RESEARCH PAPER

Confidence-Aware Melanoma Detection Using MobileNetV2 on a Collaborative Dermoscopic Dataset

[1]Sumaiya Shaikh, [2]Dr. Bhavesh Patel and [3]Dr. Vidyullata Devmane

[1,2,3]Department of Computer Engineering, Shah and Anchor Kutchhi Engineering College Mumbai, India

Email Id:[1]sumaiya.patel24@sakec.ac.in, [2] bhavesh.patel@sakec.ac.in and [3]vidyullata.devmane@sakec.ac.in

Received: 24th April, 2026; Revised: 20th May 2026; Accepted: 30th June, 2026; Available Online: 10th July, 2026

ABSTRACT

Melanoma is one of the most serious and life-threatening forms of skin cancer, so early detection is important in order to increase the chances of survival. Despite the advances of deep learning, existing automatic systems suffer from imbalance in a dataset, limited diversity, and minimal applicability in the real world. The complete implementation of a transfer learning-based melanoma detection framework using an internal architecture of three classes and a binary decision output is given with the use of a curated collaborative dataset. The system is proposed to use the MobileNetV2 architecture for classification and the conventional preprocessing and data augmentation technique for achieving good performance. Experimental testing with 1,125 test images results in a high rate of accuracy of around 89.07% and an ROC-AUC of 0.97, which corresponds to good discriminating power. Transfer learning is shown to be effective by comparing the results with a baseline CNN network. The emphasis is on the quality of the datasets, their clinical relevance and reproducibility, and the framework could be used to support decision-making in the early detection of melanoma.

***Keywords*—**Melanoma detection, transfer learning, MobileNetV2, dermoscopy, skin lesion classification, confidence-based inference, mole uncertainty, collaborated dataset, deep learning

How to cite this article: Shaikh S, Patel B, Devmane V., Confidence-Aware Melanoma Detection Using MobileNetV2 on a Collaborative Dermoscopic Dataset. Int J Drug Deliv Technol. 2026;16(76s): 621; DOI: 10.25258/ijddt.16.76s.56

Source of support: Nil.

Conflict of interest: None

I. INTRODUCTION

Melanoma is a dangerous form of skin cancer, which develops when melanocytes, which are pigment-producing cells, grow uncontrollably. The trigger for this abnormal growth is UV radiation from sunlight causing damage to the skin. Early detection is critical, as Stage IV detection has a much lower 5-year survival rate, approximately 16%-52%, as compared to Stage I survival rate, which is 97%-99%. The prognosis in Stage I is excellent, whereas Stage IV is more severe, although new treatments are improving survival.

Recent breakthroughs in deep learning and transfer learning have greatly enhanced dermoscopic image classification. MobileNetV2, ResNet50, EfficientNet and Vision Transformers are among the architectures that have proven to be successful on benchmark datasets like ISIC and HAM10000. Most of the studies that exist, however, are built on these standard datasets, which might not be sufficient to capture variations in skin tone and lesions in actual clinical settings. In addition, many systems predict without explicitly stating prediction uncertainty and are therefore not clinically reliable.

This paper introduces a confidence-aware melanoma detection framework by leveraging MobileNetV2 transfer learning techniques and a collaboratively curated dermoscopic dataset to overcome these challenges. The framework is based on publicly available ISIC images, which are carefully curated from reputable educational resources, to enhance the diversity of the dataset. The proposed system is different from the traditional binary classifiers because it has an internal uncertainty-support mechanism for recognizing the lesions that are difficult to distinguish from the image and recommends clinical review rather than making a potentially unreliable prediction.

This work has the following major contributions:

1) Creation of a dermoscopic dataset using the integration of ISIC images with curated images publicly available.

2) Development of a confidence-aware melanoma screening framework that combines MobileNetV2 transfer learning with an uncertainty-support decision mechanism for clinically ambiguous lesions.

3) A three-stage confidence-based inference strategy that identifies uncertain cases requiring clinical review instead of forcing a binary prediction.

4) Development of a web-based prototype for practical deployment of melanoma screening and evaluation.

**Author for Correspondence: sumaiya.patel24@sakec.ac.in*

The proposed framework is different from existing ones, which mainly seek to enhance the classification accuracy, in that it aims at clinically reliable decision support using confidence-aware inference. This system also explicitly classifies uncertain cases where the lesion's nature cannot be clearly determined, allowing dermatologists to examine the lesions further, reducing the risk of misdiagnosis and enhancing the safety of AI-supported melanoma detection.

II. LITERATURE SURVEY

This section reviews key studies published during 2024–2025, organized by the primary challenge each addresses.

