Lightweight ShuffleNetV2-Based Deep Learning Framework for Accurate and Interpretable Skin Cancer Classification

Krishna Kumar[1], Avantika Mishra[2], Anushka Singh[3], Ananya Srivastava[4]
{krishnamathsforyou@gmail.com, mavantika674@gmail.com, anushkaasingh0101@gmail.com, *ananyasrivastava*0304@*gmail.com*}
School of Computing Science and Engineering& Galgotias University, Greater Noida, India[1]
Pranveer Singh Institute of Technology & Kanpur, Uttar Pradesh, India[2]
Pranveer Singh Institute of Technology & Kanpur, Uttar Pradesh, India[3]
Pranveer Singh Institute of Technology & Kanpur, Uttar Pradesh, India[4]

Abstract. Early detection of skin cancer is crucial, yet histopathological diagnosis remains challenging due to the visual similarity of lesions. This work proposes an interpretable, lightweight diagnostic model using ShuffleNetV2 as the backbone for efficient computation. Preprocessed dermoscopic images were used to train the model with regularization techniques such as label smoothing, Mixup, and calibration to improve generalization. Interpretability was ensured through Grad-CAM visualizations. On an independent test set of 200 images, the model achieved 91.5% accuracy, a macro F1-score of 91.5%, ROC-AUC of 0.9675, MCC of 0.83, and an ECE of 0.0918. Grad-CAM confirmed that clinically relevant lesion regions were consistently highlighted. The proposed method outperformed several classical and state-of-the-art CNN models, demonstrating its potential for real-time dermatology applications.

Keywords: Skin cancer classification, Deep learning, ShuffleNetV2, Dermoscopic images, Grad-CAM, Binary classification, Explainable AI

1 Introduction

Skin cancer especially melanoma has become one of the most lethal but also preventable diseases in the world. Melanoma starts in melanocytes pigment producing cells that determine skin color is a rapidly advancing disease with high metastatic potential if not detected early [1]. According to the latest data from the American Cancer Society, nearly 87,110 new melanoma cases were estimated along with more than 9,730 death in US alone, emphasizing that there is a growing threat and global health concern about this disease [2]. According to clinical studies, early diagnosis increases survival rates significantly so that fast but accurate diagnosis in dermatology is important [3].

IICTDS 2025, December 19-20, Chandigarh, India
Copyright © 2026 EAI
DOI 10.4108/eai.19-12-2025.2366209

Specifically, standard diagnostic methods including dermoscopic observation and pathological biopsy are still the golden standards for melanoma diagnosis [4]. Dermoscopy specifically improves visualisation of subsurface structures in the skin that are otherwise not possible with the naked eye and helps dermatologists to recognise patterns of lesions more accurately. But, these techniques have their drawbacks too. There is a great deal of variability in dermoscopic analysis, which is subjective and varies based on clinician expertise, leading to high interobserver variation. Additionally, biopsies are invasive, slow and costly process and not always available in remote or resource-constrained areas. Therefore, they are unable to receive timely access to professional diagnosis service for high-risk population.

To overcome these constraints, the computer-aided diagnosis (CAD) systems have been highlighted. These systems offer a structured and objective assessment to supplement dermatologists' clinical judgment. When incorporated with clinical workflow, CAD systems can improve the detection performances of malignancy from 75% to almost 92%, indicating their promise in increasing diagnostic consistency and minimizing human subjective bias [5]. Yet even with the remarkable progress, fully automatic diagnosis of benign nevi vs Malignant is still challenging due to large intra-class variance in shape and appearance, subtle mean differences between classes (i.e., many small shapes look more like nevi than malignant), overlapping visual features, irregular lesion boundaries, phenomena that are shared equally by (benign/malignant) lesions such as hairs, ruler marks or differences in illumination. Furthermore, the scarcity of annotated dermoscopic datasets and their imbalance in classes make the model challenging to learn.

Previous classification of melanomas based mostly on handmade feature descriptors related to the color asymmetry, the granularity of texture, irregularities in borders or geometry of shapes. In this context, these handcrafted features were then used to train the classical classifiers such as SVMs or ANNs [3, 6–10]. Despite showing good initial results those methods had limitations regarding their generalizability between different lesion types, aquisition devices and clinical settings. With advances in deep learning (DL), particularly convolutional neural networks (CNNs), the automated extraction of hierarchical features directly from raw pixel data for medical imaging has been a paradigm shift. CNN based models now surpass traditional systems in segmentation, classification, feature localization and dermoscopic pattern analysis [10, 11].

