

Original Paper

Deep Learning–Based Classification of Melanoma Using Dermoscopic Images: Comparative Evaluation of Convolutional Neural Network and DenseNet Models

Harshal Sanghvi¹, PhD; Argyrios Agritis², BS; Isabella Alvarez², BS; Mitsu Patel², MBBS; Shuwen Wang³, PhD; Riki Patel², PhD; Raphael Banoub⁴, MD; Timonthy Makkar⁵, MD

¹Florida Atlantic University, Boca Raton, FL, United States

²Department of Technology and Clinical Trials, Advanced Research LLC, Pompano Beach, FL, United States

³Clemson University, Clemson, SC, United States

⁴Mayo Clinic in Florida, Jacksonville, FL, United States

⁵Department of Dermatology, Rutgers New Jersey Medical School, Brunswick, NJ, United States

Corresponding Author:

Harshal Sanghvi, PhD
Florida Atlantic University
777 Glades Road
Boca Raton, FL 33431
United States
Phone: 1 5615634452
Email: harshalsanghvi7@gmail.com

Abstract

Background: For melanoma, an aggressive form of skin cancer, early and accurate detection is critical to improving clinical outcomes. Deep learning–based image analysis has emerged as a promising approach for computer-assisted evaluation of dermoscopic lesions. However, model performance is often affected by limited dataset size, class imbalance, overfitting, and differences in network architecture and optimization strategies.

Objectives: This study was aimed at developing and comparatively evaluating convolutional neural network (CNN) and DenseNet-based deep learning models for automated classification of dermoscopic skin lesion images. The effects of data augmentation, network architecture, optimizer selection, and training duration on classification performance were examined.

Methods: Publicly available dermoscopic skin lesion images obtained from Kaggle were evaluated using a conventional CNN, DenseNet169, and DenseNet201. The original dataset contained 2357 images representing multiple benign and malignant skin lesion categories. Images were resized to 250×250 pixels, labeled, and randomly partitioned into training and testing sets at an 80:20 ratio. To address class imbalance, image augmentation using transformations, including $\pm 10^\circ$ shear, was performed, increasing the dataset size to 6857 images, with approximately 500 additional samples generated per class. Models were trained using adaptive moment estimation (Adam) and root mean square propagation (RMSprop) optimizers, and evaluated after 20, 60, and 100 epochs. Performance was assessed using accuracy, precision, recall (sensitivity), specificity, F_1 -score, confusion matrices, and training and validation loss.

Results: Data augmentation substantially improved model performance and reduced evidence of overfitting compared with the original dataset. On the augmented dataset, DenseNet169 and DenseNet201 achieved accuracy and F_1 -scores exceeding 94% after 100 epochs, consistently outperforming the conventional CNN. DenseNet169 achieved up to 95% validation accuracy, with only minor performance differences between the Adam and RMSprop optimizers. The CNN showed greater susceptibility to overfitting on the smaller dataset; its F_1 -score, calculated over the complete dataset, increased from 31% after 20 epochs to 59% after 60 epochs and 87% after 100 epochs before augmentation. Across the comparative experiments, the DenseNet architectures demonstrated more stable validation performance and better generalization following augmentation.

Conclusions: DenseNet169 and DenseNet201 demonstrated stronger and more stable classification performance than the conventional CNN for dermoscopic skin lesion classification, particularly after augmentation of the training dataset. Increasing the dataset size and improving class balance led to reduced overfitting and improved model performance. These findings support further investigation of DenseNet-based approaches as potential clinical decision support tools for melanoma and skin

lesion assessment. However, prospective evaluation using independent, multicenter datasets and diverse patient populations is required before clinical implementation.

JMIR Dermatol 2026;9:e104086; doi: [10.2196/104086](https://doi.org/10.2196/104086)

Keywords: deep learning ; dermoscopy ; medical image analysis ; data augmentation ; melanoma detection

Introduction

Skin cancer is one of the most prevalent malignancies globally, accounting for approximately 40%-50% of all cancer diagnoses within the United States. It is broadly categorized into melanoma and nonmelanoma skin cancers (NMSCs) [1]. NMSCs primarily originate from keratinized epithelial cells and include basal cell carcinoma, squamous cell carcinoma, and sebaceous gland carcinoma. Although melanoma is less common, it represents the most aggressive form of skin cancer and accounts for the majority of skin cancer-related deaths [2].

Melanoma arises from melanocytes, which are pigment-producing cells localized in the basal layer of the epidermis. Exposure to UV radiation induces keratinocytes to secrete α -melanocyte-stimulating hormone, which activates melanocytes via the melanocortin 1 receptor, thereby promoting melanin synthesis [3]. This physiological response confers photoprotective properties; however, chronic UV exposure can promote DNA damage and mutagenesis, contributing to malignant transformation and unregulated cellular proliferation. Epidemiological data reveal sex- and age-specific variations in melanoma incidence: incidence is higher in females than in males before the age of 40 years, whereas by the age of 75 years, incidence in males is nearly 3 times that in females. Both genetic predisposition and environmental factors contribute to melanoma pathogenesis, with UV radiation implicated in approximately 60%-70% of cases. The World Health Organization classifies UVA (315-400 nm) and UVB (280-315 nm) radiation as complete carcinogens, given their involvement in tumor initiation and progression [4].

Clinically, melanoma is classified into distinct histopathological variants, of which superficial spreading melanoma is the most common, especially in populations with lighter skin phototypes [5]. Despite advancements in understanding melanoma biology, its global incidence continues to rise, emphasizing the importance of early detection in improving patient prognosis. Diagnostic assessment relies predominantly on dermoscopic evaluation, supplemented by heuristic approaches such as the ABCDE (asymmetry, border

irregularity, color heterogeneity, diameter, and evolution) criteria. Nevertheless, these methods are inherently subjective, depend on clinical expertise, and are susceptible to interobserver variability [6].

To address these diagnostic challenges, AI and computer-aided diagnosis systems have been developed for automated skin lesion analysis. Deep learning frameworks, particularly convolutional neural networks (CNNs), have demonstrated diagnostic performance comparable to that of experienced dermatologists [7]. Model performance has been improved by incorporating transfer learning and sophisticated feature extraction techniques. Public datasets, such as those from the International Skin Imaging Collaboration and the PH² database, have been instrumental in training and validating AI models [8].

Complementary to algorithmic innovations, the use of image preprocessing techniques, such as top-hat filtering and DullRazor for hair artifact removal and noise reduction, has substantially improved lesion segmentation and classification accuracy [9]. Clinical technologies, including digital dermoscopy, total body photography, and mobile health platforms, have enhanced the accessibility of skin surveillance. However, challenges remain regarding cost, patient adherence to follow-up, and variability in diagnostic precision [10].

In this study, we propose a novel deep learning framework that integrates transfer learning with densely connected neural network (DenseNet) architectures (DenseNet169 and DenseNet201) and a conventional CNN model for automated skin lesion classification. The framework iteratively optimizes hyperparameters to enhance diagnostic accuracy and is benchmarked against established machine learning counterparts. By combining rigorous preprocessing protocols with state-of-the-art neural network architectures, this work aims to promote early melanoma detection and support clinical decision-making.

Table 1 summarizes the performance metrics of the proposed approach and relevant methods reported in the literature.

Table 1. Performance overview of the proposed method and relevant methods reported in the literature.

Reference	Methods	Dataset/validation	Key findings	Sensitivity/accuracy
This study	DenseNet169, DenseNet201, and CNN ^a	Dermoscopic images	Validation and training accuracy greater than 98%	Accuracy 98%
Cheong et al [11], 2021	Feature extraction (entropy, LBP ^b , GLCM ^c , and ABCDE ^d criteria) with SVM-RBF ^e	Dermoscopic images	Classification accuracy greater than 97.5%	Sensitivity greater than 97.5%

Reference	Methods	Dataset/validation	Key findings	Sensitivity/accuracy
Waheed et al [12], 2023	Deep learning (CNN, hybrid CNN, and transfer learning) with data augmentation	ISIC ^f dataset	Correct classification rate of 81%	Sensitivity 92.46%; specificity 92.23%
Alsaade et al [13], 2021	Hybrid ML ^g and DL ^h (segmentation, CNN, and ANN ⁱ)	ISIC 2018 and PH ^{2j}	ANN accuracy of 98% (ISIC) and 97.5% (PH ²)	Sensitivity 83%-97.5%
Indraswari et al [14], 2022	Transfer learning with MobileNetV2	ISIC Archive	Lightweight model suitable for mobile deployment	Accuracy 85%; sensitivity 85%
Tumpa and Kabir [15], 2021	Image preprocessing (MGI ^k for hair removal), texture and shape features, and ANN	ISIC and PH ²	Benign versus malignant lesions classified with 97.7% accuracy	Sensitivity ~97%
Brinker et al [16], 2019	CNN compared with dermatologists	ISIC Archive	CNN sensitivity comparable to that of dermatologists; higher specificity (68.2% vs 64.4%)	Sensitivity 89.4%
Thaajwer and Ishanka [17], 2020	Image processing (GLCM and color/shape features)+SVM ^l	Dermoscopic images	Classification accuracy of 83%	Not reported
Banerjee et al [18], 2020	Deep learning, fuzzy logic, and DCNN ^m (YOLO ⁿ)	PH ² , ISBI 2017, ISIC 2019	High accuracy across datasets	Sensitivity 97.5%-98.5%
Bakheet and El-Nagar [19], 2021	Real-time computer-aided diagnosis with DNN ^o	PH ² dataset	Accuracy of ~97.5%	Sensitivity 100%
Chang et al [20], 2022	CNN, XGB ^p classifier, and SMOTE ^q -based balancing	Stratified 10-fold cross-validation	Accuracy of ~97.5%	Sensitivity 98.1%
Thiyaneswaran et al [21], 2021	GLCM features with ML classifiers	Dermoscopic images	Accuracy of ~90%	Positive prediction 95%; negative prediction 80%

^aCNN: convolutional neural network.

^bLBP: local binary pattern.

^cGLCM: gray level co-occurrence matrix.

^dABCDE: asymmetry, border irregularity, color heterogeneity, diameter, and evolution.

^eSVM-RBF: support vector machine with the radial basis function kernel.

^fISIC: International Skin Imaging Collaboration.

^gML: machine learning.

^hDL: deep learning.

ⁱANN: artificial neural network.

^jPH²: PH² Dermoscopic Image Database.

^kMGI: morphological gradient image.

^lSVM: support vector machine.

^mDCNN: deep convolutional neural network.

ⁿYOLO: you only look once.

^oDNN: deep neural network.

^pXGB: extreme gradient boosting.

^qSMOTE: synthetic minority oversampling technique.