The transfer learning approach using lightweight architectures has been extensively used for dermoscopic image classification. Recently, Zakariah et al. [3] introduced an architecture based on MobileNet, MixNets, and transfer learning with 99.58% accuracy for binary classification tasks. However, it lacks the mechanisms for diverse dataset validation and explainability, which restrict the clinical reliability. In their work, Sri Ram and Kumar [6] compared the performance of MobileNetV2, VGG16, and ResNet50 and found VGG16 to be most stable with an accuracy of 92.7%, highlighting the importance of demographic differences in training data to be applicable in the real world. Babatunde et al. [16] tackled the issue of segmentation by using K-means clustering before classification using MobileNet, obtaining an accuracy of 98.20%, but this approach still faces the potential of overfitting on small datasets.

Another key challenge to clinical integration is model interpretability. Gajare et al. [4] developed a multiclass object detection framework with Grad-CAM and SHAP visualization, which achieved an average precision score of 92.4% with improved transparency in decision-making. Mittal et al. [13] proposed a new method called DermCDSM, which utilizes improved Chameleon Swarm Optimization along with a Convolutional Deep Spiking Neural Network with an accuracy of 95.07%. Almufareh et al. [20], however, note that current AI frameworks lack dermatology-specific concept explanation.

Several studies have examined the efficiency of the computation and resource-constrained deployment. Tchema et al. [7] demonstrated that a simple 3-layer CNN can achieve performance comparable to the more complex YOLOv7 but with significantly lower computational requirements. To realize cloud-independent inference with 94.9% accuracy, Arefi et al. [11] created quantized CNN models that can be run on FPGA-based hardware through the use of a library called hls4ml. AlexNet is also structured and pruned by Amer et al. [12], resulting in 97.48% accuracy and significantly lower model size for edge deployment.

Two major challenges are the imbalance of the datasets and the costs of annotations. Ravikumar et al. [8] applied Auxiliary Classifier GANs for the synthesis of rare classes of lesions, although the issues related to the authenticity of the generated data and ethical questions have been raised. Farhatullah et al. [9] have implemented quantum Chebyshev polynomial autoencoders and LSTM neural networks with 99.58% accuracy, but the high computational complexity of the system is not suitable for real-time applications. Nevertheless, there is a lack of generalizability across several datasets, as Jiang et al. [10] provided a framework based on self-supervised learning to reduce the amount of labeled datasets and achieved an AUC of 98.2% in a slide-based evaluation.

Multi-modal integration and cross-dataset robustness were addressed by Mohamed et al. [5], who suggested MiSC, which combines image features using

MobileNetV2 along with metadata analysis using random forest via late fusion with an accuracy of 95.60%. To get better representation across various lesion types, Khan et al. [15] developed a hybrid model of VGG16-VGG19. While deployment of CNN was done in conjunction with Harris Hawk Optimization, the accuracy of the CNN in IoT is between 96.35% and 98.44% on ISIC datasets, which were not verified in real time, nor was the privacy evaluated.

These findings are then further consolidated in systematic reviews. Jaiyeoba et al. [21] found inconsistency in processing and lack of standardization for the poor generalizability. Even with high accuracy, explainability, privacy, and integration in clinical workflows remain open challenges, as proved by Verma et al. [22]. One important point that was overlooked, as highlighted by Almufareh et al. [20], was demographic diversity, particularly dark skin color.

From this comprehensive analysis, three critical gaps emerge that directly motivate the present work: over-reliance on ISIC and HAM10000 databases that have a lack of representation of dark-skinned people class imbalance in the case of multiclass databases where there is just an 11-20% occurrence of melanoma among all images and the absence of confidence-aware inference mechanisms that communicate prediction uncertainty to clinicians which is a safety critical requirement for real-world deployment. The proposed framework targets all three gaps by providing a curated collaborative dataset, binary classification of melanoma and a three-state decision module with confidence.

The research gaps identified from the literature survey are as follows:

From the studies, it is evident that the existing systems rely heavily on standard datasets such as ISIC and HAM10000. These datasets consist of skin lesions of individuals with low skin pigmentation. This results in the underrepresentation of dark skin lesions, which leads to models being trained on this biased dataset. This will have an immediate impact on the disease prediction using the model trained on this dataset. In

this work, we have attempted to solve this issue of dataset diversity by using a collaborative dataset.

Another area that needs to be focused is the issue of class imbalance, due to which the model seems to be biased towards majority classes. The multiclass dataset contains only 11-20% of melanoma images out of the total images. The proposed system addresses this gap by curating a binary melanoma-focused dataset with near balanced class representation between benign and malignant.

The reviewed papers suggested the use of visualization techniques like Grad-CAM and SHAP for interpretability, but the current deep learning models are “black boxes" that do not meet clinical requirements. Lack of dermatology-aligned, concept-based explanation makes it difficult to integrate into the real-world implementation. This emphasizes the need for explainable AI for future integration.

III. DATASET FORMATION

The benign samples exhibit regular borders, uniform coloration, and symmetric morphology, whereas the malignant samples demonstrate characteristic dermoscopic features, including asymmetric borders, variegated pigmentation, and irregular lesion structure.