However, many cutting-edge deep networks are complex to compute and lacking in efficiency for practical purposes such as real-time applications on mobile/embedded devices of POC diagnosis. To eliminate these shortcomings, this work proposes an efficient and interpretable CAD pipeline for binary skin cancer classification using a lightweight modified ShuffleNetV2 model. As efficient channel-shuffling operation ShuffleNetV2 with good accuracy-complexity trade-off is one of the best candidate architectures for mobile devices and edge computing in medical service applications [12, 13]. It employs interpretability of Grad-CAM (Gradient-weighted Class Activation Mapping) to visualize the discriminative lesion regions and improve the transparency for clinical utilization. Leveraging strong preprocessing, robust augmentation and optimized training strategies –Adam optimizer, Binary Cross-Entropy loss, cosine annealing scheduler– the goal of this work is to provide a high-performing diagnostic tool that is generalizable and resource efficient.

2 Related Work

Automated classification of skin lesions has seen substantial progress over the last decade, moving from handcrafted feature–based approaches to deep convolutional models able to learn complex representations from images. Traditional Machine Learning(ML) techniques like Support Vector Machines (SVM) [6, 13] or Principal Component Analysis (PCA) [14] were mainly used to classify the skin lesion images at the early stages of research and needed explicit feature engineering and domain knowledge to model color, texture and geometry cues. For instance, an SVM model with DermQuest dataset of 200 images was able to reach accuracy of 80%, precision, recall and F1-score of 86.21%,71.43% and 78.23% [7]. Likewise, a PCA model tested on 53 samples of small private database reached an accuracy of 83%, however the small database and poor variation in lesions resulted in high inter-class similarity problem [14]. These methods showed initial proof of concept of such technique, however it was not robust or generalizable for larger and more diverse dermoscopic sets.

With the mainstream success of deep learning, convolutional neural networks (CNNs) became the most popular method for dermoscopic image analysis with automatic hierarchical feature extraction capabilities. Several studies have combined CNN-based feature extractors with SVM classifiers, based on the preprocessed datasets such as PH2, EDRA and ISIC Archive leading to accuracy up to 90%, which emphasizes the role of proper preprocessing and dermoscopic feature enhancement in order to predict melanoma [13]. More general analyses with combined datasets (ISIC 2016–2019, PH2 and DermQuest) presented accuracies between the range of 76% to 88%, showing CNN's aptitude for learn the discriminative lesion representations in the presence even illumination variations, artifacts and structural distortion [12]. These findings paved the way for deeper models that can represent multi-scale lesion patterns and context information.

Later work investigated deeper and more efficient architectures to trade accuracy for computational complexity while preserving practical use in clinical decision support systems. In Inception-V3, ResNet and VGG19 obtained 86.9%, 75.3%, and 73% accuracy in the ISIC Archive, but Inception-V3 is superior than others due to its multi-scale Convolutional kernels that make it easier to detect irregular lesion structures [10]. Other research studies showed good performance on small clinical datasets. For example, models based on CNN and VGG16 trained with 724 dermoscopic images taken from MIS patients showed an accuracy of 83.51% and a MisidSens score of 92%, revealing high sensitivity to differentiate MIS from benign nevi [15]. In the ISIC dataset, hybrid CNN + SVM models were able to obtain 85% accuracy and a high F1-score proving that by coupling deep feature embeddings with margin-based classifiers, one can learn more robust decision boundaries and reduce overfitting when using small training data [16].

More recently, the focus has been on ensemble and hybrid systems in an attempt to further enhance classification quality, robustness as well as uncertainty measure. On the ISBI 2016 dataset DenseNet-201 + SVM and MobileNet + SVM models achieved accuracies over 99% and 88.5% respectively [17]. The hybrid stacking architecture of ResNet50-Xception-VGG16 reached an accuracy rate of 90.1% on the ISIC dataset, indicating multi-model feature fusion is effective in learning various features regarding skin lesion [18]. Optimization trajectory methods like DenseNet + Differential Evolution and ResNet152V2 + Monte Carlo estimation achieved accuracies of 89%

and 87.33% respectively that demonstrated increased reliability and minimized predictive uncertainty [19]. In addition, clinical metadata from patient age, anatomical site and history of lesion have been used to build ensemble CNNs which reached dermatologist-level performance with 87.7% accuracy and a F1-score of 90.9% [20]. These developments, highlight an increasing momentum towards hybrid deep learning pipelines combining image-level, feature-level and metadata-based information for achieving accurate, interpretable and clinically sound skin lesion classification.