Methods

Overview

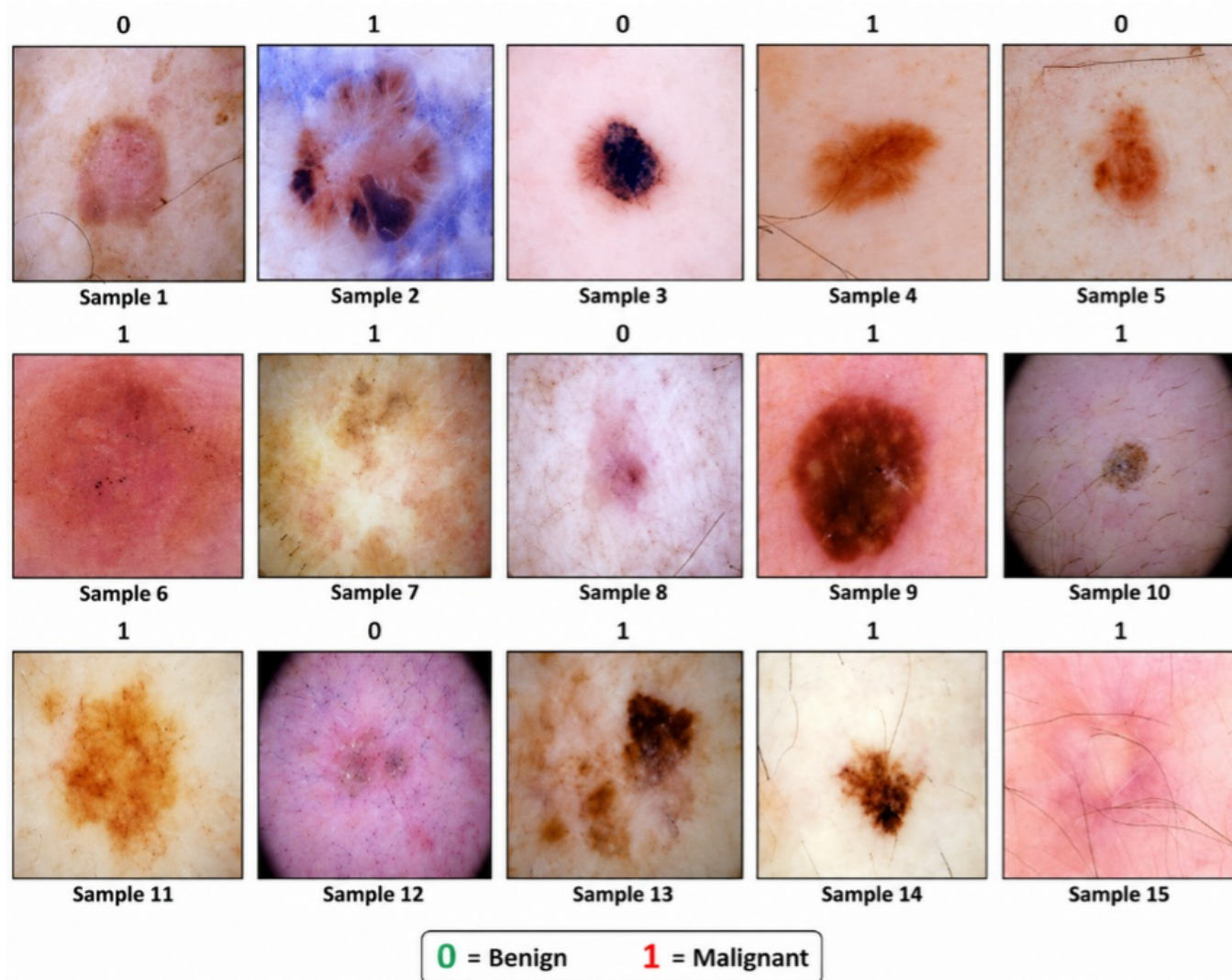
Images of skin lesions were used as inputs for the CNN, DenseNet169, and DenseNet201 models. The images were read using the OpenCV (cv2) library at a size of 250×250 pixels to speed up processing and reduce memory consumption. Subsequently, the images were labeled, and a label binarizer was used to convert the multiclass labels to binary labels. The dataset was randomly partitioned into training (80%) and validation (20%) sets. Each model was trained for 20, 60, and 100 epochs. The CNN model incorporated 3 Conv2D layers with a kernel size of 3×3 and stride of 2×2. The adaptive moment estimation (Adam) and root mean square propagation (RMSprop) optimizers were used to optimize training conditions for the models. Accuracy and loss over the training and validation data were recorded. Loss

was calculated as the difference in the actual and predicted class values.

Dataset

The dataset used for classification was obtained from Kaggle. The classification models included a CNN and 2 DenseNet architectures (DenseNet169 and DenseNet201). The dataset included images depicting several forms of skin lesions: actinic keratosis, basal cell carcinoma, dermatofibroma, melanoma, nevus, pigmented benign keratosis, seborrheic keratosis, squamous cell carcinoma, and vascular lesions.

We evaluated dataset versions 1 and 2. Both datasets contained images illustrating the same conditions. However, dataset 2 included additional images in each category and exhibited a more balanced distribution of data across conditions. Dataset 2 was used after dataset 1 to help mitigate model overfitting. Several images from both datasets are shown in Figure 1.

Figure 1. Images from the datasets.

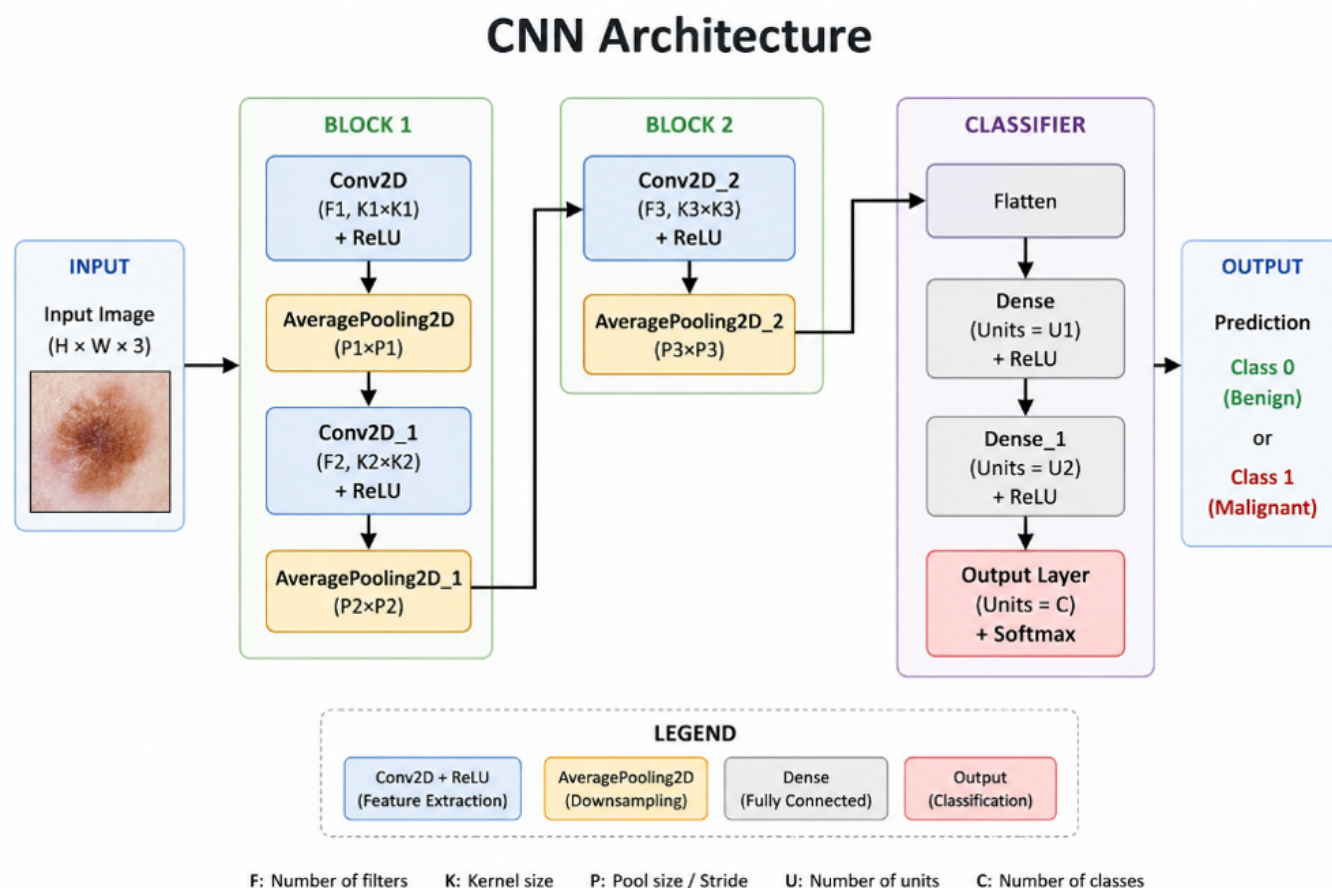
Model Architecture

CNN

The CNN architecture is shown in Figure 2. In general, CNNs are composed of convolution layers, pooling layers, and fully connected layers. The workflow involves input, feature extraction, and classification. Feature extraction is performed using convolutional layers with rectified linear unit (ReLU) activation and pooling layers. Classification is performed using a fully connected layer with softmax activation. The first convolution layer processes the input image by convoluting it through a filter that extracts image features, thereby creating a feature map. Notably, convolution layers process only the data within their receptive fields before forwarding the combined inputs to the subsequent layers. CNN inputs

are represented as tensors with specific dimensions; once these tensors pass through the convolutional layer, they are transformed to a feature map. ReLU activation helps reduce errors by improving learning efficiency. Pooling layers help mitigate overfitting by decreasing the spatial dimensions of the feature map [22,23]. Specifically, pooling layers merge the outputs of neuron clusters in one layer into a single neuron in the following layer. The CNN used in this study incorporated average pooling, a technique often used to identify minor changes in images. The fully connected layer lies at the end of the network and connects all neurons of the prior activation layers. Using a flattening process, the resulting maps can be transformed into one-dimensional data [24]. For classification, the flattened matrix is passed to a fully connected layer.

Figure 2. Architecture of the convolutional neural network. CNN: convolutional neural network; ReLU: rectified linear unit.

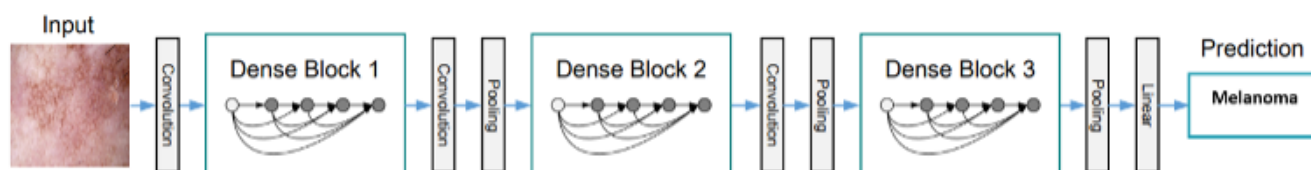


DenseNet

The general architecture of DenseNet is shown in Figure 3. DenseNets are a class of CNNs featuring dense connections between layers with similar feature-map sizes. Specifically, a feedforward connection exists between a given layer and the remaining layers constituting the network. Unlike other convolutional networks that feature a connection only

between a given layer and its following layer, DenseNets connect each layer to its previous and following layers [25]. While a standard convolutional network with L layers typically has L connections, a DenseNet with L layers has $L(L + 1)/2$ direct connections. Each layer in a dense block receives feature maps from all preceding layers and provides its output to all subsequent layers.

Figure 3. General architecture of DenseNet.