The dataset was formed through the integration of two distinct image sources:

1) **ISIC (International Skin Imaging Collaboration) Archive is a widely recognized benchmark repository containing thousands of dermoscopic images labeled across multiple skin lesion categories.**

2) **Dermoscopic images collected through controlled web scraping- The images collected through web scraping were dermatologically verified, filtered for quality, and reviewed by a domain expert. The images are publicly available educational and medical content that include non-identifiable dermoscopic images. Patient-specific confidential information was not gathered.**

The final dataset consists of 11,134 images organized across three classes: benign, malignant, and mole. The training subset contains 10,009 images, and the testing subset contains 1,125 images. The mole class, comprising 325 images in total, represents melanocytic nevi that exhibit visual overlap with early-stage melanoma. For binary classification output, mole-class predictions are handled through a confidence-based uncertainty mechanism rather than hard class assignment.

Table I. Dataset Distribution for Melanoma Screening

Dataset Split	Benign	Malignant	Mole	Total
Training Set	5,084	4,685	240	10,009
Testing Set	520	520	85	1,125
Total	**5,604**	**5,205**	**325**	**11,134**

IV. METHODOLOGY

A. System Architecture

The proposed system is implemented as a four-stage pipeline, as illustrated in Fig. 1.

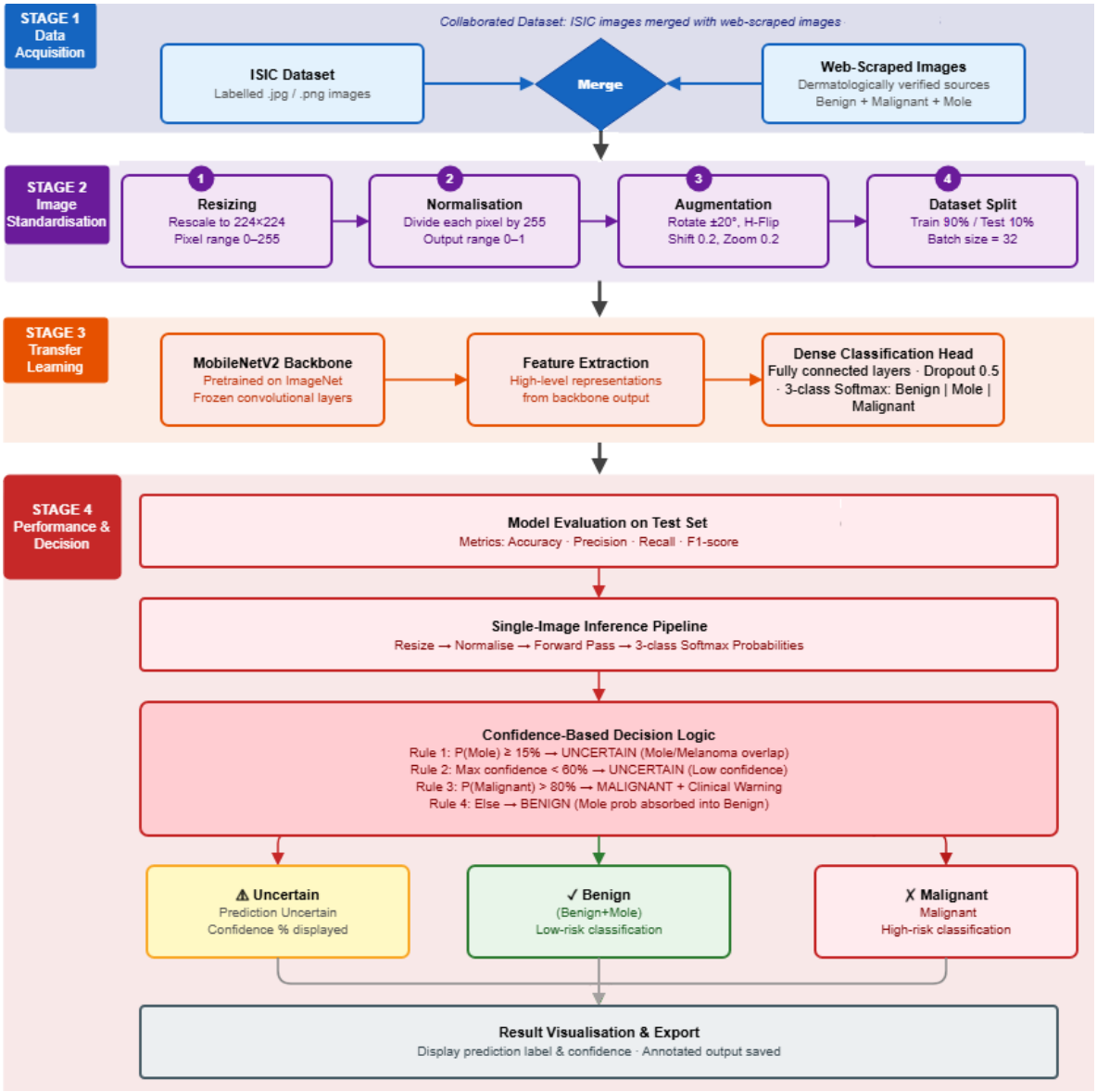

Fig. 1. Proposed confidence-aware melanoma detection framework.

