International Journal of Robotics and Control Systems
Vol. 6, No. 1, 2026, pp. 507-527
ISSN 2775-2658
http://pubs2.ascee.org/index.php/ijrcs

Deep Learning Architecture Optimization for Skin Cancer Image Classification on Multi-Source Dataset

Nurul Khasanah [a,1,*], Taopik Hidayat [a,2], Elly Firasari [a,3], Laela Kurniawati [a,4], Eni Heni Hermaliani [a,5]

[a] Department of Technology Informatic, Universitas Nusa Mandiri, Jakarta, Indonesia
[1] nurul.nuk@nusamandiri.ac.id; [2] taopik.toi@nusamandiri.ac.id; [3] elly.efa@nusamandiri.ac.id; [4] laela@nusamandiri.ac.id; [5] enie_h@nusamandiri.ac.id
*Corresponding Author

ARTICLE INFO

Article history
Received December 07, 2025
Revised January 30, 2026
Accepted February 12, 2026

Keywords
Skin Cancer;
Image Classification;
Deep Learning;
Densenet169

ABSTRACT

Early detection of skin cancer is crucial to reduce diagnostic delays and improve patient outcomes, yet existing automated systems often suffer from limited generalization due to single-source datasets and restricted model evaluation. This study presents a deep learning based skin cancer image classification system using a multi-source dataset of 13,902 dermatoscopic images categorized into benign and malignant classes. In the first phase, a benchmarking study was conducted on seven pre-trained CNN architectures, namely MobileNetV2, InceptionV3, Xception, DenseNet169, ResNet50, VGG16, and VGG19. The results indicate that DenseNet169 achieved the best baseline performance with a test accuracy of 89.30%. In the second phase, the DenseNet169 architecture was optimized through structural modification of dense layers, application of dropout regularization, and selective fine-tuning of backbone layers. The optimized model improved the test accuracy to 91.20% and achieved an AUC-ROC of 97.14%, demonstrating enhanced robustness and sensitivity in detecting malignant lesions. The novelty of this work lies in the integration of multi-source datasets combined with targeted architectural optimization of DenseNet169 to reduce false-negative rates in malignant detection. These findings highlight the potential of the proposed model as a reliable non-invasive clinical decision support tool for early skin cancer diagnosis, and emphasize the need for further validation using prospective clinical datasets and real-world deployment scenarios to ensure its practical applicability in clinical environments.

© 2025 The Authors.
Published by Association for Scientific Computing Electrical and Engineering.
This is an open-access article under the CC–BY-NC license.

1. Introduction

Skin cancer is one of the fastest growing types of cancer worldwide and ranks sixth in terms of prevalence [1]-[4]. In Indonesia, skin cancer is associated with a high mortality rate, largely due to delayed diagnosis [5], [6]. Clinically, skin cancer is classified into benign and malignant types, where malignant lesions are invasive and potentially metastatic [7]-[9]. Early detection is therefore critical, as timely medical intervention significantly improves patient survival rates [10]-[14]. However, current diagnostic practices primarily rely on visual examination and biopsy procedures, which are

invasive, time-consuming, costly, and may cause patient discomfort [15]-[21]. These limitations highlight the urgent need for faster, more accessible, and non-invasive diagnostic tools [22].