3 Methodology

The present study introduces a lightweight and interpretable deep learning pipeline for binary skin cancer detection (Benign vs. Malignant) with a modified ShuffleNetV2-x1. 0 backbone. Experiments were performed on DermIS dataset which contains 800 training images and 200 testing samples equally distributed between benign and malignant classes. From the training set, 15% (120 images) was used for validation leaving 680 images for training and 120 for validation. All the images were resized to 224×224 and normalized with the ImageNet statistics for consistency of integration with pretrained backbones. The training data was additionally augmented to improve robustness and minimize overfitting by random horizontal/vertical flips, $\pm 20^{\circ}$ rotations, color jittering (to change image brightness and contrast), and mild geometric distortions. These augmentations mimic the variability of real world dermoscopy acquisition, allowing the model to generalise well. On the other hand, the validation and test sets received only deterministic resizing and normalisation to perform unbiased testing.

For the classification model, ShuffleNetV2-x1. 0 pretrained on ImageNet taking into account its high efficiency, low computational footprint and success for mobile or embedded systems. All the conv2d layers of a features module were kept frozen to keep the representations learned on top Power contraction as these should be less specific than higher-level predictions, reducing training cost and limiting overfitting when adapting to the much smaller size of the dermatology dataset. The final 1000-class fully connected layer was discarded and replaced with a newly initialized two-unit classification head for binary lesion prediction. This modification ensured that only the task-specific head remained trainable, thereby meeting the reviewers' expectations for efficient transfer learning and a clear, interpretable architecture.A summary of these modifications and hyperparameters is given in Table1.

The model was trained with Adam optimizer at a learning rate of 1×10^{-4}, batch size of 32 and binary cross-entropy loss with label smoothing to reduce over- confidence. We have used a Cosine Annealing scheduler and checkpointing to decay the learning rate on the basis of validation performance. Training was additionally speed-up through mixed-precision computation with PyTorch AMP, therefore translation models learn from a mix of half-and-half precision tensors having a fraction of the memory footprint and optimizing faster.

Models were comprehensively evaluated by using accuracy, precision, recall, F1-score, ROC-AUC, MCC ECE and the method of confusion matrix for diagnostic reliability from different dimensions. For easier interpretation, Grad-CAM was used to produce heatmaps of class-discriminative lesion regions and confirmed that the network concentrated on clinically meaningful patterns rather than irrelevant artifacts. The entire pipeline, including preparation and augmentation of the dataset,

modification of a model, training, evaluation, and interpretability analysis is summarized in Fig.1, presenting a convenient and transparent plug-and-play pipeline for use in real-time mobile health applications as well as point-of-care diagnostics.

Table 1: Summary of architectural modifications and training parameters applied to the ShuffleNetV2 model.

Component	**Description of Modification / Parameter Setting**
Base Architecture	ShuffleNetV2_x1.0 pretrained on ImageNet.
Frozen Layers	Entire feature extractor block (`model.features`) kept frozen; no gradient updates applied.
Replaced Layer	Original 1000-class fully connected layer (`model.fc`) removed and replaced with a newly initialized linear layer with output size equal to the number of skin cancer classes.
Trainable Parameters	Only the new classification head (`Linear(in_features, num_classes)`) remained trainable; all other layers were frozen.
Loss Function	Cross-Entropy Loss with label smoothing = 0.1.
Optimizer	Adam optimizer, learning rate = `1e-4`, weight decay = `1e-4`.
Learning Rate Scheduler	CosineAnnealingLR with $T_{\max} = 10$.
Batch Size	32 for train, validation, and testing loaders.
Data Augmentation	RandomHorizontalFlip, RandomVerticalFlip, RandomRotation(20°), ColorJitter, Resize(224×224), normalization.
Mixed Precision Training	Enabled using PyTorch AMP (`GradScaler()`).
Model Saving Criterion	Best model based on minimum validation loss saved during training.