1) Stage 1- Data Acquisition: The dataset used in this study was compiled from two complementary data sources, recognizing that a balance among variety, clinical data, and quantity was needed. The main source was the ISIC (International Skin Imaging Collaboration) Archive. To add more images and enhance class representation, we collected additional images through organized web scraping from dermatologically verified online sources. Both sources were stored as raw .jpg or .png images in one dataset, which formed the foundation for all future processes.

2) Stage 2- Image Standardization Pipeline: In order to ensure consistency and generalization capabilities of the model, all images in the collaborated dataset were subjected to a standardized preprocessing procedure before the training. First, each image was resized to 224 × 224 pixels to match the input requirements of MobileNetV2. The values of the pixels were then re-scaled to [0, 1] by dividing each pixel intensity by 255. To reduce overfitting and improve robustness, data augmentation techniques, including random rotation (±20°), horizontal flipping, width and height shifts (0.2), and zooming (0.2) were applied to the training set. Lastly, the dataset was split into 90% training (10,009 images) and 10% test (1,125 images) without altering the three class categories (benign, malignant, and mole).

3) Stage 3- *Transfer Learning-Based Feature Extraction*: Feature extraction was performed using the MobileNetV2 architecture, a lightweight convolutional neural network pre-trained on the ImageNet dataset. The technique of transfer learning was utilized to take advantage of the rich hierarchical representations that were already built in the convolutions of the network, thus decreasing the amount of the labeled training set needed for the training process and increasing the speed of convergence.

The backbone of the MobileNetV2 architecture was used as a fixed feature extractor with the weights of the convolutional layers frozen during the first training phase in order to keep the learned representations such as edges, textures, and color gradients. The feature map that comes out of the backbone architecture is fed into the custom dense classification head, which consists of the fully connected layers with dropout regularization followed by a softmax layer producing a probability distribution over the target classes. The entire model was then optionally fine-tuned in subsequent epochs with a reduced learning rate to allow the upper convolutional layers to adapt to dermoscopic image characteristics.

4) Stage 4- Confidence-Based Decision Module: Single-image inference was implemented as a standalone pipeline, allowing the system to classify an unseen dermoscopic image at runtime. An input image is preprocessed using the same set of transforms used for training, namely resizing and normalization, and then is fed into the trained model to generate raw logits, which are further converted to probabilities by applying the softmax function.

Three-stage confidence-based decision-making was employed during single image inference to address prediction uncertainty and the clinical ambiguity associated with moles. Internally, the model outputs probability estimates belonging to three classes: benign, malignant, and mole.

Decision rules were as follows: when the probability of mole was greater than 15%, the prediction was recorded as 'Uncertain: Maybe Mole / Maybe Melanoma' and manual checking was required, as there is no visual difference between moles and early melanoma. If mole probability stays below 15%, mole probability is summed up into the benign class, and a regular binary prediction is returned. Irrespective of their class, the predictions with overall maximum confidence below 60% are also classified as unsure. Malignant predictions are given with a clinical warning when they have a high confidence level of above 80%. The empirical choice of the confidence thresholds during model development relied on the observation of prediction confidence on the validation set. Conservative threshold values were adopted to prioritize sensitivity and reduce the likelihood of confidently misclassifying potentially malignant lesions.

This solves the issue of clinical safety highlighted in the literature review where hard binary predictions without uncertainty communication create deployment risk, particularly at the benign-malignant boundary where mole misclassification is most likely to occur.

5) *Stage 5*- Performance Evaluation: For the evaluation of the model performance, four standard classification metrics were used on the test dataset, including accuracy, precision, recall, and the F1-score, which were calculated on the model's three-class outputs (benign, malignant, and mole) so that the confidence-aware uncertainty mechanism could be validated alongside standard melanoma classification performance, rather than being obscured by remapping mole predictions into the benign class. While accuracy provides a general measure for the correctness of prediction, precision and recall are measures that show the balance of false positives and false negatives, respectively. The F1-score, which is the harmonic mean of precision and recall, has been chosen as the main performance metric due to the possible imbalances of classes in the skin lesion data.

Lastly, all outcomes are presented through a results visualizer and exporter that keeps a record of the output image along with its annotation and prediction.

B. **System Interface**

To demonstrate the practical implementation of the proposed framework, a web-based user interface was developed that enables users to upload dermoscopic images and obtain automated melanoma predictions. The interface is shown in Fig. 2.