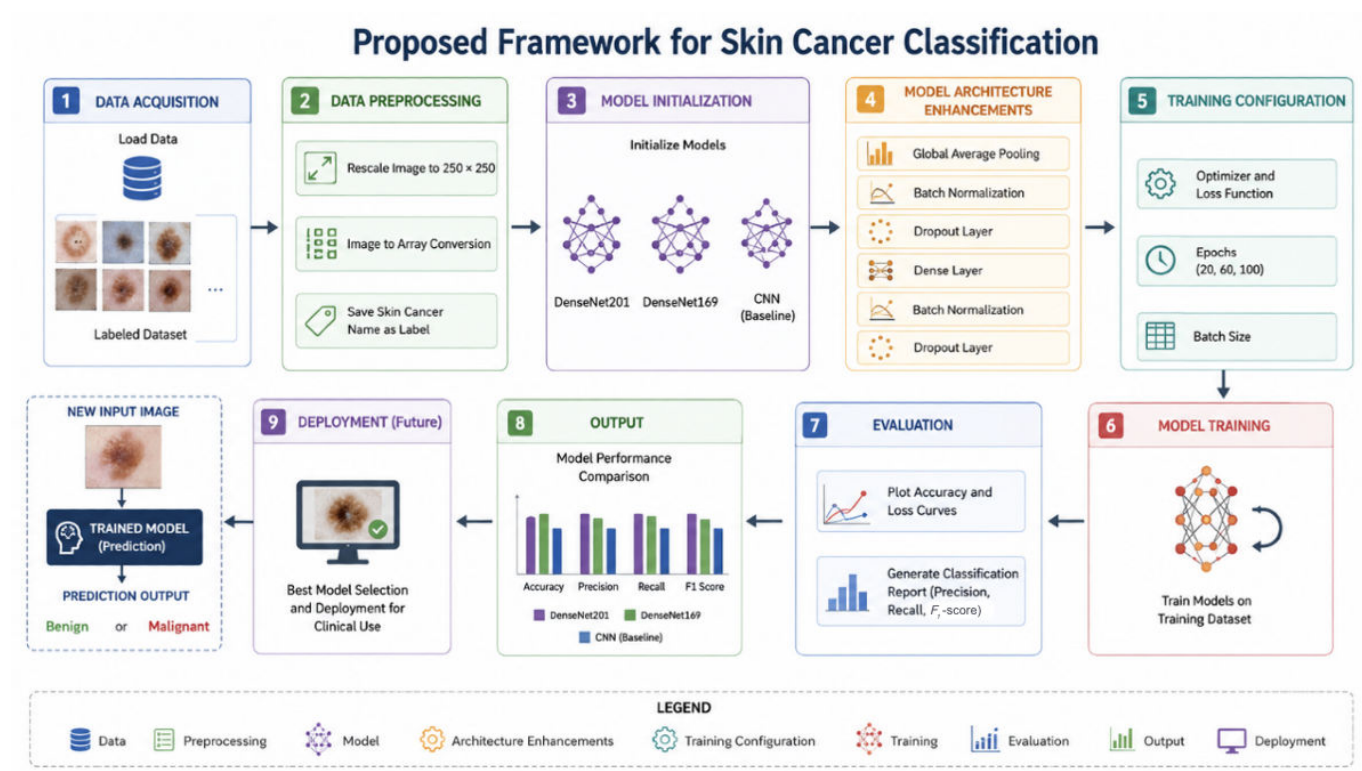


In DenseNets, feature maps obtained from the previous layers are combined through concatenation rather than summing. This architecture offers several advantages, including mitigation of the vanishing gradient problem commonly encountered in CNN frameworks, enhanced strength of connections, more efficient propagation of information, feature reuse, and reduced parameter requirements [26]. The ability of DenseNet to efficiently propagate information

facilitates training through implicit deep supervision [27]. DenseNet169 and DenseNet201 include 4 blocks with 6, 12, 32, and 32 layers and 6, 12, 48, and 32 layers, respectively.

Proposed Framework

Figure 4 illustrates the framework used to detect skin lesions in this work.

Figure 4. Flowchart of the proposed framework for skin lesion classification. CNN: convolutional neural network.

Ethical Considerations

This study involved secondary analysis of a publicly available dermoscopic image dataset obtained from Kaggle. No participants were prospectively recruited, and no new patient data or biological specimens were collected. The investigators did not have access to direct patient identifiers or identifiable private information in the dataset used for the analysis. The data were analyzed solely for research purposes in accordance with the terms and conditions of the data source. Because the study was based exclusively on publicly available, deidentified secondary data and involved no direct interaction with or intervention in human participants, informed consent was not obtained for this study. No compensation was provided because no participants were recruited or directly involved.

Results

Overview

The models were trained for 20, 60, and 100 epochs. Loss was calculated by subtracting the predicted class values from the actual class values.

The CNN model was first evaluated using the Adam optimizer. At 20 and 60 epochs, training and validation loss decreased gradually, with negligible differences between the 2 values. This indicates that the model did not exhibit overfitting or underfitting at these durations. However, at 100 epochs, training and validation loss exhibited different trends. The validation loss initially decreased but began to increase at 40 epochs. In contrast, the training loss continued to decrease even after 40 epochs. The difference between training and validation loss at 100 epochs indicated

overfitting. Similarly, during training for 100 epochs, the training accuracy continued to increase, whereas validation accuracy remained constant at 50%. The F_1 -scores on the validation set were 40%, 50%, and 60% at 20, 60, and 100 epochs, respectively. The F_1 -score over the complete dataset increased with training duration from 20 to 100 epochs.

The optimizers assessed in this work were RMSprop and Adam. These optimizers affect model performance by influencing the training process. Both optimizers use moving averages of gradients. Although Adam is similar to RMSprop, it integrates momentum-based quantities. Adam has been reported to be more efficient than other optimizers owing to its reduced memory and time consumption, as well as more effective management of scattered gradients in scenarios involving noise [24].

Experiments were conducted using 3 neural network architectures: a conventional CNN, DenseNet169, and DenseNet201. The models were assessed using 2 optimizers (Adam and RMSprop) and over two dataset versions: the original dataset containing 2357 images (version 1) and an augmented dataset containing 6857 images (version 2). Models were trained for 15 to 100 epochs using GPU acceleration. The runtime per epoch ranged from approximately 33 to 64 seconds.

Model Performance Without Optimizers

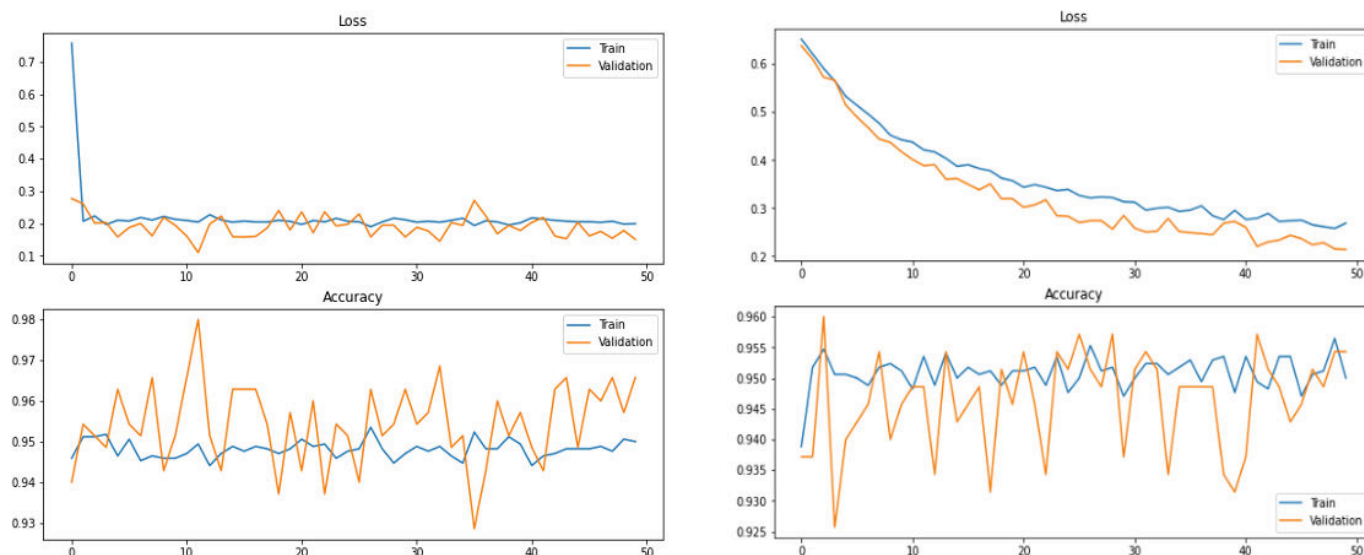
DenseNet169

The model was trained for 50 epochs. The first epoch took 59 seconds, while the remaining epochs took 33–39 seconds with GPU acceleration. At the 50th epoch, the model demonstrated strong performance, with a training accuracy

of 95.07% and validation accuracy of 94.60%, along with low loss values (training loss of 0.2014 and validation loss of 0.2101), indicating stable learning. The model achieved a high binary accuracy of 0.9616, along with perfect recall (1.0)

and strong precision (0.962). These results highlight that the model identified all true positives with few false positives. Figure 5A shows the training and validation accuracy and loss by epoch for DenseNet169.

Figure 5. Training and validation accuracy and loss by epoch: (A) DenseNet169, (B) DenseNet201.



DenseNet201

The model was trained for 50 epochs. The first epoch took 64 seconds, while the subsequent epochs took 35 seconds with GPU acceleration. At the 50th epoch, the model performed well, achieving a training accuracy of 95% and validation accuracy of 95.43% with low loss values, indicating stable learning. Additionally, the model achieved a binary accuracy of 94.86%, with perfect recall (1.0) and high precision (0.953), indicating it correctly identified all true positives with a low rate of false positives. Figure 5B shows the training and validation accuracy and loss by epoch for DenseNet201.

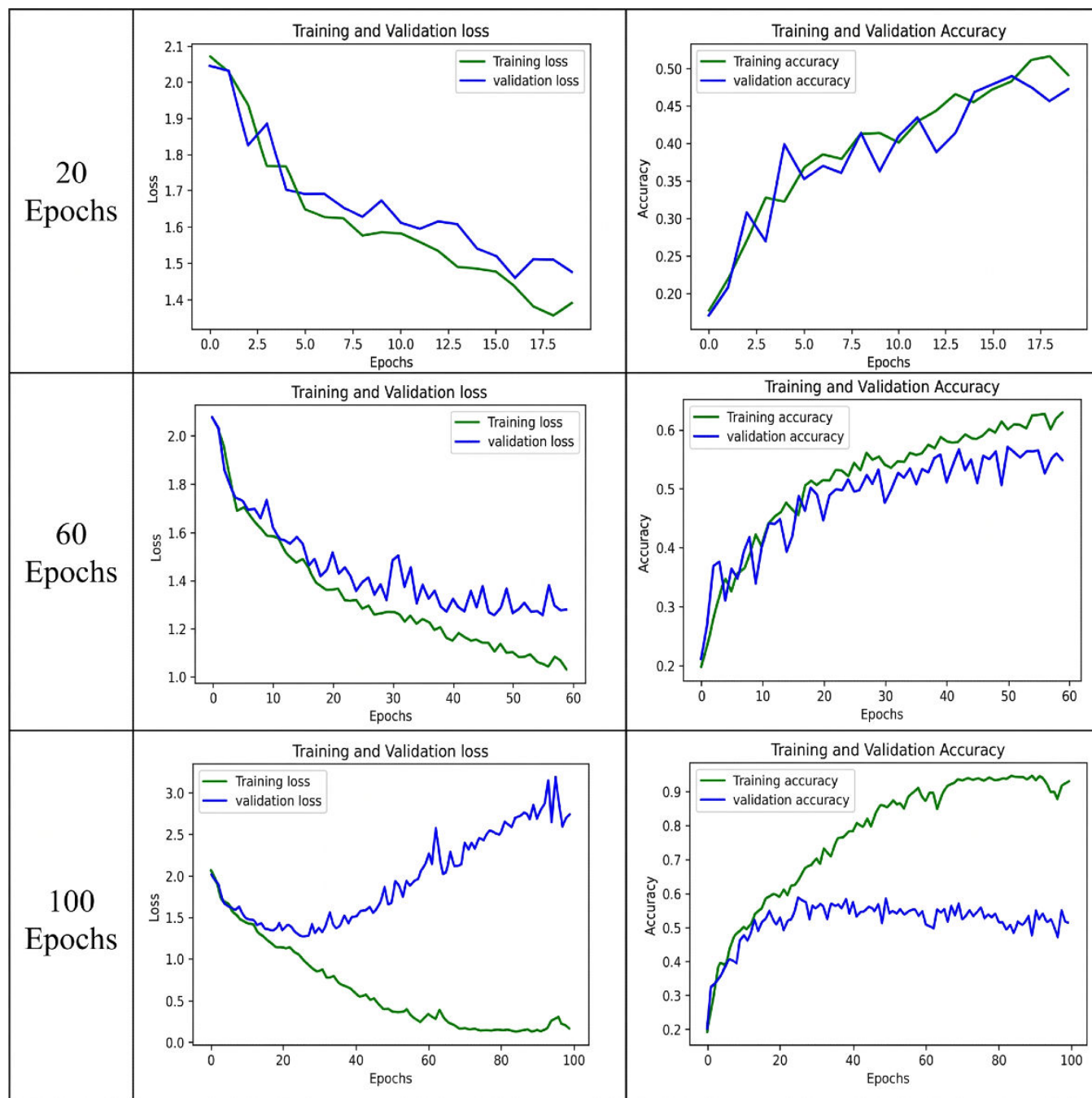
Model Performance With Optimizers

The CNN model showed improvements across epochs and optimizers, as discussed below. Adam is an optimization algorithm that can be used as an alternative to classical stochastic gradient descent, iteratively updating network weights based on training data. With the Adam optimizer, validation accuracy on the augmented dataset increased with the training duration, reaching the maximum value

of approximately 85% after 100 epochs. RMSprop yielded comparable accuracy but exhibited fluctuations in loss and validation accuracy at higher epochs, suggesting overfitting.