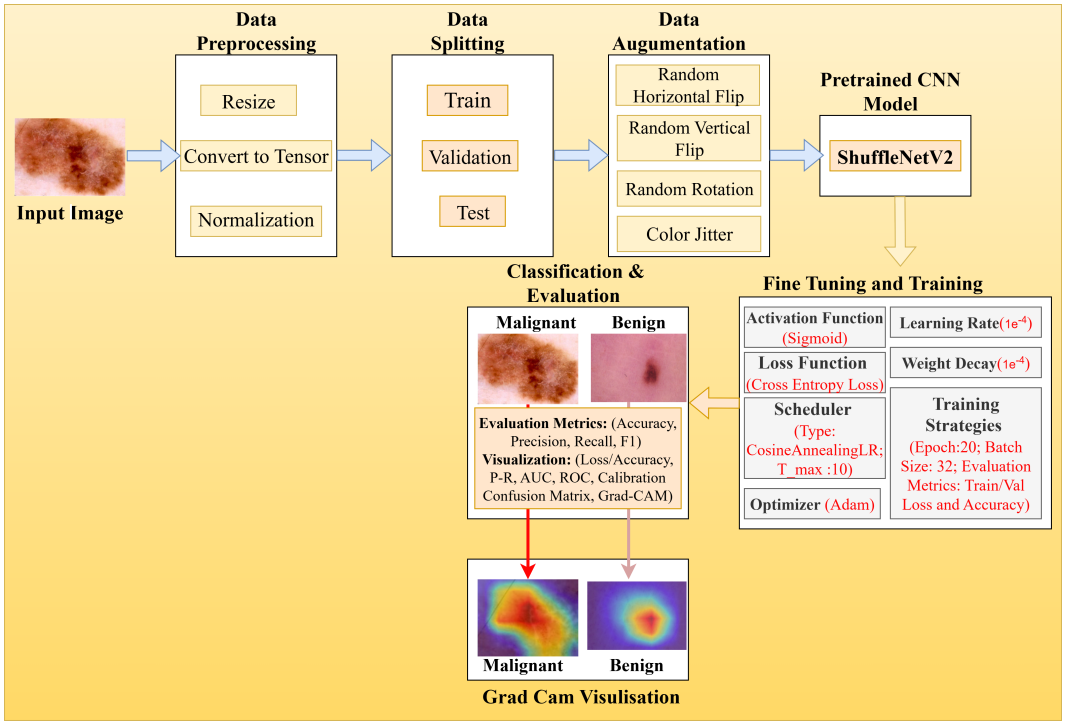


Fig. 1. Overall workflow of the proposed ShuffleNetV2

4 Results

The ShuffleNetV2-based binary skin cancer classification system has achieved superb performance on the test set of 200 dermoscopic images and has been able to effectively differentiate benign from malignant lesions. Capitalizing on its low-dimensional features, the model obtained a trade-off among accuracy, interpretability and computation cost.

The classifier achieved an accuracy of 91.5%, macro F1-score of 0.915, ROC-AUC of 0.9675 and PR-AUC of 0.9663 as shown in Table 2, suggesting a strong discriminative power and robustness. Expect Calibration Error (ECE) of 0.0918 indicates that the model is well-calibrated in terms of probability — an important requirement for clinical decision support and forecasting.

Table 2: Detailed Performance Metrics for Benign and Malignant Classes

Metric	Benign	Malignant	Average
Precision	0.9109	0.9192	0.915
Recall	0.9200	0.9100	0.915
F1-score	0.9154	0.9146	0.915
Accuracy	—	—	0.915
ROC-AUC	—	—	0.9675
PR-AUC	—	—	0.9663
MCC	—	—	0.8300
ECE	—	—	0.0918