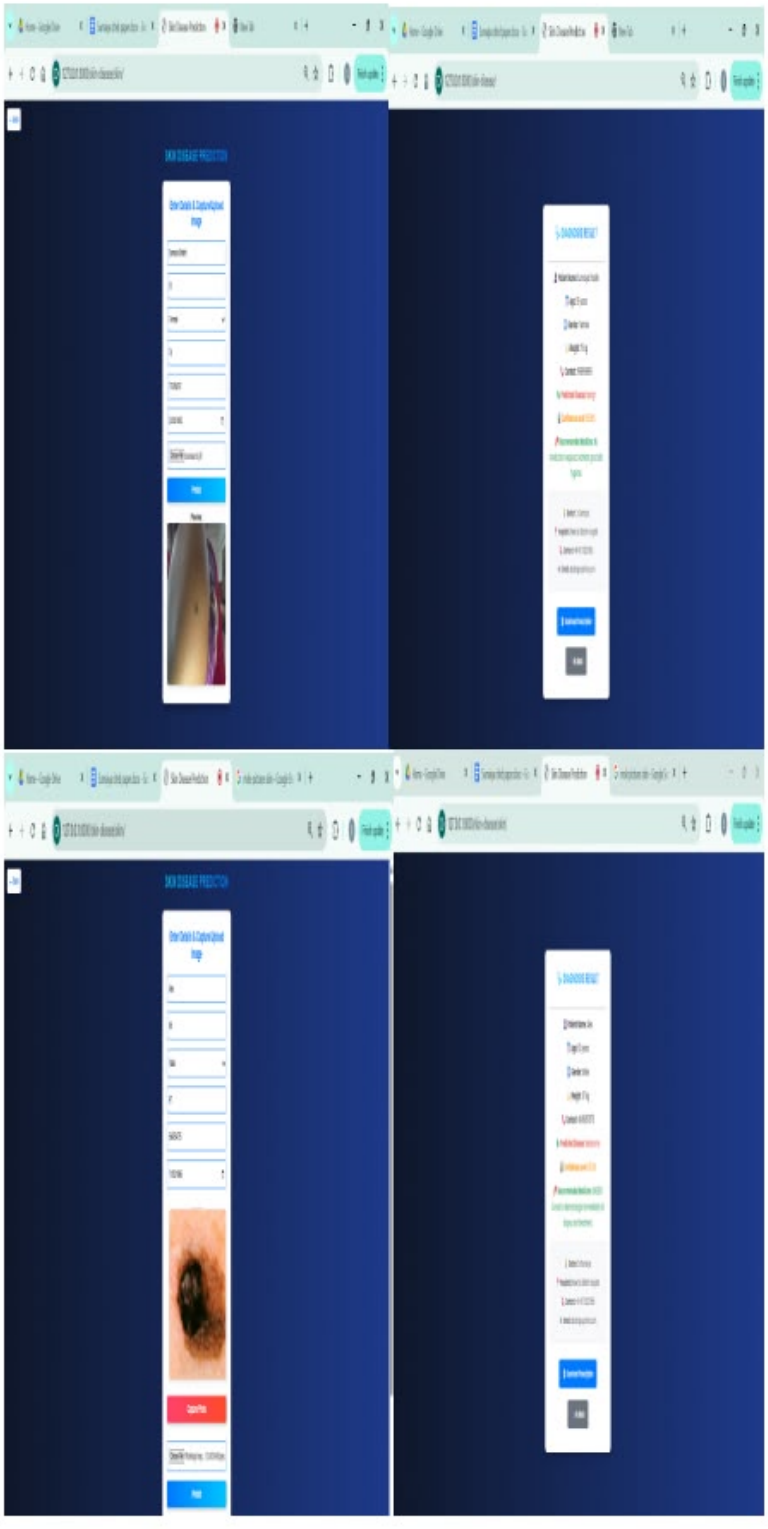

Fig. 2. Web application interface: (a) patient detail entry and image upload form with image preview, (b) diagnosis result for a benign case showing predicted label, confidence score 99.88%, and clinical recommendation, (c) diagnosis result for a malignant case showing Melanoma prediction at 100% confidence with urgent clinical warning.

C. Experimental Environment

All experiments were conducted on Jupyter Notebook using an NVIDIA Tesla T4 GPU. The framework was implemented in Python 3.10 with TensorFlow 2. x, Keras, and Matplotlib. Training completed in approximately 7 minutes per epoch.

V. EXPERIMENTAL RESULTS

D. Training Convergence

Training accuracy increased steadily from 0.82 to 0.88 over 10 epochs. Validation accuracy remained consistently higher than training accuracy throughout, stabilizing around 0.89–0.90. Training loss decreased steadily, while validation loss showed minor oscillation after epoch 5 but did not trend upward, indicating no significant overfitting.

E. Confusion Matrix Analysis

A three-class confusion matrix was computed on all 1,125 test images. By keeping the mole class separate from the other classes and not mapping it to benign, the confidence-aware uncertainty mechanism could be assessed directly without introducing any changes into the primary benign-versus-malignant clinical assessment. The normalized confusion matrix revealed a classification accuracy of 92% for benign images, where 7% were misclassified as malignant and 1% as mole. The malignant class received a recall of 0.89, indicating strong classification of melanomas, as evidenced by the approximately 11% false-negative rate.