To address the limitations of conventional diagnostic approaches, deep learning has emerged as a non-invasive solution that enables faster and more objective analysis. Advances in deep learning techniques have been extensively applied in medical image analysis, demonstrating strong potential for automated and reliable interpretation [23]-[31]. One of its most prominent applications is in detecting and classifying various types of skin cancer [22]-[33]. In its application, various Convolutional Neural Network architectures have been widely applied to their ability to extract and recognize complex visual patterns in skin lesion images accurately [34]-[38]. Several previous studies have utilized models such as DenseNet169, VGG19, MobileNet, and Xception. The DenseNet169 model has been shown to achieve an accuracy of over 91% on the HAM10000 dataset [5]. Meanwhile, VGG19 shows promising results with 93% accuracy on the ISIC 2019 dataset and 92% on ISIC 2020 [39]. In another study, the MobileNet and Xception architectures managed to achieve accuracies of 96% and 97% respectively in classifying 5 types of skin diseases [40]. The results that have been achieved show the great potential of CNN architecture in skin cancer classification. Although many studies have benchmarked CNN models on the ISIC dataset, most are still limited to the use of a single data source or the evaluation of only a few architectures, resulting in a gap in generalization ability when data comes from heterogeneous sources with variations in color, lighting, and image quality, especially for applications on real-world clinical data. Accordingly, this study addresses three key research gaps: (1) dataset variability across different acquisition conditions, (2) limited multi-architecture benchmarking in prior works, and (3) insufficient exploration of architectural optimization strategies that improve diagnostic performance without substantially increasing computational cost.

To tackle these challenges, this study utilizes a multi-source dataset with 13,902 images from two public datasets originating from Kaggle, namely Skin Cancer: Malignant vs. Benign [41] and Melanoma Skin Cancer Dataset of 10,000 Images to build a two-class (malignant vs. benign) classification model [42]. This study will compare seven state-of-the-art CNN architectures initialized with ImageNet pretrained weights [43]. The best architecture is modified by layers to produce better accuracy. Each model's performance is evaluated using accuracy, recall, precision, confusion matrix analysis, F1-score and ROC Curve [2], [34].

The contributions of this study are threefold: integrating publicly available multi-source datasets to improve lesion variability and model robustness, conducting a comprehensive evaluation of seven CNN architectures, and proposing architectural modifications with hyperparameter tuning to reduce critical errors, particularly false negatives in malignant classes, while improving diagnostically relevant metrics such as AUC-ROC. Although numerous studies have benchmarked CNN models on ISIC datasets, a gap remains in addressing generalization across heterogeneous data sources. This study fills that gap through multi-source benchmarking and targeted architectural optimization.

Overall, the expected outcome of this research is a high-accuracy skin cancer classification model that is ready to be integrated into clinical workflows as a non-invasive diagnostic support tool and can serve as a foundation for further artificial intelligence-based diagnostic system development and structured clinical validation [44].

2. Related Work

Table 1 shows some previous studies that are relevant to this study. These studies focus on the development of skin cancer classification methods based on machine learning and deep learning approaches. The level of accuracy achieved in each study varies, depending on the model architecture and type of dataset used. Research [39] implemented a deep learning approach using the VGG16, VGG19 architectures, as well as a hybrid model combining VGG19 and SVM. Using the ISIC 2019 and 2020 datasets, the accuracy produced by VGG19 combined with the SVM classifier is 96%,

which confirms the effectiveness of the combined approach between deep learning and traditional classification. Research [5] proposed a DeepSkin model using the DenseNet169 and ResNet50 architectures. The evaluation results showed that DenseNet169 excelled with an accuracy of 91.20%, while ResNet50 only achieved 83%. This shows the superiority of dens connection in enhancing feature extraction in dermatoscopic images [45]. Research [46] explore the influence of Image Super Resolution (ISR) techniques on CNN model performance. The findings indicate that enhancing image resolution substantially improves model performance. Specifically, the accuracy of the InceptionV3 model increased from 83.48% to 91.26% following the application of image super-resolution (ISR).

Research [40] evaluating lightweight architectures such as MobileNet and Xception for skin disease classification. Each model achieved accuracies of 96% and 97%, respectively, indicating that even small-sized models are capable of delivering high performance with good computational efficiency. Firasari et al. [47] conducted a comparative study evaluating the performance of ResNet50 and MobileNetV2 using various optimization algorithms. Both models achieve the same accuracy of 88%. However, MobileNetV2 has a lower loss value (0.3884 compared to 0.4674), indicating a more stable training process.