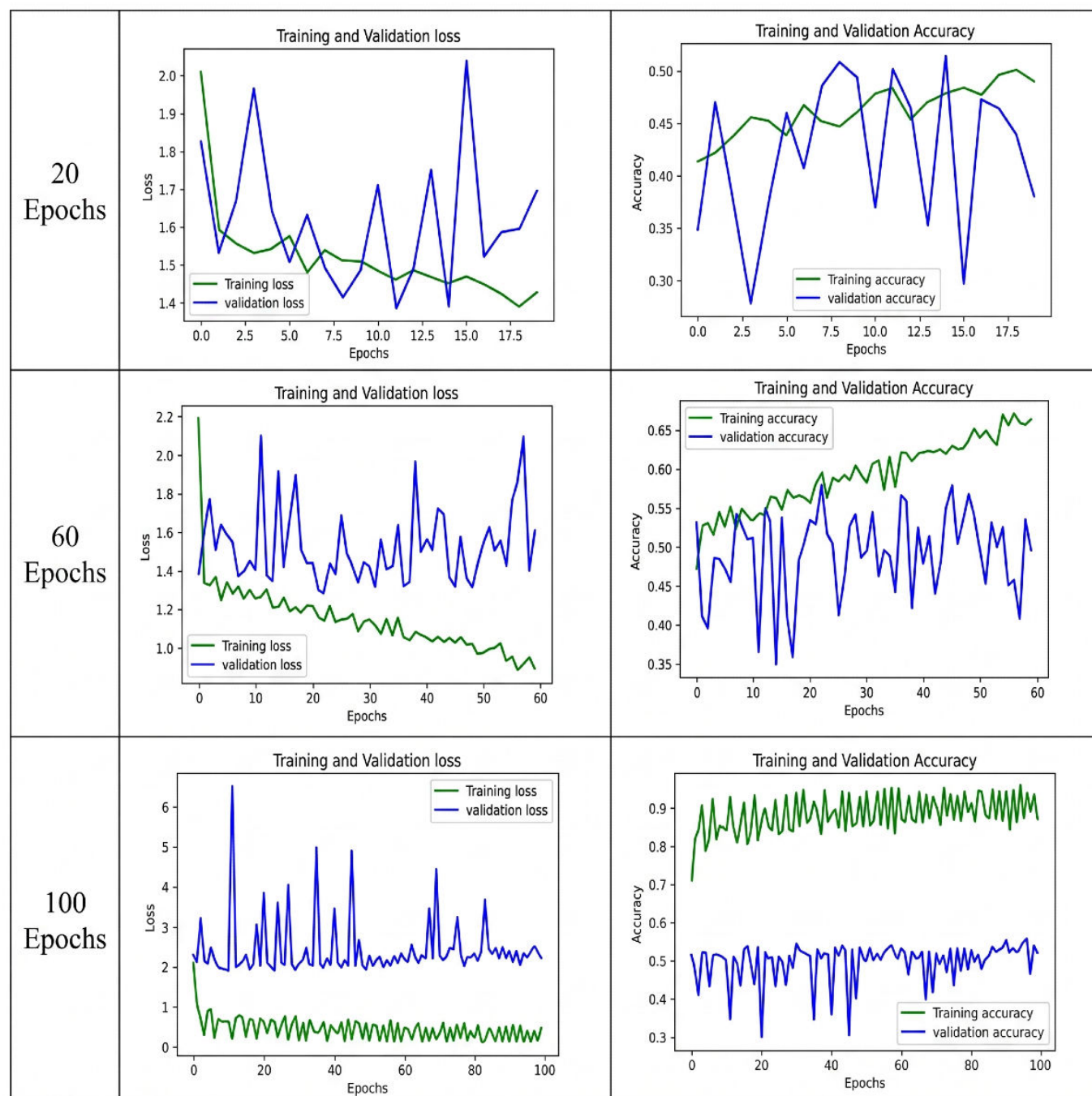
Comparison of Results Using RMSprop and Adam

Figure 6 shows the training and validation loss and accuracy of the model at 20, 60, and 100 epochs using the Adam optimizer. At 20 and 60 epochs, training and validation loss gradually decreased, with minimal difference between them. These results indicate that the model did not exhibit overfitting or underfitting. However, at 100 epochs, the validation loss initially decreased but began to increase at approximately 40 epochs. In contrast, training loss continued to decrease after 40 epochs. A similar trend was observed in training and validation accuracy. Training accuracy continuously increased with increasing epoch count, but validation accuracy remained approximately 50%. This deviation indicates overfitting. Figure S1 in Multimedia Appendix 1 shows the confusion matrix for the Adam optimizer.

Figure 6. Training and Validation Loss and Accuracy curves for the Adam optimizer. Adam: adaptive moment estimation.

The RMSprop optimizer can be considered similar to the gradient descent algorithm with momentum. This optimizer restricts oscillations in the vertical direction. Therefore, by increasing the learning rate, the model can take larger steps in

the horizontal direction, resulting in accelerated convergence. Figure 7 shows the training and validation loss and accuracy of the model at 20, 60, and 100 epochs using the RMSprop Optimizer.

Figure 7. Training and validation loss and accuracy curves for the RMSprop optimizer. RMSprop: root mean square propagation.

The training and validation loss and accuracy curves indicate limited learning efficiency of the model, with sharp fluctuations in both loss and accuracy suggesting overfitting. Figure S2 in [Multimedia Appendix 1](#) shows the confusion matrix for the RMSprop optimizer.

For the complete dataset, the F_1 -scores were 31%, 59%, and 87% at 20, 60, and 100 epochs, respectively. Even though the F_1 -score was 87% after 100 epochs, the model performance was expected to be unsatisfactory on unseen data (data not in the training and validation sets).

The results indicate that the model overfit after 100 epochs. Lack of sufficient data may contribute to overfitting. Both the Adam and RMSprop optimizers displayed similar

F_1 -scores after 100 epochs. Although both optimizers exhibit momentum-like behavior (Adam through its bias-corrected moving average and RMSprop with momentum by explicitly adding a fraction of the previous step), the mechanisms are different, which may influence the hyperparameter learning rate.

To enhance model performance, data augmentation was performed using shear rotations of $\pm 10^\circ$. Although this transformation led to only minor changes in the existing data, the model treated them as newly input images. Approximately 500 images were added per class to ensure a balanced class distribution.

The CNN was then trained using the augmented dataset. Figures 8 and 9 show the loss and accuracy associated with the augmented training and validation data at 20, 60, and 100 epochs using the Adam and RMSprop optimizers,

respectively. Figures S3 and S4 in Multimedia Appendix 1 present the confusion matrices for the augmented validation data using the Adam and RMSprop optimizers, respectively.

Figure 8. Training and validation loss and accuracy curves for the augmented dataset using the Adam optimizer at 20, 60, and 100 epochs. Adam: adaptive moment estimation.

