A Two-Stage Deep Learning Framework for Skin Lesion Detection and Classification Using ResNet18 and EfficientNet-B4

Abid Muhammad
Dept. of Computer Science
Amal Jyothi College of Engineering
(Autonomous)
Kottayam, India
abidmuhammad2025@cs.ajce.in

Alan Abdul Gafar
Dept. of Computer Science
Amal Jyothi College of Engineering
(Autonomous)
Kottayam, India
alanabdulgafar2025@cs.ajce.in

Abin Sunny
Dept. of Computer Science
Amal Jyothi College of Engineering
(Autonomous)
Kottayam, India
abinsunny2025@cs.ajce.in

Abin Melvin
Dept. of Computer Science
Amal Jyothi College of Engineering
(Autonomous)
Kottyam, India
abinmelvin2025@cs.ajce.in

Mr. Bibin Varghese
Dept. of Computer Science
Amal Jyothi College of Engineering
(Autonomous)
Kottyam, India
bibinvarghese@amaljyothi.ac.in

***Abstract*—Skin diseases encompass a wide range of conditions that require an early and accurate diagnosis for effective treatment. This paper presents a two-stage deep learning framework for automated skin lesion detection and classification using deep convolutional neural networks. The first stage uses a ResNet18 model to detect the presence of a lesion in dermoscopic images. If a lesion is detected, the image is transferred to an EfficientNet-B4 model for multiclass classification. Our approach integrates data augmentation, hair removal preprocessing, learning rate scheduling, and early stopping to enhance model performance and robustness. The framework is trained and evaluated on the HAM10000 dataset, addressing challenges such as class imbalance, model fine-tuning, and overfitting. Experimental results demonstrate the effectiveness of this method in accurately identifying and categorizing skin lesions, contributing to the advancement of deep learning-based dermatological diagnosis.[1]**

***Index Terms*—Deep learning, Convolutional Neural Networks, Skin Lesion Detection, Skin Lesion Classification, ResNet18, EfficientNet-B4, Medical image analysis, Computer-aided diagnosis, Dermatology, HAM10000 data set, image preprocessing, data enhancement.**

I. Introduction

A. Background:

Skin diseases affect millions of people worldwide, ranging from benign conditions to life-threatening malignancies. Early and accurate diagnosis is crucial for effective treatment and improved patient outcomes. Traditionally, dermatologists have relied on clinical and dermoscopic examinations to identify and classify skin lesions. However, manual diagnosis is time-consuming, requires expertise, and is subject to variability between observers. With advances in deep learning and computer vision, automated systems have shown significant potential to assist dermatologists by providing an objective and consistent analysis of skin lesions. Convolutional Neural Networks (CNNs) have become the state-of-the-art approach for medical image analysis, demonstrating high accuracy in lesion detection and classification tasks.

B. Problem Statement:

Despite the success of deep learning in medical imaging, the classification of skin lesions remains challenging due to several factors:

1) *Class Imbalance:* Some types of lesion are underrepresented in the data sets, leading to biased predictions.
2) *Variability in Lesion Appearance:* Lesions differ in size, shape, color, and texture, making classification difficult.
3) *Overfitting:* Due to limited training data, deep learning models can memorize patterns instead of generalizing well.
4) *Computational Efficiency:* Deploying deep learning models on real-world devices, such as mobile applications, requires optimization for speed and efficiency.

To address these challenges, we propose a two-stage deep learning framework that first detects the presence of a lesion and then classifies it into predefined categories. This hierarchical approach ensures that only relevant images are passed to the classification stage, improving the efficiency and accuracy of the model.

C. Objectives:

The primary objective of this study is to develop a robust and efficient deep learning-based framework for the detection and classification of skin lesions. Specifically, we aim to:

DOI: 10.5281/zenodo.15529405

1) Develop a ResNet18-based model for accurate lesion detection in dermoscopic images.
2) Train an EfficientNet-B4 model for multi-class skin lesion classification.
3) Implement preprocessing techniques, including hair removal and data augmentation, to improve generalization.
4) Address class imbalance using techniques such as stratified k-fold cross-validation and loss function adjustments.
5) Optimize model training and evaluation with learning rate scheduling and early stopping to prevent overfitting.