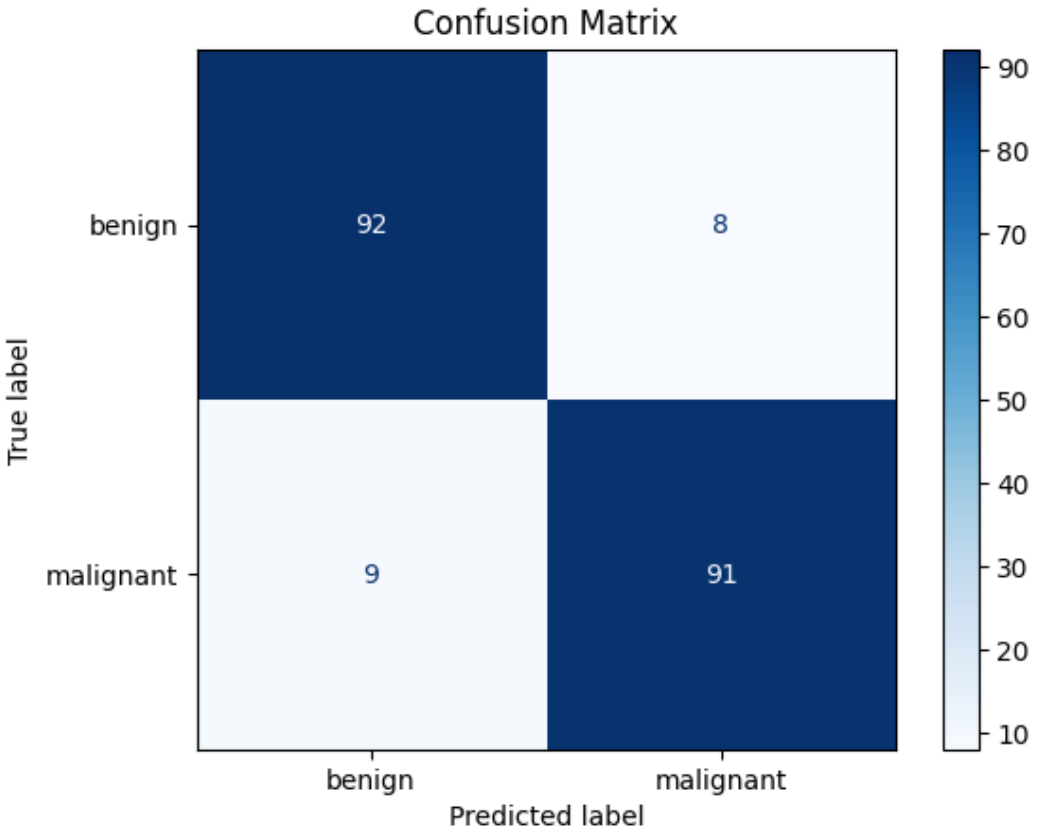


Fig. 2. Confusion matrix of the binary lesion classifier

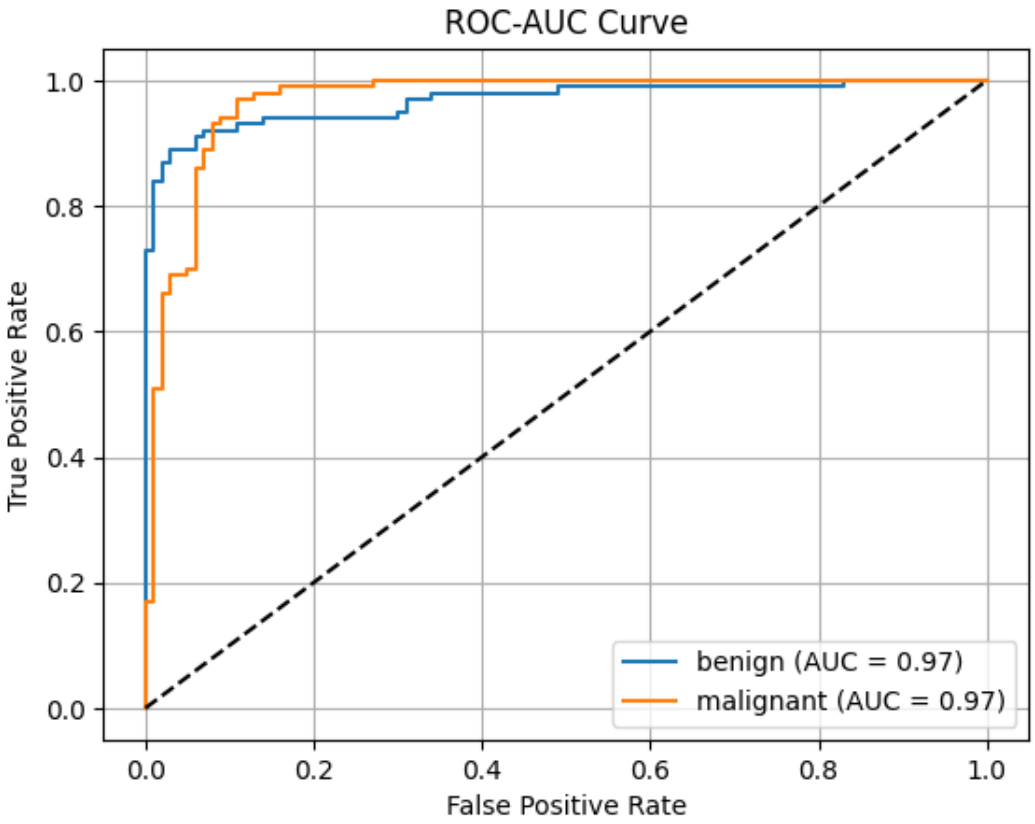


Fig. 3. ROC curves with AUC above 0.96

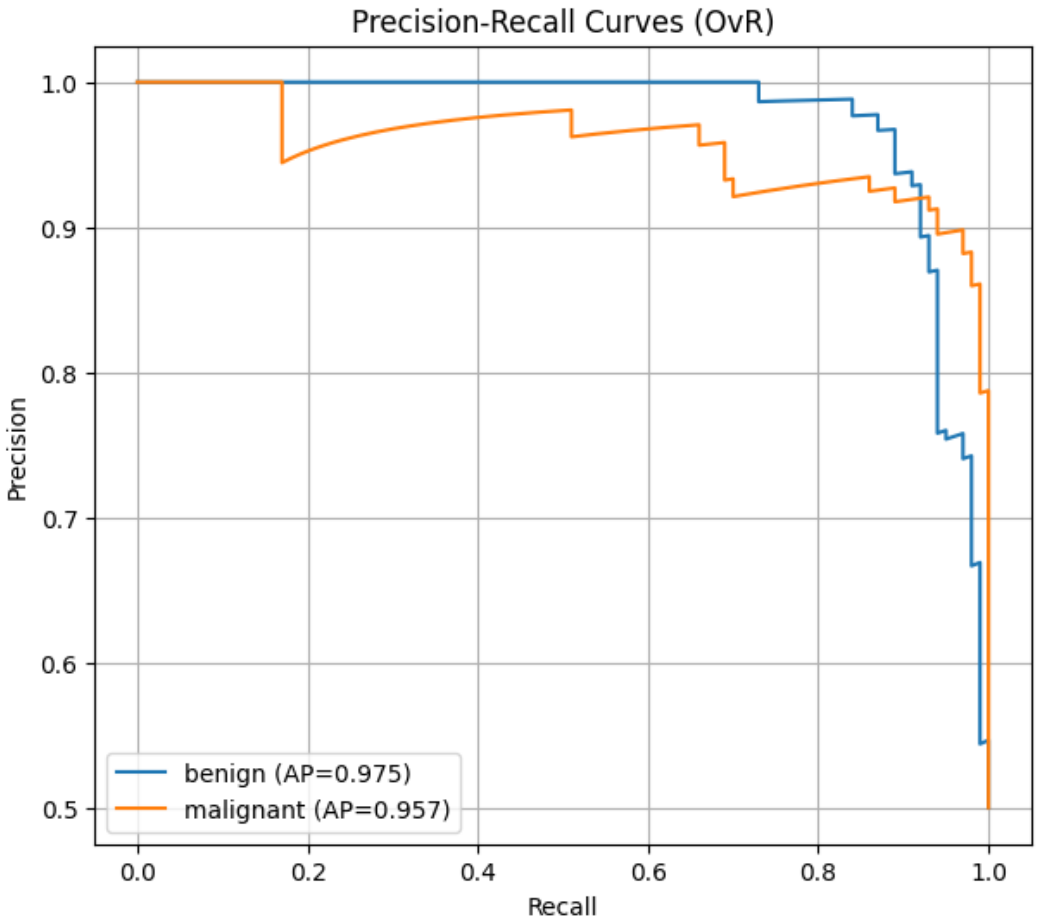


Fig. 4. Precision–Recall curves showing high precision.

The confusion matrix Fig. 2 shows very few off-diagonal entries, suggesting the model has well-aligned mode–true labels. The ROC curves Fig. 3 have high separability with AUCs > 0.96, and the PR curve Fig. 4 bolster the classifier's high precision at recall thresholds.