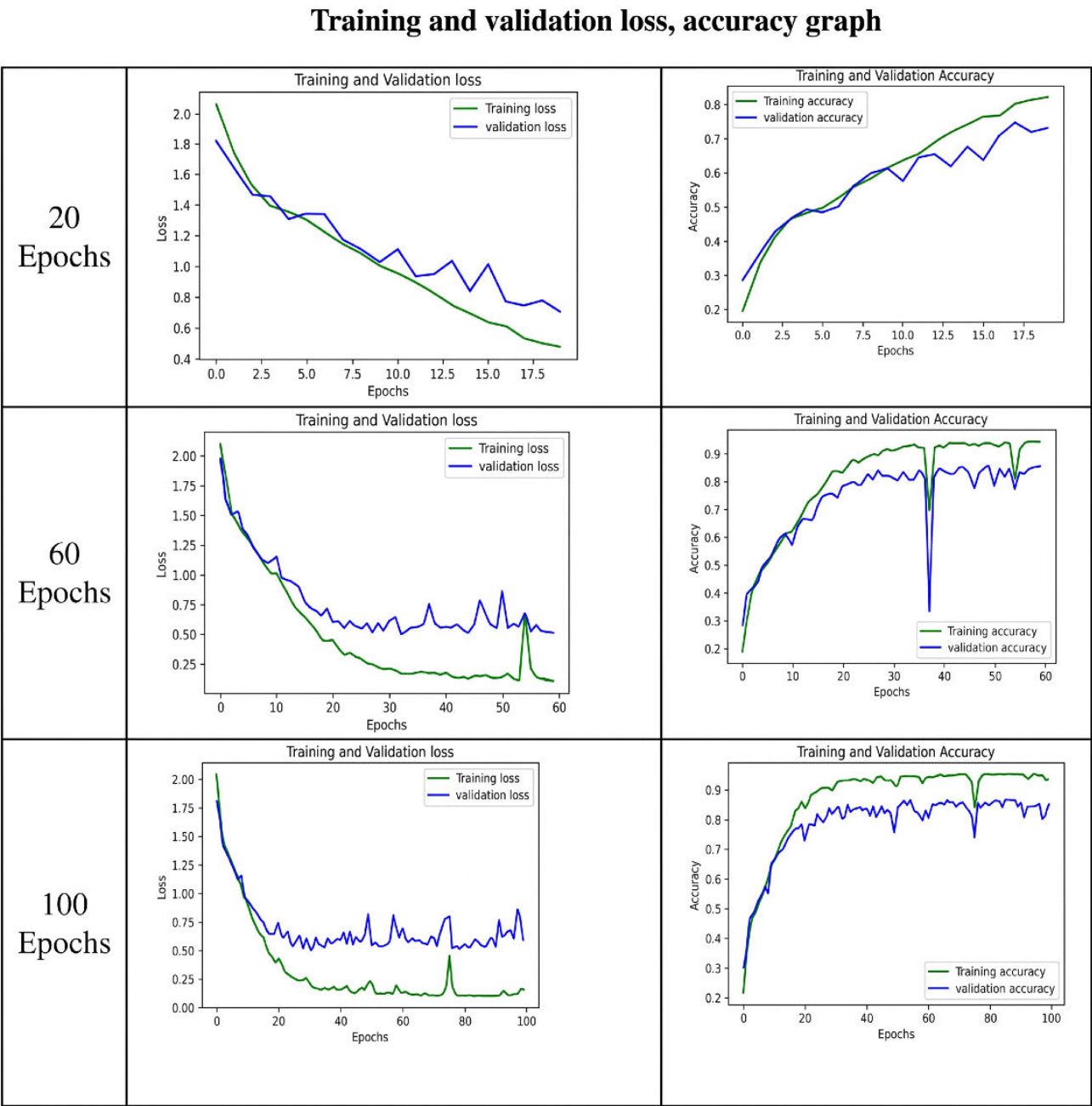
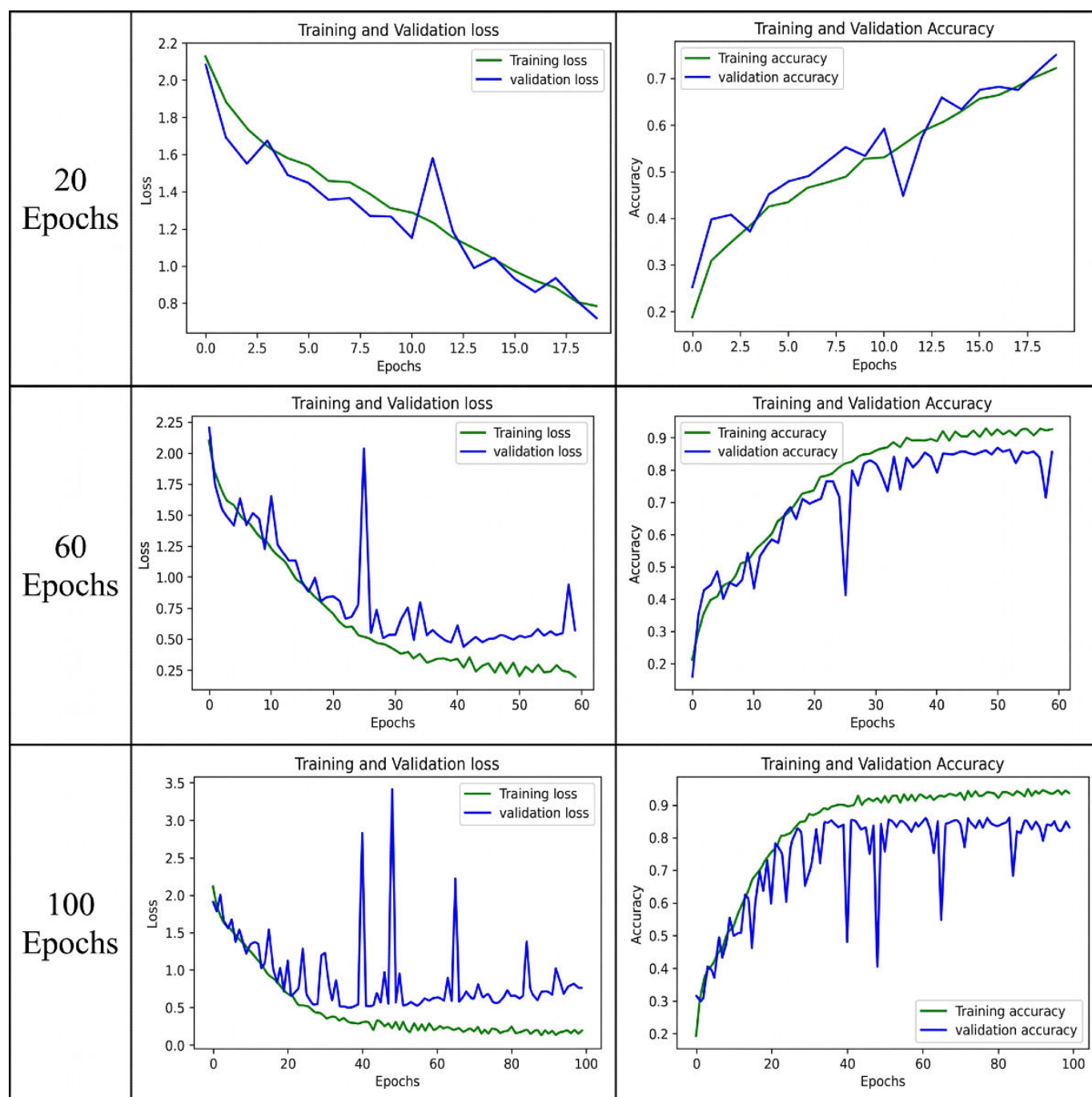


Figure 9. Training and validation loss and accuracy curves for the augmented dataset using the RMSprop optimizer at 20, 60, and 100 epochs. RMSprop: root mean square propagation.



As shown in Figure 8, training and validation loss decreased compared with the results obtained using the original dataset. At 100 epochs, the training and validation accuracy exceeded 90% and 85%, respectively. Although both loss and accuracy improved with the RMSprop optimizer as well (Figure 9), sudden increases in loss and corresponding decreases in accuracy were noted.

DenseNet169

Figures 10 and 11 and Figures S5 and S6 in Multimedia Appendix 1 summarize the performance of DenseNet169 in

terms of accuracy, stability, and generalizability. Figures 10 and 11 present the training and validation loss and accuracy curves using the Adam and RMSprop optimizers, respectively. Figures S5 and S6 in Multimedia Appendix 1 present the confusion matrices for the validation set using the Adam and RMSprop optimizers, respectively.

Figure 10. Training and validation loss and accuracy curves for DenseNet169 using the Adam optimizer at 20, 60, and 100 epochs. Adam: adaptive moment estimation.

Training and validation loss, accuracy graph

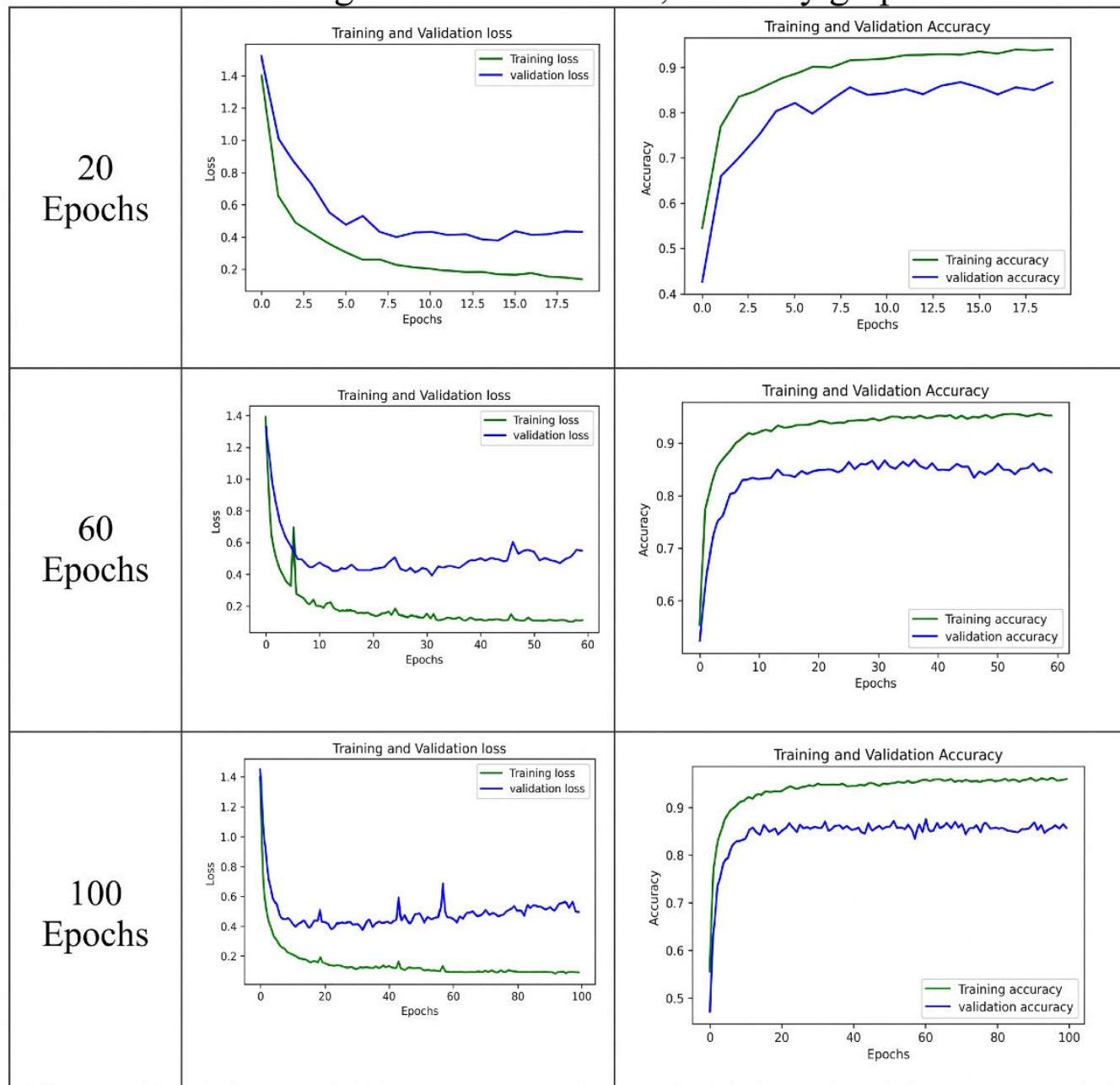
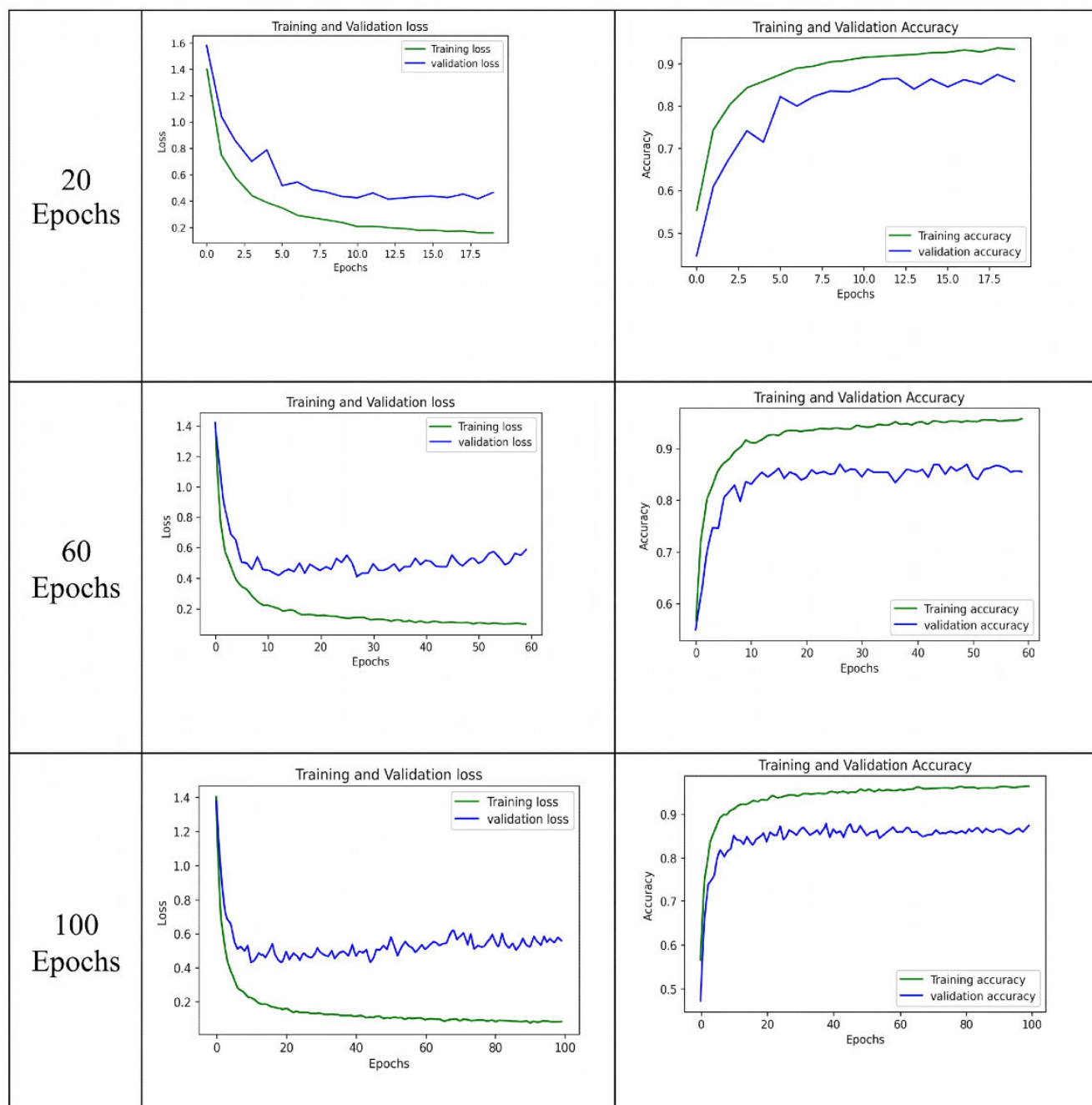


Figure 11. Training and validation loss and accuracy curves for DenseNet169 using the RMSprop optimizer at 20, 60, and 100 epochs. RMSprop: root mean square propagation.