By achieving these objectives, we contribute to the advancement of computer-aided dermatological diagnosis, providing a reliable and scalable solution for automated skin lesion analysis.

II. Related Works

Skin lesion detection and classification have been widely explored using deep learning techniques, particularly convolutional neural networks (CNNs). Traditional machine learning approaches relied on handcrafted features, but these methods often struggled with variations in lesion appearance, lighting conditions, and skin tone. With the rise of deep learning, pretrained models and transfer learning have significantly improved classification accuracy. Researchers have experimented with different architectures, data augmentation techniques, and loss functions to address challenges such as class imbalance and overfitting. Despite these advancements, there is still a need for efficient and accurate models that can generalize well across diverse skin lesion types.

Ahammed et al. [1] proposed a machine learning-based approach to the detection and classification of skin diseases using image segmentation. Their method involved digital hair removal using morphological filtering, feature extraction through the Gray Level Co-occurrence Matrix (GLCM), and classification using Decision Tree (DT), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). The study demonstrated that SVM outperformed the other classifiers, achieving high accuracy on the ISIC 2019 and HAM10000 datasets.

Rafay et al. [2] introduced EfficientSkinDis, an EfficientNet-based deep learning model for the classification of skin diseases. The study utilized a manually curated dataset comprising 31 skin diseases, merged from Atlas Dermatology and ISIC datasets. The authors applied transfer learning with CNN architectures, including EfficientNet, ResNet, and VGG, and found that EfficientNet-B2 achieved the highest accuracy of 87.15% after data augmentation. The model was later deployed on a web server for real-time diagnosis [2].

Galisot et al. [3] proposed an interactive machine learning framework for medical image segmentation that integrates both visual and spatial anatomical knowledge. Their approach combines region-specific voxel classifiers and a spatial relationship learning mechanism to improve segmentation accuracy and adaptability. The method demonstrated robustness in segmenting various anatomical structures with limited training data, reducing the need for global image registration and minimizing user interaction time [3].

Kumar et al. [4] proposed a novel mixed-domain hand-crafted feature extraction approach for skin disease recognition using a multiheaded 1-D Convolutional Neural Network (CNN). Their method integrates spatial and spectral domain features, including spectrogram and cepstrum analysis, to enhance classification accuracy. The proposed model was evaluated on the HAM10000 and Dermnet datasets, achieving classification accuracies of 89.71% and 88.57%, respectively, outperforming existing state-of-the-art techniques [4].

Inthiyaz et al. [5] proposed an automated deep learning-based approach for skin disease detection and classification using Convolutional Neural Networks (CNNs). The study utilized VGG16 and VGG19 architectures with a binary cross-entropy loss function for image classification. The model was trained on the Xiangya-Derm dataset and demonstrated an accuracy improvement from 81.36% to 87.42% using optimized image pre-processing and feature extraction techniques [5].

Reddy et al. [6] proposed an enhanced skin disease detection framework utilizing optimized region-growing segmentation with Grey Wolf Optimization (GWO) and an autoencoder-based classification model. Their approach extracted texture features using the Gray Level Co-occurrence Matrix (GLCM) and Weber Local Descriptor (WLD), reducing dimensionality with autoencoders before classification using a convolutional neural network (CNN). The experimental results demonstrated improved accuracy compared to traditional deep learning classifiers such as CNN, RNN, and LSTM [6].

Velasco et al. [7] explored the effectiveness of transfer learning in Convolutional Neural Networks (CNNs) for skin disease classification. Their study collected 3,400 images of seven common skin diseases and evaluated multiple pre-trained CNN architectures, including VGG16, VGG19, ResNet50, InceptionV3, DenseNet, and MobileNet. MobileNet achieved the highest accuracy of 94.1%, demonstrating its suitability for real-time skin disease detection applications on mobile devices [7].