Khasanah et al. [48] assessed the performance of the InceptionV3 model for binary skin cancer classification using a transfer learning strategy. Total dataset 2,637 images, with input images measuring 224×224 pixels. When training the model reached an accuracy of 86.65%, during testing the accuracy dropped to 50.30%, indicating overfitting. This behavior is further evidenced by persistently high loss values despite an increase in the number of training epochs. Precision and recall for both classes ranged from 45% to 56%, indicating that the model is not yet optimal and still requires hyperparameter tuning and expansion of training data to improve generalization. Research by Kavitha et al. [49] using CNN and ResNet50 architecture for image-based skin cancer classification from the ISIC dataset (2,357 images, 9 classes). The proposed model was constructed through several stages, including image preprocessing (hair removal, noise reduction, and data augmentation), segmentation. The resulting accuracy is 91.32%.

Based on previous studies, the CNN architecture are effective in classifying skin cancer. However, most studies still have limitations, such as not merging datasets and only evaluating one to three model architectures. This study aims to overcome these limitations by combining two large-scale public datasets and conducting a comprehensive evaluation of seven CNN architectures. In addition, modifications were made to the best model to explore the potential for increasing accuracy through fine-tuning strategies and adding network layers.

3. Method

This research consists of several main stages, namely image acquisition, which is done by combining two large-scale public datasets, then preprocessing image to improve image quality through normalization and resizing. Furthermore, the application of a deep learning model is carried out by comparing seven CNN architectures. After obtaining the best model, the architecture is modified through a fine-tuning strategy and adding layers. The final stage is validation and evaluation to measure model validation and evaluation using performance metrics. Fig. 1 presents the complete research workflow starting from multi-source dataset acquisition, preprocessing image, model training, best model modification, until evaluation model.

3.1. Image Acquisition

The study commenced with the dermatoscopic images of skin cancer as the primary data source [50]. The dataset comprised 13,402 images categorized into 2 primary classes: benign and malignant. The data were obtained by combining two public datasets from the Kaggle platform that have different visual characteristics and number of images. The first dataset is Skin Cancer: Malignant vs.

Benign which consists of 3,297 images with two classes and includes testing and training data [41]. The second dataset is the Melanoma Skin Cancer Dataset of 10,000 Images which provides 10,105 images with the same label and more complex visual variations [42]. The purpose of combining these two datasets is to increase the diversity of visual features, strengthen the model's generalization ability, and minimize potential bias due to limited data from a single source [14], [51].

The distribution of the number of images by category and dataset source after class balancing using undersampling techniques is presented in Table 2. The amount of data used was adjusted to ensure balance between classes. Fig. 2 shows visual examples images [41], [42].

Table 1. Related works

Researcher	Architecture / Method	Dataset	Accuracy	Research Findings
Mudassir Saeed et al. [39]	VGG16, VGG19, VGG19 + SVM	ISIC 2019 & 2020	VGG19+SVM: 96%	Hybrid approach improves classification accuracy
Gururaj et al. [5]	DenseNet169, ResNet50	Dataset: HAM10000 and ISIC	DenseNet169: 91.20% ResNet50: 83%	DenseNet169 is superior in feature extraction
Ashutosh Lembhea et al. [46]	InceptionV3 + ISR	ISIC	Sebelum ISR: 83.48%Setelah ISR: 91.26%	Increasing image resolution improves model accuracy.
Rifat Sadik et al. [40]	MobileNet, Xception	Darmnet dan HAM10000	MobileNet: 96%Xception: 97%	The lightweight model provides high accuracy and efficiency
Firasari et al. [47]	ResNet50, MobileNetV2 + berbagai optimizer	ISIC	Keduanya: 88%	MobileNetV2 shows lower loss values and more stable learning.
Khasanah et al. [48]	InceptionV3	ISIC	InceptionV3: 86.65%	Model is not optimal; more data and further tuning are needed
Kavitha et al. [49]	ResNet50	ISIC	ResNet50: 91.32%	The use of ResNet50 with deep learning architecture significantly improves classification efficiency and accuracy.