With increase in the dataset size, DenseNet169 achieved higher accuracy and smaller loss than those for the CNN architecture at 20, 40, and 60 epochs. Validation accuracy was 85%

DenseNet201

Figures 12 and 13 and Figures S7 and S8 in Multimedia Appendix 1 summarize the performance of DenseNet201 in

terms of accuracy, stability, and generalizability across the two optimization strategies. Figures 12 and 13 present the training and validation loss and accuracy curves using the Adam and RMSprop optimizers, respectively. Figures S7 and S8 in Multimedia Appendix 1 present the confusion matrices for the validation set using the Adam and RMSprop optimizers, respectively.

Figure 12. Training and validation loss and accuracy curves for DenseNet201 using the Adam optimizer at 20, 60, and 100 epochs. Adam: adaptive moment estimation.

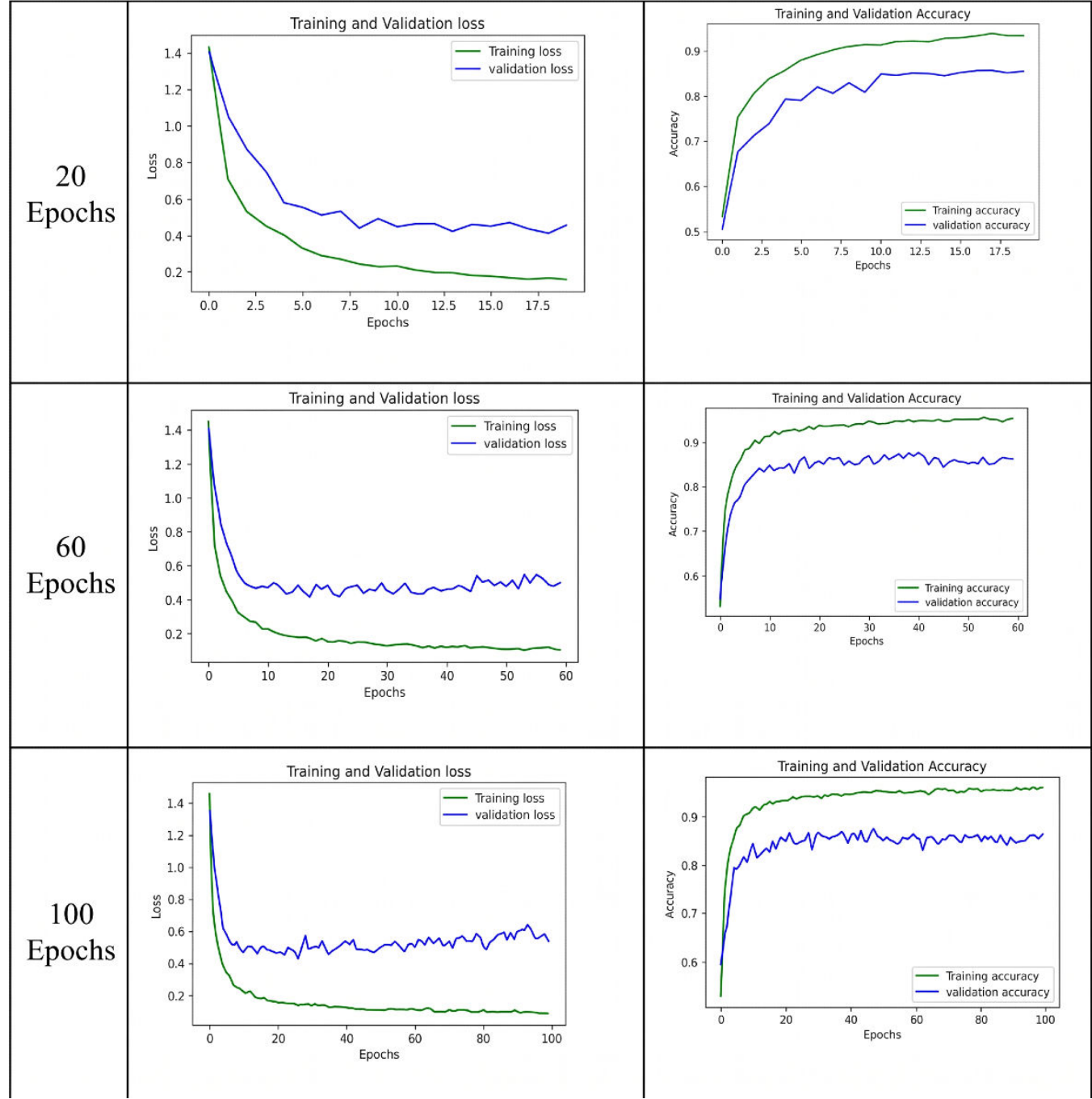
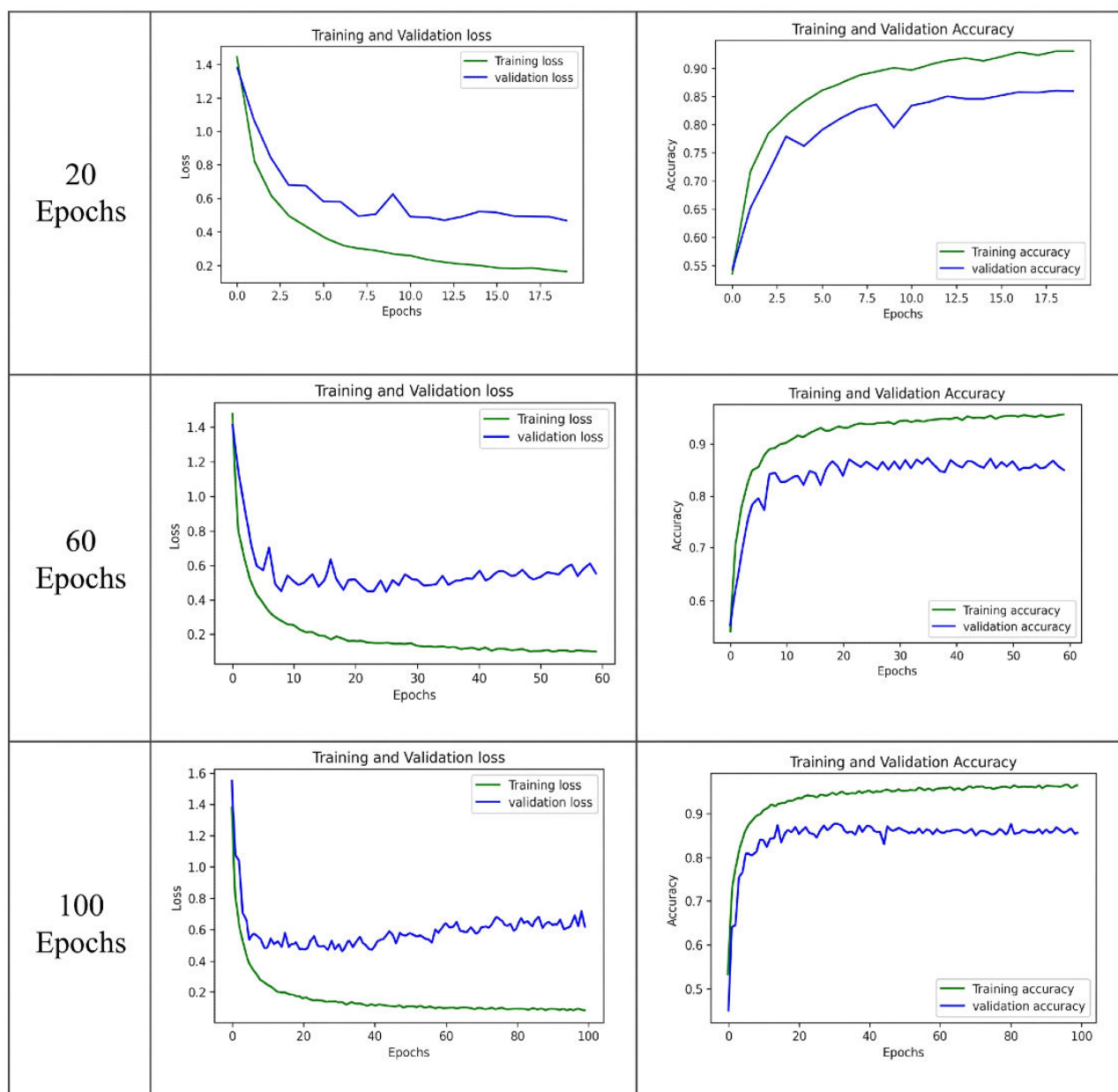


Figure 13. Training and validation loss and accuracy curves for DenseNet201 using the RMSprop optimizer at 20, 60, and 100 epochs. RMSprop: root mean square propagation.

Training and validation loss, accuracy graph



Comparative Analysis

Both DenseNet variants consistently outperformed the CNN baseline. On the augmented data, both models achieved accuracy and F_1 -scores exceeding 94% after 100 epochs, with only minor differences between the Adam and RMSprop optimizers. DenseNet169 achieved a validation accuracy of up to 95% and demonstrated superior generalization benefits compared with both CNN and DenseNet201, attributable to its dense connectivity pattern.

Augmentation involved generating approximately 500 synthetic samples per class using image transformations such as $\pm 10^\circ$ shear, resulting in improved class balance and model generalizability. The augmented dataset (version 2) was associated with increased accuracy and more stable validation loss curves, suggesting potential reduction in overfitting. Figure 14 shows heatmaps of skin lesion images from the dataset.

Figure 14. Heatmaps of skin lesion images from the dataset.