Bhaminie et al. [8] proposed a modified ResNet-50 architecture for the classification of skin diseases, focusing on three skin conditions: basal cell carcinoma, benign keratosis, and melanoma — using the ISIC 2019 dataset. Their study applied transfer learning by fine-tuning ResNet-50 with global average grouping and dense layers, achieving improved diagnostic accuracy. Additionally, data augmentation techniques such as rotation, shifting, and zooming were used to enhance model generalization. The results demonstrated that the modified ResNet-50 model effectively improved skin disease classification performance compared to traditional deep learning approaches [8].

III. Methodology

A. Overall Framework

The proposed system is a two-stage deep learning framework for automated skin lesion detection and classification. The first stage uses a ResNet18 model to detect if a lesion is present in the input image. If a lesion is detected, the image is passed to an EfficientNet-B4 model, which classifies it into a specific category of skin lesion. This hierarchical approach improves computational efficiency by filtering out non-lesion images before classification. The workflow consists of four major steps:

1) Dataset Preparation: Sorting, balancing, and preprocessing dermoscopic images.

2) Lesion Detection: Identifying lesion presence using ResNet18.

3) Lesion Classification: Categorizing lesions using EfficientNet-B4.

4) Model Optimization: Enhancing performance with data augmentation, learning rate scheduling, and early stopping.

B. Dataset Preparation

The HAM10000 dataset was used, a widely used benchmark for the classification of skin lesions. The dataset contains 10,015 dermoscopic images covering seven different types of lesion. However, the dataset is highly imbalanced, with some classes significantly underrepresented. To address this, we implemented data preprocessing as follows:

1) Dataset Sorting: The raw dataset consists of images stored in a single directory with corresponding metadata. We sorted the images into diagnosis-specific folders based on the metadata file, making it easier to manage class-wise samples during training.

2) Dataset Balancing: To mitigate class imbalance, we applied data augmentation using the imgaug library. Augmentations included:

- Geometric Transformations – Random flipping and rotation (±20°).
- Blur Effects – Gaussian blurring to simulate real-world variability.
- Brightness Adjustments – Random brightness scaling (0.8× to 1.2×).
- Noise Injection – Adding Gaussian noise to increase robustness.

Augmentation was selectively applied to underrepresented classes until all categories had an equal number of images, ensuring a balanced dataset for training.

C. Model Architecture

1) Lesion Detection (ResNet18): A ResNet18 model was used for binary classification (lesion vs. normal). The model was pre-trained on ImageNet and fine-tuned for this task. The final fully connected layer was replaced with a single neuron output activated by a sigmoid function.

- Input Size: 224×224
- Loss Function: Binary Cross-Entropy with Logits Loss (BCEWithLogitsLoss)
- Optimizer: Adam
- Activation Function: Sigmoid

The model was trained for 5 epochs with a batch size of 32 and a learning rate of 0.001. If the model classified an image as containing a lesion, it was passed to the classification model.