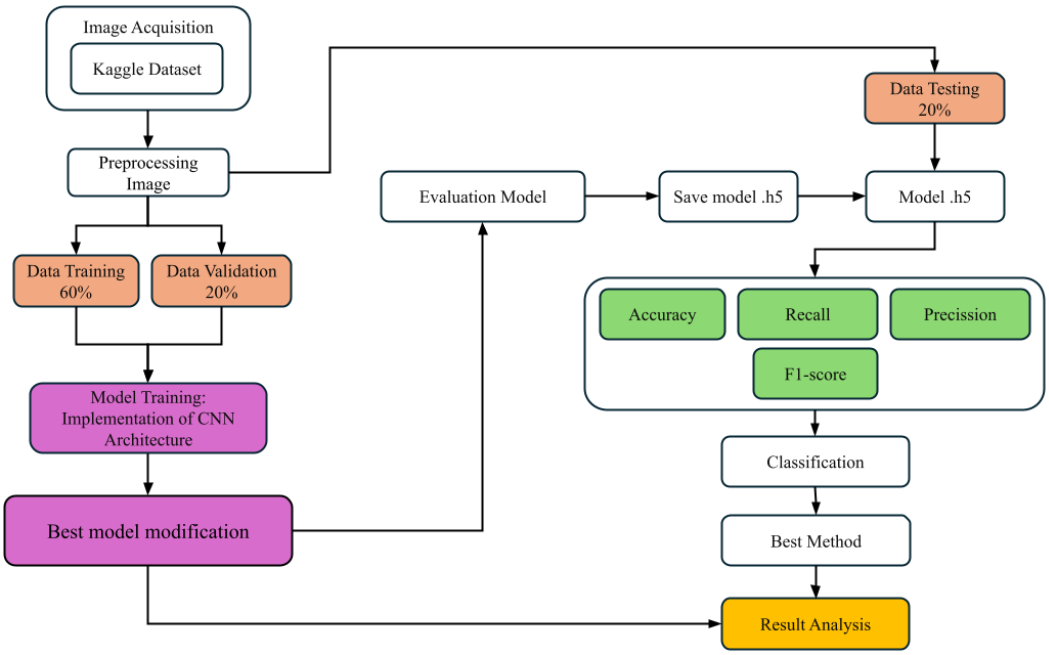
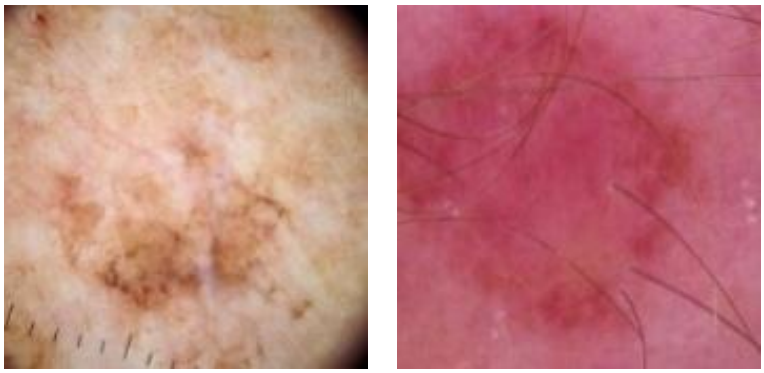


Fig. 1. Research method

Table 2. Dataset distribution by class

No	Class	Total images
1	Malignant	6,587
2	Benign	6,587



Malignant Benign

Fig. 2. Sample image for each class

3.2. Preprocessing Image

All datasets were standardized to a fixed resolution of 224×224 pixels, then normalized to stabilize the training process. The data is then divided using a tiered scheme, according to the allocations shown in Table 3 [52]. To reduce the risk of data leakage due to the emergence of very similar or duplicate images, an image similarity checking process is carried out before data distribution, namely by applying file hashing to remove identical images and perceptual hashing to detect images that have high similarity, so that the potential distribution of duplicate images can be minimized. The training process is carried out using v6e-1 TPU.

