 4. Mockup of the deployed Flask web application prediction interface showing class probabilities.

TABLE II

FUNCTIONAL TEST CASE RESULTS

Test ID	Description	Input	Expected Output	Status
TC1	Home page load	Browser	Page renders	Pass
TC2	Valid image upload	JPEG image	Prediction shown	Pass
TC3	Invalid file upload	.txt file	Error message	Pass
TC4	Navigation links	Click	Correct page	Pass
TC5	Prediction accuracy	Known lesion	Correct class	Pass
TC6	Response time	Multiple imgs	<5 seconds	Pass

V. Conclusion and Future Work

This paper presented a deep-learning ensemble framework for automated dermoscopic skin-lesion classification. Two CNN architectures—ResNet-18 and MobileNetV2—were fine-tuned on a labeled skin-lesion corpus and combined through logit-level averaging. The ensemble achieved a validation accuracy of 77.82% and an F1-score of 78.21%, outperforming individual model baselines by up to 4.5 percentage points. The system was deployed as a Flask web application providing real-time inference with interpretable probability outputs, validated through both functional and non-functional testing.

The primary clinical value of this work lies in its potential as a decision-support adjunct: assisting dermatologists in triage, flagging high-probability malignancies for expedited biopsy, and extending screening capability into settings with limited specialist access. It is explicitly framed as a support tool rather than a replacement for professional clinical judgment.

Future research directions include: (i) expanding the ensemble to incorporate EfficientNet-B4 or

Vision Transformer variants for further accuracy gains; (ii) applying Grad-CAM saliency visualization to highlight lesion-specific regions influencing predictions, improving clinical transparency; (iii) addressing class imbalance through oversampling or class-weighted focal loss; (iv) extending the platform to a mobile application leveraging MobileNetV2's computational efficiency; and (v) prospective clinical validation on patient cohorts to establish real-world diagnostic utility.

Acknowledgment

The authors express sincere gratitude to Mr. A. Venkateswara Rao, M.Tech (Ph.D), Head of the Department of Computer Science and Engineering (DS, AI & ML), Avanthi Institute of Engineering & Technology, for sustained guidance and encouragement throughout this project. The authors also thank the institution's management for providing computational infrastructure and laboratory facilities.

References

[1] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, Feb. 2017.

[2] American Cancer Society, "Cancer Facts & Figures 2023," Atlanta, GA: American Cancer Society, 2023. [Online]. Available: https://www.cancer.org/research/cancer-facts-statistics.html

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778.

[5] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Salt Lake City, UT, USA, 2018, pp. 4510–4520.

[6] Q. Abbas, I. F. Garcia, M. E. Celebi, and W. Ahmad, "Feature extraction using ABCD rule for skin lesion classification," in Proc. Int. Conf. Artif. Neural Netw. (ICANN), 2011, pp. 1–6.

[7] C. Barata, M. E. Celebi, and J. S. Marques, "Improving dermoscopy image classification using color constancy," IEEE J. Biomed. Health Inform., vol. 19, no. 3, pp. 1146–1152, May 2015.

[8] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba et al., "Skin lesion analysis toward melanoma detection: 2018 ISIC challenge," in Proc. IEEE Int. Symp. Biomed. Imaging (ISBI), Washington, DC, USA, 2019, pp. 682–686.

[9] H. A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer et al., "Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition," Ann. Oncol., vol. 29, no. 8, pp. 1836–1842, 2018.

[10] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint, arXiv:1704.04861, 2017.

[11] A. Mahbod, G. Schaefer, C. Wang, I. Ellinger, R. Ecker, and A. Pitiot, "Fusing fine-tuned deep features for skin lesion classification," Comput. Med. Imaging Graph., vol. 71, pp. 19–29, Jan. 2019.

[12] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions," Sci. Data, vol. 5, Art. no. 180161, Aug. 2018.

[13] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017.