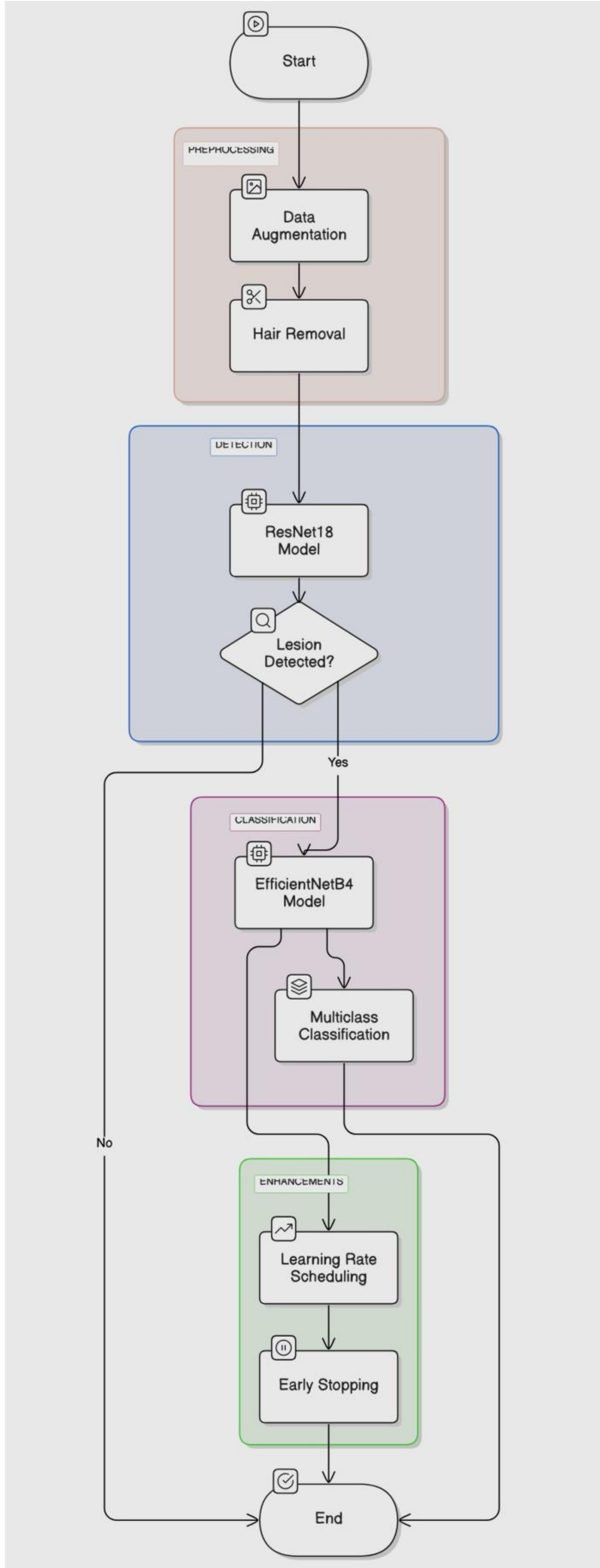

Fig. 1. Proposed Architecture

2) Lesion Classification (EfficientNet-B4): For multi-class lesion classification, we employed an EfficientNet-B4 model, which provides a good balance between accuracy and efficiency. The model was fine-tuned on the dataset, with the final classification layer modified to match the number of skin lesion categories.

- Input Size: 224×224
- Loss Function: Cross-Entropy Loss
- Optimizer: AdamW
- Learning Rate Scheduler: StepLR (step size = 5, gamma = 0.1)
- Regularization: Weight Decay (1e-4), Dropout, Early Stopping

To further enhance performance, we applied hair removal preprocessing, ensuring that the model focused on lesion features rather than irrelevant artifacts.

D. Training Strategy

1) Learning Rate Scheduling: A StepLR scheduler was applied to reduce the learning rate by a factor of 0.1 every 5 epochs, preventing overfitting and ensuring stable convergence.

2) Early Stopping: To prevent overfitting, we implemented an early stopping mechanism with a patience of 5 epochs. If the validation loss did not improve for 5 consecutive epochs, training was halted.

3) Model Evaluation: The trained models were evaluated using accuracy, precision, recall, and F1-score. The lesion detection model was assessed for binary classification performance, while the classification model was evaluated using a confusion matrix to analyze per-class accuracy.

IV. Discussions and Analysis

A. Detection Model Performance

The first stage of the proposed system is the detection model, responsible for distinguishing between normal skin and skin lesions. The model achieved:

- Accuracy: 99.64%
- Precision: 99.34%
- Recall: 100%
- F1-Score: 99.67%

The confusion matrix shows that the model correctly classified 123 normal images while misclassifying only 1 as a lesion. Additionally, all 150 lesion images were correctly identified, resulting in zero false negatives. This indicates an extremely high sensitivity (recall = 1.000), ensuring that no lesions go undetected. The low false positive rate (1 misclassified normal image) further supports the model's robustness.
The high performance can be attributed to the binary classification setup, which simplifies the problem compared to multi-class classification. The use of ResNet18, along with preprocessing techniques like hair removal and data augmentation, has contributed to this strong performance.