Although a recall of 0.89 indicates good sensitivity, the remaining false-negative cases highlight the need for clinical review of uncertain predictions, which motivated the confidence-aware decision mechanism.

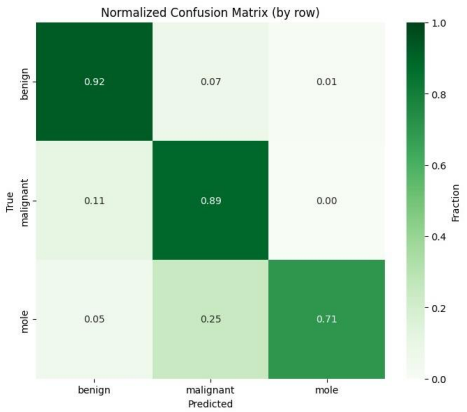

Fig. 3. Normalized Confusion Matrix for the Test Dataset.

The confusion matrix is illustrated in Fig. 3. The matrix shows that the benign class got a recall of 92%, which means that most of the benign lesions were correctly classified, and there were not many false positive predictions. The recall rate for the malignant class is 89%, and this is an effective measure of correctly identifying melanoma. However, this also identifies a small number of false-negative cases. The diagonal dominance of the matrix shows a high overall level of correctness of the classification of the model and a uniform predictive power in both classes. The confusion matrix also confirms the efficiency of the proposed transfer learning framework based on MobileNetV2 to classify dermoscopic lesions into benign or malignant and mole categories while supporting confidence-aware clinical decision-making. The comparative precision, recall, and F1-score for benign, malignant, and mole classes are presented in Fig. 4.

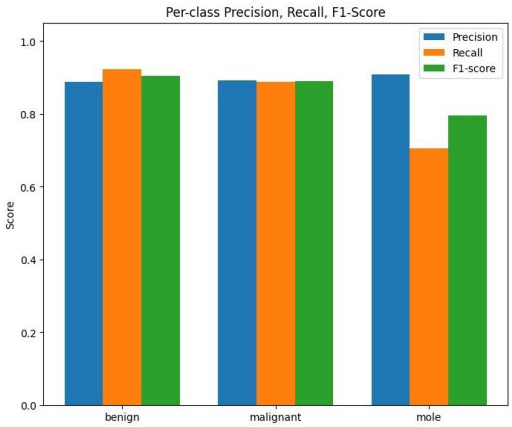

Fig. 4. Comparison of Per-class Precision, Recall, F1-score of benign, malignant and mole classes

The benign class had the maximum recall score of 0.92, indicating that the model has a better ability to accurately classify non-cancerous lesions.The precision value for the malignant class was 0.89, confirming reliable melanoma detection performance. The mole class included to evaluate the uncertainty detection mechanism rather than as a primary classification target, achieved a precision of 0.91 and an F1-score of 0.79.

F. **ROC CURVE**

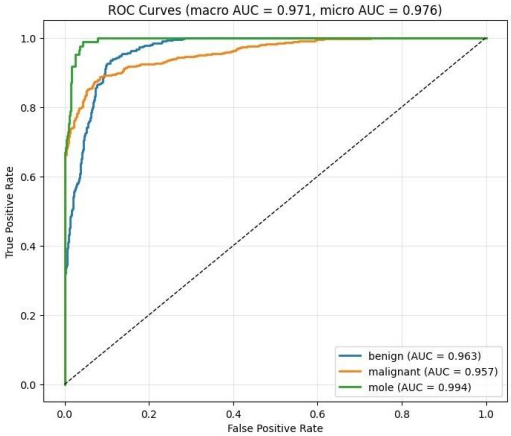

Fig. 5. One-vs-Rest ROC curves for benign, malignant, and mole classes (Macro AUC = 0.971, Micro AUC = 0.976).

The ROC curve in Fig. 5. shows the classification performance of the proposed melanoma detection framework at various thresholds. The curve demonstrates the trade-off between the True Positive Rate (Sensitivity) and False Positive Rate (1 − Specificity) for each class. The model developed using the proposed MobileNetV2 architecture attained AUC values of 0.963 and 0.957 for the benign and malignant classes respectively, indicating strong discriminative capability and reliable class separation. The mole class, evaluated solely to validate the confidence-aware uncertainty mechanism, achieved an AUC of 0.994. The ROC curves demonstrate high sensitivity and low false-positive rates for the primary benign-versus-malignant classification task.

VI. Comparison with Existing Methods

A CNN model baseline was used to evaluate the effectiveness of transfer learning compared to conventional training methods. The CNN model baseline included three convolutional layers using ReLU activation with max-pooling operations in between, followed by the classification layer. Both models were trained for 10 epochs with a batch size of 32 in the same 90:10 training–testing split.