This data partitioning strategy was designed to provide sufficient samples for model training, enable effective validation during the learning process, and ensure evaluation using an independent test set [53]. This approach results in an accurate and objective evaluation of model performance, while reducing potential bias and increasing the reliability of skin cancer image classification results [54].

Table 3. Distribution of images across dataset subsets

Subset	Total
Training (60%)	7,904
Validation (20%)	2,635
Testing (20%)	2,635
Total	13,174

3.3. Model Training: Implementation of the CNN Model

The benchmarking phase employs a transfer learning strategy to compare seven CNN architectures. Each model is initialized with pre-trained weights from the ImageNet dataset [55]. Model training employed the Adam optimizer with a learning rate of 0.0001 for stable transfer learning, while the batch size was selected to balance efficiency and training stability. Binary Crossentropy was used as the loss function for the binary classification task [56]. Model performance was monitored at each training epoch using accuracy and loss metrics computed on the training and validation datasets [57]. Model performance was monitored at each training epoch by assessing accuracy and loss on both the training and validation datasets. The best model of each architecture is saved in .h5 file format for use in the testing phase.

Fig. 3 illustrates the transfer learning workflow applied to the seven CNN architectures, offering an overview of the training and implementation stages used in this study. The stage begins by inputting a 224×224×3 image into the model, which is the standard input size for ImageNet-pretrained CNN architectures used in this study, enabling fair benchmarking while preserving sufficient spatial information for classifying benign and malignant lesions. The models were trained for 20 epochs using the Binary Crossentropy loss function, which is appropriate for binary classification tasks [58]. To prevent overfitting, a dropout technique of 0.5 and an early stopping strategy with a patience value of 5 epochs were applied, which automatically stops training if there is no increase in performance on the validation data for five consecutive epochs [59], [60]. At the end of the architecture, a fully connected layer with 128 neurons and ReLU activation is incorporated, followed by a sigmoid-activated output layer to generate binary class predictions [61].