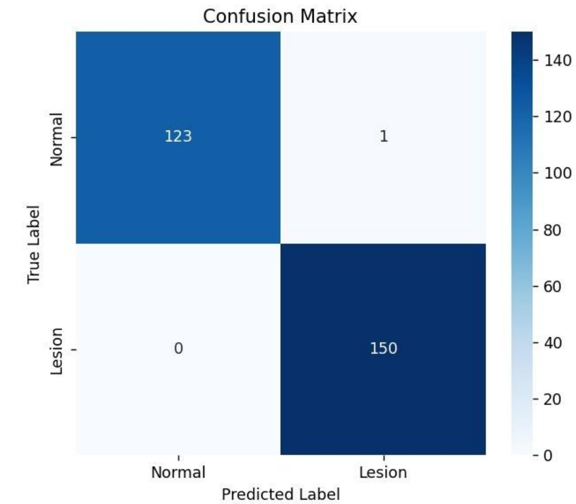

Fig. 2. Confusion matrix of the lesion detection model, showing its ability to differentiate between normal and lesion images with high accuracy.

B. Classification Model Performance

The second stage of the pipeline classifies detected lesions into specific categories. The classification model achieved an overall accuracy of 95%, with the following per-class performance:

TABLE I
PERFORMANCE METRICS FOR SKIN LESION CLASSIFICATION

Class	Precision	Recall	F1-Score
akiec	0.98	0.92	0.95
bcc	0.91	0.98	0.94
bkl	0.96	0.87	0.92
df	0.99	0.99	0.99
mel	0.88	0.95	0.91
nv	0.95	0.93	0.94
vasc	0.99	1.00	0.99

The confusion matrix highlights key insights:

- Strong classification performance for df (dermatofibroma) and vasc (vascular lesions) with almost perfect precision and recall.
- Melanoma (mel) and bkl (benign keratosis-like lesions) show more misclassifications, which may be due to their similarity with other types of lesion.
- Some confusion exists between bkl and nv (nevus), which could be improved with additional data or better augmentation strategies.
- The model maintains balanced performance across classes, with no single category performing drastically worse than others.

C. Overall Performance and Future Directions

The two-stage pipeline proves to be an effective approach, demonstrating high sensitivity in lesion detection to ensure

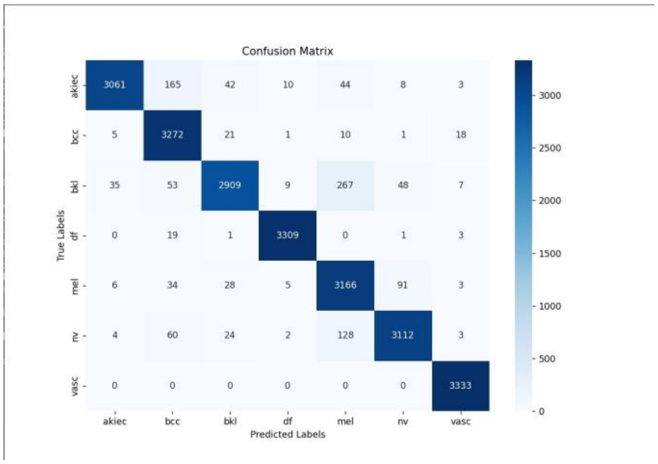

Fig. 3. Confusion matrix of the classification model, illustrating classification performance across multiple skin lesion categories.

that all potential lesions are accurately classified. Furthermore, the classification model achieves strong performance, with an overall accuracy of 95%. The deep learning models used in this pipeline efficiently balance accuracy with computational efficiency, making them well-suited for real-world applications. However, the following aspects should be considered for further enhancement:

1) External validation on real-world datasets to confirm generalizability.
2) Data augmentation improvements targeting misclassified classes.
3) Integration of attention mechanisms to highlight lesion regions.
4) Potential semi-supervised learning for better feature learning in limited data scenarios.

V. Conclusion

The proposed two-stage deep learning pipeline designed for skin lesion detection and classification has demonstrated remarkable effectiveness in accurately identifying and categorizing a wide range of skin lesions. The detection model exhibited exceptionally high sensitivity, ensuring that nearly all potential lesions, regardless of their type, size, or complexity, were correctly identified during the testing phase. The classification model achieved an outstanding precision of 95%, accompanied by consistently high precision, recall, and F1-scores across all lesion classes, which further establishes the reliability of the proposed system.