Table II. Baseline Model Comparison

Reference	Model used	Dataset used	Accuracy	Clinical Uncertainty	Macro AUC
Zakariah et al. [3]	MobileNet + MixNets	ISIC + HAM10000	99.58%	No	-
Sriram & Kumar [6]	MobileNetV2, VGG16, ResNet50	ISIC	92.70%	No	-
Mohamed et al. [5]	MiSC (MobileNetV2 + RF)	PAD- UFES-20	95.60%	No	-
Baseline CNN (3-layer)	Custom CNN	ISIC	89%	No	-
Proposed	**MobileNetV2**	**ISIC + Web Scraped**	**89.07%**	**Yes**	**0.971**

Zakariah et al. [3] reported higher accuracy using a different dataset configuration. However, their study did not address confidence-aware inference or demographic diversity, both of which are incorporated into the proposed framework. While the proposed framework's accuracy (89.07%) is comparable to the baseline CNN (89%), it achieves substantially stronger discriminative performance, with a macro-average AUC of 0.971 and micro-average AUC of 0.976 indicating more reliable separation between benign and malignant lesions. This result is consistent with findings from Zakariah et al. [3], Sriram & Kumar [6], and Mohamed et al. [5], demonstrated the effectiveness of MobileNetV2-based architectures for dermoscopic image analysis.

VII. LIMITATIONS

Despite the promising results achieved by the proposed framework, several limitations warrant acknowledgement. Firstly, the recall of the malignant class for the test set is 0.89, which means that approximately 11% of melanoma cases were falsely classified as benign.

This can be considered a considerable problem in the clinical setting due to high risks associated with false negatives and the necessity of calibration of the threshold before using the framework. Secondly, although the web-scraped images were dermatologically checked and selected based on the quality standards, the images cannot be compared with the standardized image sets such as the ISIC dataset due to the domain differences. Thirdly, the dataset has limited geographical and phenotypic diversity and requires external validation on multi-center clinical datasets. Fourthly, the framework lacks explainability using techniques such as Grad-CAM for clinical interpretation. Finally, the mole subset consists of only 325 images (240 for training and 85 for testing). Expanding this subset in future work would improve the reliability of the uncertainty detection mechanism.

VIII. CONCLUSION

This paper presented a complete, implemented AI framework for early melanoma detection using a carefully curated collaborative dataset combining ISIC Archive images with ethically web-scraped dermoscopic sources. The framework addressed the following three major gaps that exist in the literature:

1) The lack of representation of melanomas in the training datasets
2) Demographic homogeneity in the benchmarking datasets
3) Lack of confidence-aware inference methods in the context of clinical usage.

The MobileNetV2-based transfer learning model trained on 10,009 images with three categories and tested on 1,125 images reached 89.07% accuracy, a weighted F1-score of 0.89, a macro-average ROC-AUC of 0.971, and a micro-average ROC-AUC of 0.976. Comparative analysis with the baseline CNN showed the advantages of transfer learning through substantially stronger discriminative performance (AUC) despite comparable overall accuracy. The confidence-based decision layer, which explicitly accounts for the uncertainty in the mole, is the main innovation provided in this paper, allowing reliable classifications to be distinguished from cases requiring clinical review. The lightweight architecture of MobileNetV2 allows reaching the appropriate level of computational efficiency for use in resource-limited clinical environments.

This work establishes a reproducible, transparent, and clinically oriented foundation for AI-driven melanoma screening. Future research will be directed toward optimizing the convolutional base for domain adaptation, integrating explainable AI approaches like Grad-CAM for clinical visualization, expanding the demographic spectrum of the dataset and external validation on multi-center clinical datasets.

REFERENCES

[1] American Cancer Society, “Key Statistics for Melanoma Skin Cancer,” 2024.

[2] World Health Organization, "Skin Cancers," WHO, 2024.

[3] M. Zakariah, M. Al-Razgan, and T. Alfakih, "Skin cancer detection with MobileNet-based transfer learning and MixNets for enhanced diagnosis," Neural Comput. Appl., vol. 36, pp. 21383–21413, 2024, doi: 10.1007/s00521-024-10227-w.

[4] N. Gajare and P. Mane, "A Robust CNN Approach for Multi-Class Skin Disease Detection and Classification," in Proc. 2025 Int. Conf. Intelligent and Cloud Computing (ICoICC), Bhubaneswar, India, 2025, pp. 1–5, doi: 10.1109/ICoICC64033.2025.11052018.

[5] E. H. Mohamed, N. Abdu, M. Khalil et al., "MiSC: A hybrid multi-modal deep learning approach for accurate skin cancer detection," Multimed. Tools Appl., vol. 84, pp. 45321–45345, 2025, doi: 10.1007/s11042-025-20951-7.

[6] G. Sriram and N. Siva Kumar, "Skin Cancer Detection with Images Processing in Deep Learning," in Proc. 2025 3rd Int. Conf. Self-Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India, 2025, pp. 959–964, doi: 10.1109/ICSSAS66150.2025.11080769.