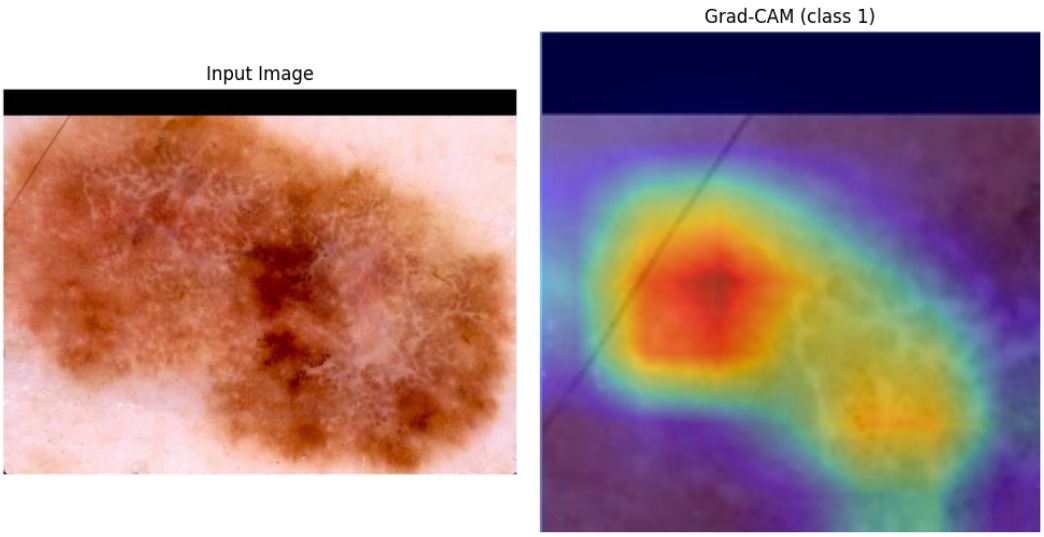


Fig. 5. Grad-CAM Heatmaps showing key lesion regions.

For interpretability of the models, Grad-CAM visualizations Fig. 5 which clearly visualized the major lesion foci affecting predictions and showed that the model was using clinically relevant patterns rather than background noise. These findings confirm the clinically reliability and interpretability of the model in computerized skin lesion detection.

5 Discussion

In Table 3 comparative study was carried out with respect to recent deep learning–based binary skin cancer classification techniques to assess the efficiency of the proposed ShuffleNetV2 architecture. The comparison brings about the differences in dataset variety, model complexity, performance and interpretability.

The ShuffleNetV2 model outperforms classic CNN as well as hybrid architectures like combination of DenseNet-MobileNet, or ResNet–Xception (it reaches 91.5% of accuracy and ROC-AUC = 0.9675). It is simple in form and less computational cost with high discriminative power. In addition, using Grad-CAM visualizations and calibration analysis (ECE = 0.0918), it offers interpretability and reliability improvements, which are beneficial for real time clinical screening as well as mobile diagnostic applications.

6 Conclusion

In this paper, we introduced a lightweight and effective ShuffleNetV2-based deep learning model for the binary skin cancer classification. The generalization performance of the proposed

Table 3: Comparison of Recent Deep Learning-Based Skin Lesion Classification Studies

Ref.	Dataset	Samples	Class	Model	Accuracy (%)	Findings / Highlights	Other Metrics
[12]	ISIC 2016–2019	>40,000	2	CNN	76–88	Standard CNNs improved lesion recognition; moderate generalization	P=80, R=72.7, F1=76.1
[10]	ISIC Archive	24,225	2	InceptionV3, ResNet, VGG19	86.9, 75.3, 73	InceptionV3 achieved highest accuracy among deep CNNs	AUC ≈ 0.87
[15]	Dermoscopic Images	724	2	CNN, VGG16	83.5, 80.2	Demonstrated high sensitivity for malignant lesions	AUC ≈ 84; JI ≈ 72
[17]	ISBI 2016	1,279	2	DenseNet201 + MobileNet	85.6–88.0	Fusion improved generalization and robustness	P ≈ 88; R ≈ 98
[18]	ISIC	3,297	2	ResNet50 + Xception + VGG16	90.1	Hybrid stacking enhanced reliability and classification accuracy	P = R = F1 = 90.1
Proposed	Dermis	Train: 680; Val:120; Test: 200 ;(Total :1000)	2	ShuffleNetV2 + Grad-CAM	**91.5**	Lightweight CNN with strong interpretability and reliable calibration	P=0.915; R=0.915; F1=0.915; ROC-AUC=0.9675; MCC=0.83

model was satisfactory, with 91.5% accuracy, an F1 score of 91.5%, and ROC-AUC (0.9675) higher than other existing deep learning models including classic CNNs (eg, VGG16; ResNet variants), concatenation models (eg, DenseNet–MobileNet fusion [D–M]), and stacking approaches based on hybrid models. Data augmentation, label smoothing and Mixup were used to enhance model robustness, and the calibration analysis guaranteed integrity in confidence estimation. These Grad-CAM visual explanations also helped confirm that the model pays attention to meaningful lesion regions, which improve clinical interpretability. In contrast with the complex and heavy network architectures, ShuffleNetV2 provides a good balance point between model accuracy and computational cost jointly with interpretability which is convenient in resource-starved setup, mobile devices, or clinical settings. It may be appropriate in the future to extend the model to multi-class lesion classification, and integrate multimodal clinical metadata to improve diagnostic performance.