Furthermore, one of the notable strengths of the proposed deep learning models is their ability to strike a favorable balance between high classification accuracy and computational efficiency. This aspect is particularly important when considering the deployment of such systems in real-world clinical or telemedicine environments where fast and accurate results are imperative. The use of advanced image preprocessing techniques, effective data augmentation strategies, and carefully fine-tuned deep neural networks has significantly contributed to the resolution of some of the major challenges in medical image analysis, such as class imbalance, overfitting, and noise in image data.

Despite the promising results obtained from the proposed approach, there remains significant scope for further improvement. Future work could focus on incorporating more sophisticated ensemble learning techniques, which could potentially enhance the predictive performance of the system by combining the outputs of multiple complementary models.

In conclusion, the proposed deep learning-based pipeline offers a highly effective and computationally efficient solution for the automated detection and classification of skin lesions. Its demonstrated ability to accurately differentiate between benign and malignant lesions, along with its potential for real-world deployment, makes it a valuable tool in the field of dermatology and telemedicine.

References

[1] Ahammed, M., Mamun, M. A., & Uddin, M. S. (2022). A machine learning approach for skin disease detection and classification using image segmentation. Healthcare Analytics, 2, 100122.

[2] A. Rafay and W. Hussain, ”EfficientSkinDis: An EfficientNet-based classification model for a large manually curated dataset of 31 skin diseases,” Biomedical Signal Processing and Control, vol. 85, p. 104869, 2023. [Online]

[3] G. Galisot, J.-Y. Ramel, T. Brouard, E. Chaillou, and B. Serres, ”Visual and structural feature combination in an interactive machine learning system for medical image segmentation,” Machine Learning with Applications, vol. 8, p. 100294, 2022.

[4] A. Kumar, A. Vishwakarma, V. Bajaj, and S. Mishra, ”Novel Mixed Domain Hand-Crafted Features for Skin Disease Recognition Using Multiheaded CNN,” IEEE Transactions on Instrumentation and Measurement, vol. 73, 2024, Art. no. 5013013.

[5] S. Inthiyaz, B. R. Altahan, S. H. Ahammad, V. Rajesh, R. R. Kalangi, L. K. Smirani, M. A. Hossain, and A. N. Z. Rashed, ”Skin disease detection using deep learning,” Advances in Engineering Software, vol. 175, p. 103361, 2023.

[6] D. A. Reddy, S. Roy, S. Kumar, and R. Tripathi, ”A Scheme for Effective Skin Disease Detection using Optimized Region Growing Segmentation and Autoencoder-based Classification,” Procedia Computer Science, vol. 218, pp. 274–282, 2023.

[7] J. S. Velasco, J. V. Catipon, E. G. Monilar, V. M. Amon, G. C. Virrey, and L. K. S. Tolentino, ”Classification of Skin Disease Using Transfer Learning in Convolutional Neural Networks,” International Journal of Emerging Technology and Advanced Engineering, vol. 13, no. 4, pp. 1-7, Apr. 2023.

[8] P. Bhaminie and Y. C., ”Skin Disease Classification using Transfer Learning with a Modified ResNet-50 Architecture,” 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), 2024.

[9] MP.N.Srinivasu, J.G.Sivasai, M.F.Ijaz, A.K.Bhoi, W.Kim,and J.J.Kang, ”Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM”, Sensors, vol. 21, no. 8, pp. 2852, Apr. 2021

[10] J. Alam, ”An efficient approach for skin disease detection using deep learning”, in Proc. IEEE Asia–Pacific Conf. Comput. Sci. Data Eng.(CSDE), pp. 1–8, 2021

[11] Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results, Institute for Health Metrics and Evaluation (IHME), Seattle, WA, USA, 2020

[12] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, ”MobileNets: Efficient convolutional neural networks for mobile vision applications”, 2017, arXiv:1704.04861.