[7] R. Bogne Tchema, A. C. Polycarpou, and M. Nestoros, "Skin cancer classification using machine learning," Multimed. Tools Appl., vol. 84, pp. 3239–3256, 2025, doi: 10.1007/s11042-025-20595-7.

[8] A. Ravikumar, H. Sriraman, C. Chadha, and V. Kumar Chattu, "Alleviation of Health Data Poverty for Skin Lesions Using ACGAN: Systematic Review," IEEE Access, vol. 12, pp. 122702–122723, 2024, doi: 10.1109/ACCESS.2024.3417176.

[9] Farhatullah, X. Chen, D. Zeng, J. Xu, R. Nawaz, and R. Ullah, "Classification of Skin Lesion With Features Extraction Using Quantum Chebyshev Polynomials and Autoencoder From Wavelet-Transformed Images," IEEE Access, vol. 12, pp. 193923–193936, 2024, doi: 10.1109/ACCESS.2024.3502513.

[10] Z. Jiang et al., "A Self-Supervised Learning Based Framework for Eyelid Malignant Melanoma Diagnosis in Whole Slide Images," IEEE/ACM Trans. Comput. Biol. Bioinform., vol. 21, no. 4, pp. 701–714, Jul.–Aug. 2024, doi: 10.1109/TCBB.2022.3207352.

[11] E. Arefi and M. Mousazadeh, "Hardware implementation of a quantized CNN model for early detection of skin cancer cells using hls4ml tool,"

Neural Comput. Appl., vol. 37, pp. 21147–21171, 2025, doi: 10.1007/s00521-025-11411-2.

[12] Y. A. Amer, H. I. Saleh, and O. A. Nasr, "Optimized AlexNet Pruning for Edge-Based Medical Diagnostics," IEEE Access, vol. 13, pp. 139129–139140, 2025, doi: 10.1109/ACCESS.2025.3593453.

[13] R. Mittal, F. Jeribi, R. J. Martin, V. Malik, S. J. Menachery, and J. Singh, "DermCDSM: Clinical Decision Support Model for Dermatosis Using Systematic Approaches of Machine Learning and Deep Learning," IEEE Access, vol. 12, pp. 47319–47337, 2024, doi: 10.1109/ACCESS.2024.3373539.

[14] Z. Kuş and M. Aydin, "MedSegBench: A comprehensive benchmark for medical image segmentation in diverse data modalities," Sci. Data, vol. 11, p. 1283, 2024, doi: 10.1038/s41597-024-04159-2.

[15] M. A. Khan, D. Rastogi, P. Johri et al., "Hybrid deep CNN model for multi-class classification of skin lesion," Neural Comput. Appl., vol. 37, pp. 19479–19499, 2025, doi: 10.1007/s00521-025-11409-w.

[16] A. N. Babatunde, B. F. Balogun, A. J. Babatunde et al., "Improving skin lesion classification using k-means, transfer learning and Mobilenet architecture," Netw. Model. Anal. Health Inform. Bioinforma., vol. 14, p. 70, 2025, doi: 10.1007/s13721-025-00569-3.

[17] N. J. F. Jaber and A. Akbas, "Melanoma skin cancer detection based on deep learning methods and binary Harris Hawk optimization," Multimed. Tools Appl., vol. 84, pp. 25709–25722, 2025, doi: 10.1007/s11042-024-19864-8.

[18] V. S. S. B. T. Sathvika et al., "Pipelined Structure in the Classification of Skin Lesions Based on Alexnet CNN and SVM Model With Bi-Sectional Texture Features," IEEE Access, vol. 12, pp. 57366–57380, 2024, doi: 10.1109/ACCESS.2024.3387533.

[19] D. Christopoulos, S. Spanos, V. Ntouskos, and K. Karantzalos, "TRACE: Transformer-Based Risk Assessment for Clinical Evaluation," IEEE Access, vol. 13, pp. 101721–101734, 2025, doi: 10.1109/ACCESS.2025.3577973.

[20] M. F. Almufareh, "Unveiling the Spectrum of UV-Induced DNA Damage in Melanoma: Insights From AI-Based Analysis of Environmental Factors, Repair Mechanisms, and Skin Pigment Interactions," IEEE Access, vol. 12, pp. 64837–64860, 2024, doi: 10.1109/ACCESS.2024.3395988.

[21] O. Jaiyeoba, E. Ogbuju, G. Ataguba et al., "State-of-the-art skin disease classification: a review of deep learning models," Netw. Model. Anal. Health Inform. Bioinforma., vol. 14, p. 9, 2025, doi: 10.1007/s13721-024-00495-w.

[22] S. Verma, A. Kumar, and M. Kumar, "A Review on Deep Learning Techniques for Skin Disease Detection," in Proc. 2024 1st Int. Conf. Advanced Computing and Emerging Technologies (ACET), Ghaziabad, India, 2024, pp. 1–6, doi: 10.1109/ACET61898.2024.10729965.