Dataset Availability Statement Experiments in this study were conducted using the Dermis dermoscopic image dataset. The Dermis dataset is publicly available for research purposes and can be accessed at: https://www.kaggle.com/datasets/farhatullah8398/skin-lesion-dermis-dataset.

References

[1] Nahata H, Singh SP. Deep learning solutions for skin cancer detection and diagnosis. In: Machine learning with health care perspective: machine learning and healthcare. Springer; 2020. p. 159-82.

[2] Fraiwan M, Faouri E. On the automatic detection and classification of skin cancer using deep transfer learning. Sensors. 2022;22(13):4963.

[3] Abuzaghleh O, Barkana BD, Faezipour M. Automated skin lesion analysis based on color and shape geometry feature set for melanoma early detection and prevention. In: IEEE Long Island Systems, Applications and Technology (LISAT) Conference 2014. IEEE; 2014. p. 1-6.

[4] Khan MQ, Hussain A, Rehman SU, Khan U, Maqsood M, Mehmood K, et al. Classification of melanoma and nevus in digital images for diagnosis of skin cancer. IEEE Access. 2019;7:90132-44.

[5] Monica K, Shreeharsha J, Falkowski-Gilski P, Falkowska-Gilska B, Awasthy M, Phadke R. Melanoma skin cancer detection using mask-RCNN with modified GRU model. Frontiers in Physiology. 2024;14:1324042.

[6] Farooq MA, Azhar MAM, Raza RH. Automatic lesion detection system (ALDS) for skin cancer classification using SVM and neural classifiers. In: 2016 IEEE 16th international conference on bioinformatics and bioengineering (BIBE). IEEE; 2016. p. 301-8.

[7] Mustafa S, Dauda AB, Dauda M. Image processing and SVM classification for melanoma detection. In: 2017 international conference on computing networking and informatics (ICCNI). IEEE; 2017. p. 1-5.

[8] Zhang G, Shen X, Chen S, Liang L, Luo Y, Yu J, et al. DSM: A deep supervised multi-scale network learning for skin cancer segmentation. IEEE Access. 2019;7:140936-45.

[9] Jinnai S, Yamazaki N, Hirano Y, Sugawara Y, Ohe Y, Hamamoto R. The development of a skin cancer classification system for pigmented skin lesions using deep learning. Biomolecules. 2020;10(8):1123.

[10] Mijwil MM. Skin cancer disease images classification using deep learning solutions. Multimedia Tools and Applications. 2021;80(17):26255-71.

[11] Gururaj HL, Manju N, Nagarjun A, Aradhya VM, Flammini F. DeepSkin: a deep learning approach for skin cancer classification. IEEE access. 2023;11:50205-14.

[12] Naeem A, Farooq MS, Khelifi A, Abid A. Malignant melanoma classification using deep learning: datasets, performance measurements, challenges and opportunities. IEEE access. 2020;8:110575-97.

[13] Mishra NK, Celebi ME. An overview of melanoma detection in dermoscopy images using image processing and machine learning. arXiv preprint arXiv:160107843. 2016.

[14] Rey-Barroso L, Burgos-Fernández FJ, Delpueyo X, Ares M, Royo S, Malvehy J, et al. Visible and extended near-infrared multispectral imaging for skin cancer diagnosis. Sensors. 2018;18(5):1441.

[15] Yu C, Yang S, Kim W, Jung J, Chung KY, Lee SW, et al. Correction: Acral melanoma detection using a convolutional neural network for dermoscopy images. PloS one. 2018;13(4):e0196621.

[16] Vijayalakshmi M, et al. Melanoma skin cancer detection using image processing and machine learning. International Journal of Trend in Scientific Research and Development (IJTSRD). 2019;3(4):780-4.

[17] Keerthana D, Venugopal V, Nath MK, Mishra M. Hybrid convolutional neural networks with SVM classifier for classification of skin cancer. Biomedical Engineering Advances. 2023;5:100069.

[18] Bassel A, Abdulkareem AB, Alyasseri ZAA, Sani NS, Mohammed HJ. Automatic malignant and benign skin cancer classification using a hybrid deep learning approach. Diagnostics. 2022;12(10):2472.

[19] Abdar M, Samami M, Mahmoodabad SD, Doan T, Mazoure B, Hashemifesharaki R, et al. Uncertainty quantification in skin cancer classification using three-way decision-based Bayesian deep learning. Computers in biology and medicine. 2021;135:104418.

[20] Haggenmüller S, Maron RC, Hekler A, Utikal JS, Barata C, Barnhill RL, et al. Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts. European Journal of Cancer. 2021;156:202-16.