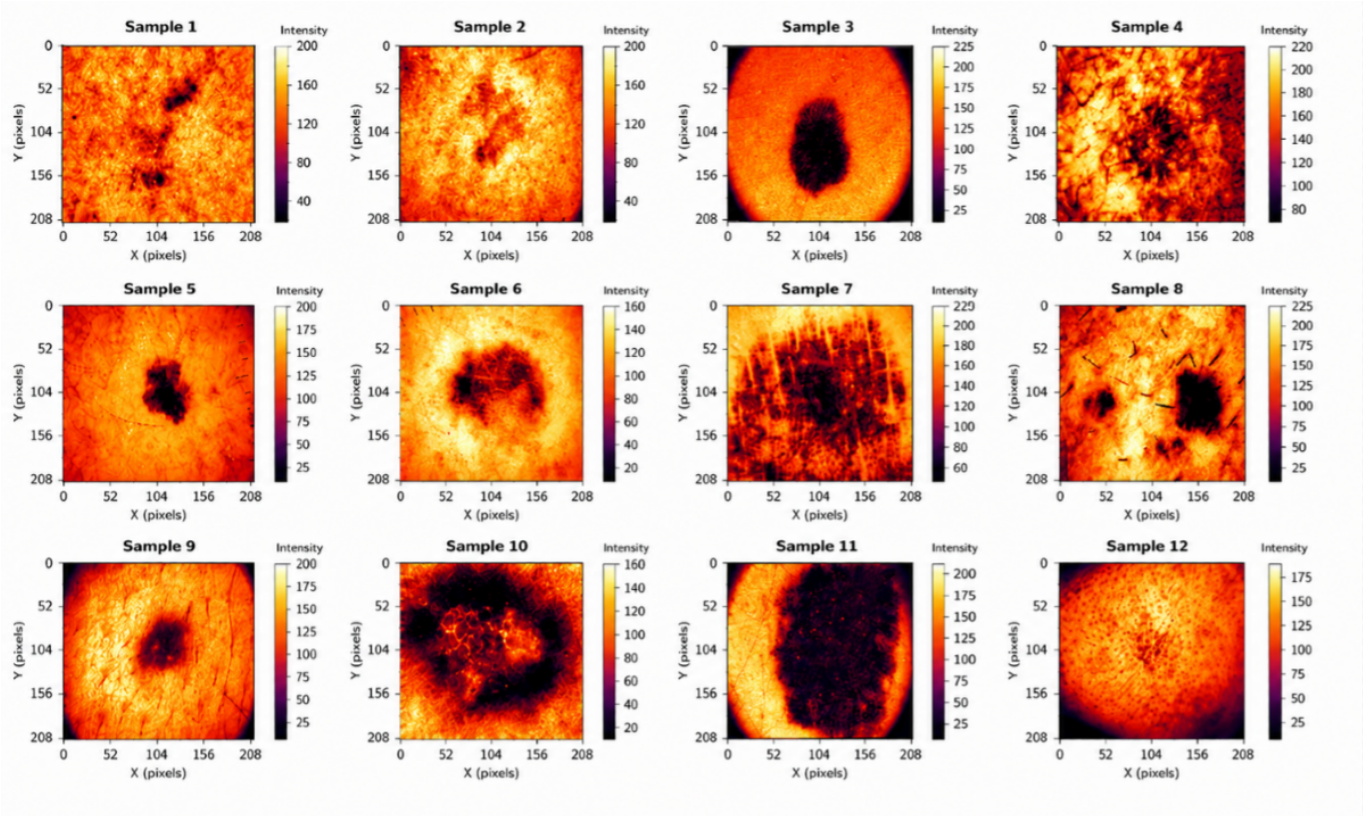


Table 2 presents the accuracy values of the CNN, DenseNet169, and DenseNet201 models. Data were obtained from Kaggle, and the models were also tested on an augmented dataset containing additional samples for each class. The dataset was divided using an 80:20 split. The validation set corresponded to the 20% portion, and the “whole dataset” included both the training and validation sets (80%+20%).

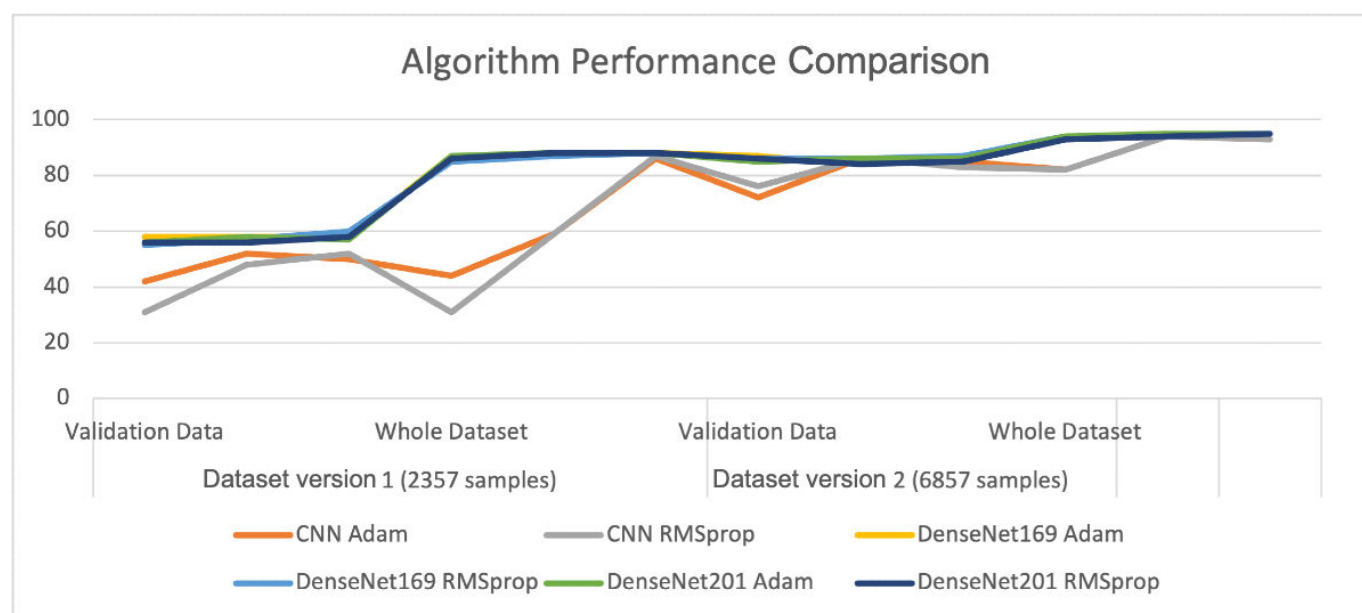
Table 2. Overall performance (accuracy, %) of the algorithms.

Model and optimizer	Dataset (2357 samples)						Dataset (6857 samples)					
	Validation data			Whole dataset			Validation data			Whole dataset		
	20 epochs	60 epochs	100 epochs	20 epochs	60 epochs	100 epochs	20 epochs	60 epochs	100 epochs	20 epochs	60 epochs	100 epochs
CNN ^a												
Adam ^b	42	52	50	44	59	86	72	86	85	82	94	93
RMSprop ^c	31	48	52	31	59	87	76	86	83	82	94	93
DenseNet169												
Adam	58	58	58	87	88	88	87	84	86	94	94	95
RMSprop	55	57	60	85	87	88	86	86	87	94	95	95
DenseNet201												
Adam	56	58	57	87	88	88	85	86	86	94	95	95
RMSprop	56	56	58	86	88	88	86	84	85	93	94	95

^aCNN: convolutional neural network.
^bAdam: adaptive moment estimation.
^cRMSprop: root mean square propagation.

Figure 15 compares the accuracy of the three studied models, highlighting the superior accuracy of DenseNet169 and DenseNet201 compared with the conventional CNN architecture.

Figure 15. Algorithm performance comparison. Adam: adaptive moment estimation; CNN: convolutional neural network; RMSprop: root mean square propagation.



Discussion

Implications

AI is increasingly transforming medicine by enhancing diagnostic accuracy, enabling personalized treatment strategies, and improving overall patient outcomes. Its ability to rapidly process large datasets, identify complex patterns, and support clinical decision-making is reshaping health care delivery. Notably, AI systems can improve diagnostic reliability, reduce human error, and streamline clinical workflows across medical specialties.

In dermatology, AI has garnered considerable attention for its potential to improve the early detection of skin cancer, particularly melanoma. Advanced deep learning models, including CNNs and DenseNet architectures, have demonstrated performance comparable to, and in some cases, exceeding that of dermatologists in the analysis of dermoscopic and clinical images. These approaches can help address inherent challenges, such as subjective interpretation and interobserver variability, offering scalable and consistent solutions for large-scale skin cancer screening.

The integration of CNN and DenseNet architecture into a unified diagnostic framework is especially valuable in clinical practice. When embedded within dermatology workflows as a decision support tool, such systems can aid physicians by automating lesion analysis. Such automation could help reduce diagnostic uncertainty, enhance clinician confidence, and facilitate the efficient classification of multiple lesion types, enabling prioritization of high-risk cases and timely decision-making.

The proposed automated framework incorporates advanced image preprocessing and augmentation techniques to improve model robustness and generalizability. By minimizing reliance on subjective visual assessment and complementing

traditional dermoscopy, the system can support earlier and more consistent melanoma detection. Dense connectivity facilitates the extraction of subtle yet clinically relevant features, improving the discrimination of benign and malignant lesions.

AI-assisted analysis can also contribute to improved clinical efficiency. Notably, rapid and accurate lesion classification can accelerate diagnostic workflows, reduce the time spent on image interpretation, and shorten patient visits, potentially yielding an increase in patient throughput while maintaining diagnostic quality. Additionally, AI-driven triage may allow trained technical staff to assist with image acquisition and preliminary screening, optimizing resource utilization within clinical settings.

From an economic perspective, AI implementation in dermatology may have the potential to reduce health care costs. Early and accurate detection of melanoma could decrease unnecessary biopsies and limit the need for costly treatments associated with advanced-stage disease. By identifying high-risk lesions earlier, the proposed framework could support more effective resource allocation, ultimately improving both economic aspects and patient outcomes.

Moreover, automation of routine image analysis may allow clinical staff to devote more time to patient coordination and care delivery. This redistribution of responsibilities would enhance operational efficiency and may help reduce provider burnout.

The compatibility of the proposed framework with dermoscopic imaging may support the use of AI in tele-dermatology. Patients could remotely submit lesion images for preliminary assessment, enabling prioritization of urgent cases and enhancing access to care in underserved or remote regions. This capability could strengthen continuity of care through remote monitoring and timely intervention.

Future work could be focused on improving model interpretability and performance through the integration of attention mechanisms into DenseNet architectures. These mechanisms could allow the model to highlight salient image regions, enhancing both diagnostic accuracy and transparency, which may increase clinician trust. Additionally, incorporating multimodal data—such as patient demographics and clinical history—alongside imaging features may further improve diagnostic precision. Validation across diverse populations and real-world clinical settings will be essential to ensure generalizability and support widespread clinical adoption.

Although the findings of this study are encouraging and suggest the potential clinical utility of the proposed framework, further validation is necessary before routine clinical implementation. Specifically, prospective evaluation across multiple institutions using larger and independent patient cohorts will be essential to assess generalizability across patient populations, imaging systems, clinical environments, and real-world workflows. Such multicenter validation may also help identify potential sources of variability and bias that may not be fully captured within the current dataset. Future studies should therefore prioritize prospective external validation to establish the robustness, reproducibility, and clinical reliability of the framework before its integration into routine clinical practice.

Clinical Translation and Future Deployment

Although the proposed CNN and DenseNet architectures demonstrated strong diagnostic performance on the evaluated dataset, clinical deployment requires validation beyond retrospective benchmarking. Before implementation in routine dermatology practice, the models should undergo prospective, multicenter external validation using images acquired from diverse populations, imaging devices, and clinical environments to assess generalizability. Integration into clinical workflows will also require consideration of

computational efficiency, interoperability with electronic health record systems, regulatory approval, data privacy, and clinician acceptance. Continuous performance monitoring and periodic model retraining will be essential to mitigate performance degradation resulting from distribution shifts in real-world data. Future work should therefore focus on external validation, explainability, uncertainty estimation, and workflow integration to facilitate the safe and effective adoption of these models as clinical decision support systems.

Conclusion

This study presented a comprehensive comparative evaluation of CNN and DenseNet-based deep learning architectures for automated classification of melanoma using dermoscopic skin lesion images. In experiments conducted under multiple training configurations, the DenseNet models consistently demonstrated superior classification performance compared with the baseline CNN architecture, with higher accuracy, sensitivity, specificity, precision, and F_1 -score. These results highlight that dense feature propagation and feature reuse may benefit the identification of malignant skin lesions while reducing false positive and false negative predictions. The presented comparison of commonly used deep learning architectures can guide researchers and clinicians in selecting appropriate models for computer-aided melanoma diagnosis. The proposed framework can serve as a clinical decision support tool, assisting dermatologists in the early detection of melanoma, thereby facilitating timely intervention and improving patient outcomes. Although further prospective, multicenter external validation is required before routine clinical implementation, the findings demonstrate the potential of deep learning-based image analysis in enhancing diagnostic accuracy and supporting precision dermatology. Overall, this work contributes to the growing body of evidence supporting the integration of AI into dermatological practice and establishes a strong foundation for future research on robust, interpretable, and clinically deployable skin cancer screening systems.

Generative AI Use

The authors declare that no generative AI tools were used in the preparation, writing, and revision of this manuscript.

Funding

The authors declared no financial support was received for this work.

Data Availability

The data analyzed in this study were obtained from a publicly available dermoscopic skin lesion dataset hosted on Kaggle [28]. The dataset contains images representing multiple categories of benign and malignant skin lesions and was used in accordance with the terms and conditions of the source repository. No new patient-level clinical data were collected specifically for this study.

Authors' Contributions

Conceptualization: HS, TM

Data curation: AA, IA, MP, RP

Formal analysis: AA, IA, MP, SW, RP

Investigation: AA, IA, RB, RP

Methodology: HS, MP, SW, TM

Project administration: HS

Resources: RB

Software: AA, MP
 Supervision: HS, TM
 Validation: HS, MP, SW, RB, TM
 Visualization: AA, IA, SW, RP
 Writing – original draft: HS, AA, IA
 Writing – review & editing: HS, MP, SW, RB, RP, TM
 All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest:

None declared.

Multimedia Appendix 1

Supplemental figures S1 to S8, showing confusion matrices for model validation performance using adaptive moment estimation (Adam) and root mean square propagation (RMSprop) optimizers.

[\[DOCX File \(Microsoft Word File\), 1512 KB-Multimedia Appendix 1\]](#)

References

1. Dildar M, Akram S, Irfan M, et al. Skin cancer detection: a review using deep learning techniques. *Int J Environ Res Public Health*. May 20, 2021;18(10):5479. [doi: [10.3390/ijerph18105479](#)] [Medline: [34065430](#)]
2. Linares MA, Zakaria A, Nizran P. Skin cancer. *Prim Care*. Dec 2015;42(4):645-659. [doi: [10.1016/j.pop.2015.07.006](#)] [Medline: [26612377](#)]
3. Sample A, He YY. Mechanisms and prevention of UV-induced melanoma. *Photodermatol Photoimmunol Photomed*. Jan 2018;34(1):13-24. [doi: [10.1111/php.12329](#)] [Medline: [28703311](#)]
4. Saginala K, Barsouk A, Aluru JS, Rawla P, Barsouk A. Epidemiology of melanoma. *Med Sci (Basel)*. Oct 20, 2021;9(4):63. [doi: [10.3390/medsci9040063](#)] [Medline: [34698235](#)]
5. Davis LE, Shalin SC, Tackett AJ. Current state of melanoma diagnosis and treatment. *Cancer Biol Ther*. 2019;20(11):1366-1379. [doi: [10.1080/15384047.2019.1640032](#)] [Medline: [31366280](#)]
6. Rastrelli M, Tropea S, Rossi CR, Alaibac M. Melanoma: epidemiology, risk factors, pathogenesis, diagnosis and classification. *In Vivo*. 2014;28(6):1005-1011. [Medline: [25398793](#)]
7. Dzwierzynski WW. Melanoma risk factors and prevention. *Clin Plast Surg*. Oct 2021;48(4):543-550. [doi: [10.1016/j.cps.2021.05.001](#)] [Medline: [34503715](#)]
8. D'Orazio J, Jarrett S, Amaro-Ortiz A, Scott T. UV radiation and the skin. *Int J Mol Sci*. Jun 7, 2013;14(6):12222-12248. [doi: [10.3390/ijms140612222](#)] [Medline: [23749111](#)]
9. Smaoui N, Bessassi S. A developed system for melanoma diagnosis. *Int J Comput Vis Signal Process*. 2013;3:10-17. URL: <https://cennser.org/IJCVSP/finalPaper/030102.pdf> [Accessed 2026-10-12]
10. Marchetti MA, Liopyris K, Dusza SW, et al. Computer algorithms show potential for improving dermatologists' accuracy to diagnose cutaneous melanoma: results of the International Skin Imaging Collaboration 2017. *J Am Acad Dermatol*. Mar 2020;82(3):622-627. [doi: [10.1016/j.jaad.2019.07.016](#)] [Medline: [31306724](#)]
11. Cheong KH, Tang KJW, Zhao X, et al. An automated skin melanoma detection system with melanoma-index based on entropy features. *Biocybern Biomed Eng*. Jul 2021;41(3):997-1012. [doi: [10.1016/j.bbe.2021.05.010](#)]
12. Waheed SR, Saadi SM, Mohd Rahim MS, et al. Melanoma skin cancer classification based on CNN deep learning algorithms. *Mal J Fund Appl Sci*. 2023;19(3):299-305. [doi: [10.11113/mjfas.v19n3.2900](#)]
13. Alsaade FW, Aldhyani THH, Al-Adhaileh MH. Developing a recognition system for diagnosing melanoma skin lesions using artificial intelligence algorithms. *Comput Math Methods Med*. 2021;2021:9998379. [doi: [10.1155/2021/9998379](#)] [Medline: [34055044](#)]
14. Indraswari R, Rokhana R, Herulambang W. Melanoma image classification based on MobileNetV2 network. *Procedia Comput Sci*. 2022;197:198-207. [doi: [10.1016/j.procs.2021.12.132](#)]
15. Tumpa PP, Kabir MA. An artificial neural network based detection and classification of melanoma skin cancer using hybrid texture features. *Sens Int*. 2021;2:100128. [doi: [10.1016/j.sintl.2021.100128](#)]
16. Brinker TJ, Hekler A, Enk AH, et al. A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. *Eur J Cancer*. Apr 2019;111:148-154. [doi: [10.1016/j.ejca.2019.02.005](#)] [Medline: [30852421](#)]
17. Thaajwer MAA, Ishanka UAP. Melanoma skin cancer detection using image processing and machine learning techniques. Presented at: 2020 2nd International Conference on Advancements in Computing (ICAC); Dec 10-11, 2020:363-368; Malabe, Sri Lanka. [doi: [10.1109/ICAC51239.2020.9357309](#)]
18. Banerjee S, Singh SK, Chakraborty A, Das A, Bag R. Melanoma diagnosis using deep learning and fuzzy logic. *Diagnostics (Basel)*. Aug 9, 2020;10(8):577. [doi: [10.3390/diagnostics10080577](#)] [Medline: [32784837](#)]

19. Bakheet S, El-Nagar A. A deep neural approach for real-time malignant melanoma detection. Appl Math Inf Sci. Jan 1, 2021;15(1):89-96. [doi: [10.18576/amis/150111](https://doi.org/10.18576/amis/150111)]
20. Chang CC, Li YZ, Wu HC, Tseng MH. Melanoma detection using XGB classifier combined with feature extraction and k-means SMOTE techniques. Diagnostics (Basel). Jul 19, 2022;12(7):1747. [doi: [10.3390/diagnostics12071747](https://doi.org/10.3390/diagnostics12071747)] [Medline: [35885650](https://pubmed.ncbi.nlm.nih.gov/35885650/)]
21. Thiyaneswaran B, Anguraj K, Kumarganesh S, Thangaraj K. Early detection of melanoma images using gray level co-occurrence matrix features and machine learning techniques for effective clinical diagnosis. Int J Imaging Syst Technol. Jun 2021;31(2):682-694. [doi: [10.1002/ima.22514](https://doi.org/10.1002/ima.22514)]
22. Young AT, Vora NB, Cortez J, et al. The role of technology in melanoma screening and diagnosis. Pigment Cell Melanoma Res. Mar 2021;34(2):288-300. [doi: [10.1111/pcmr.12907](https://doi.org/10.1111/pcmr.12907)] [Medline: [32558281](https://pubmed.ncbi.nlm.nih.gov/32558281/)]
23. Kim P. MATLAB Deep Learning. Apress; 2017. [doi: [10.1007/978-1-4842-2845-6](https://doi.org/10.1007/978-1-4842-2845-6)]
24. Fu'adah YN, Pratiwi NC, Pramudito MA, Ibrahim N. Convolutional neural network (CNN) for automatic skin cancer classification system. IOP Conf Ser: Mater Sci Eng. Dec 1, 2020;982(1):012005. [doi: [10.1088/1757-899X/982/1/012005](https://doi.org/10.1088/1757-899X/982/1/012005)]
25. Gajera HK, Nayak DR, Zaveri MA. A comprehensive analysis of dermoscopy images for melanoma detection via deep CNN features. Biomed Signal Process Control. Jan 2023;79:104186. [doi: [10.1016/j.bspc.2022.104186](https://doi.org/10.1016/j.bspc.2022.104186)]
26. Huang G, Liu Z, Maaten LVD, Weinberger KQ. Densely connected convolutional networks. arXiv. Preprint posted online on Jan 28, 2018. [doi: [10.48550/arXiv.1608.06993](https://doi.org/10.48550/arXiv.1608.06993)]
27. Lee CY, Xie S, Gallagher P, Zhang Z, Tu Z. Deeply-supervised nets. arXiv. Preprint posted online on Sep 25, 2014. [doi: [10.48550/arXiv.1409.5185](https://doi.org/10.48550/arXiv.1409.5185)]
28. Hussain SI. ISIC 2019 skin cancer detection dataset. Kaggle. URL: <https://www.kaggle.com/dsv/16086084> [Accessed 2026-10-26]

Abbreviations

ABCDE: asymmetry, border irregularity, color heterogeneity, diameter, and evolution

Adam: adaptive moment estimation

CNN: convolutional neural network

DenseNet: densely connected neural network

NMSC: nonmelanoma skin cancer

ReLu: rectified linear unit

RMSprop: root mean square propagation

Edited by Robert Dellavalle; peer-reviewed by John Meisenheimer, Xiaolong Liang; submitted 08.Jun.2026; final revised version received 02.Sep.2026; accepted 23.Sep.2026; published 27.Oct.2026

Please cite as:

Sanghvi H, Agritis A, Alvarez I, Patel M, Wang S, Patel R, Banoub R, Makkar T

Deep Learning–Based Classification of Melanoma Using Dermoscopic Images: Comparative Evaluation of Convolutional Neural Network and DenseNet Models

JMIR Dermatol 2026;9:e104086

URL: <https://derma.jmir.org/2026/1/e104086>

doi: [10.2196/104086](https://doi.org/10.2196/104086)

© Harshal Sanghvi, Argyrios Agritis, Isabella Alvarez, Mitsu Patel, Shuwen Wang, Riki Patel, Raphael Banoub, Timothy Makkar. Originally published in JMIR Dermatology (<http://derma.jmir.org>), 27.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Dermatology, is properly cited. The complete bibliographic information, a link to the original publication on <http://derma.jmir.org>, as well as this copyright and license information must be included.