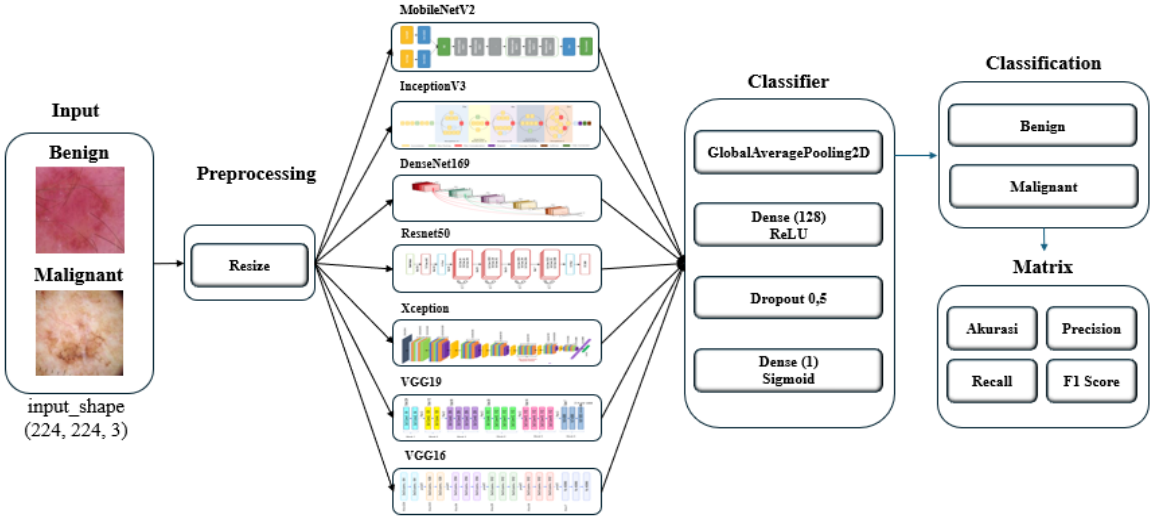


Fig. 3. Architecture of the proposed CNN model

3.4. Selection and Optimization of the CNN Model

After completing the benchmarking phase, the best-performing architecture is selected for further optimization. Model optimization is performed by applying a fine-tuning strategy on the backbone layers and adding architectural components to improve performance [62]. To prevent overfitting, a dropout technique of 0.5 and an early stopping strategy with a patience value of 5 epochs were applied, which automatically stops training if there is no increase in validation performance for five consecutive epochs.

At the end of the optimized architecture, a dense layer with 128 neurons and a ReLU activation function is added, followed by an output layer with sigmoid activation to produce binary class predictions. This optimization phase aims to enhance model robustness and improve diagnostically relevant performance, particularly in reducing false-negative predictions [62]. Fig. 4 illustrates the architecture of the optimized CNN model.

3.5. Evaluation Model

Model performance was assessed on the validation dataset using accuracy, recall, precision, and F1-score as evaluation metrics [63]. If necessary, parameter tuning and model optimization are performed to improve and enhance performance. If the model performance is deemed inadequate, a model optimization and tuning process can be performed, which involves adjusting hyperparameters [64], modification of model architecture, or application of regularization techniques to improve performance and prevent overfitting [65]. Next, after the seven models are evaluated, determine the best model. Indicators used to measure achievements:

1. Accuracy

Accuracy is a performance metric that represents the proportion of correctly classified instances relative to the total number of predictions produced by the model [66]. This metric reflects how well the model classifies the data as a whole [67].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \; x \; 100 \tag{1}$$

2. Precision and Recall

Precision quantifies the reliability of the model's positive predictions by indicating the proportion of correctly identified positive instances among all predicted positives [68]. Recall, also referred to as sensitivity, measures the model's ability to correctly identify all true positive instances [69]. Recall focuses on how many positive data the model successfully detects out of all the positive data. These two metrics provide insight into the balance between correct detections and false detections [70].

$$Precision = \frac{TP}{TP + FP} \; x \; 100 \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \; x \; 100 \tag{3}$$

3. F1-Score

An evaluation metric that combines precision and recall to provide a more balanced picture of the performance of a classification model, especially when there is a trade-off between the two [71].

$$F1 - score = 2x \frac{Recall \; x \; Precision}{Recall + \; Precision} \; x \; 100 \tag{4}$$

4. Confusion Matrix

The confusion matrix offers a detailed representation of the model's classification outcomes by summarizing correct predictions and misclassifications for each class [53], [72], [73]. Confusion matrix representation shown in Table 4.

Table 4. Confusion matrix representation [74]

	Predicted positive	**Predicted Negative**
Actual Positive	TP	FP
Actual negative	FN	TN

4. Results and Discussion

4.1. Main Findings

This research compares the performance of seven CNN architectures using a multi-source dataset comprising 13,174 dermatoscopic images. To ensure a consistent evaluation, all models were trained using identical hyperparameter configurations. The baseline results of these architectures are presented in Table 5. Fig. 5 shows comparison of classification accuracy across the evaluated CNN architectures and Fig. 6 shows comparison of loss values across the evaluated CNN architectures.

As reported in Table 5, DenseNet169 achieved the highest test accuracy of 89.30%. This superior performance is attributed to DenseNet's dense connectivity mechanism, which enables effective feature reuse and improved gradient flow, allowing the model to retain discriminatory

lesion characteristics and maintain a good balance between precision and sensitivity. The Xception and MobileNetV2 models demonstrated competitive performance despite their relatively lighter architectures. In contrast, the ResNet50 model demonstrated suboptimal performance, with an accuracy of only 74.90% and a low F1 score for the malignant class. This suggests possible overfitting or a mismatch of the architecture to the characteristics of the target data.

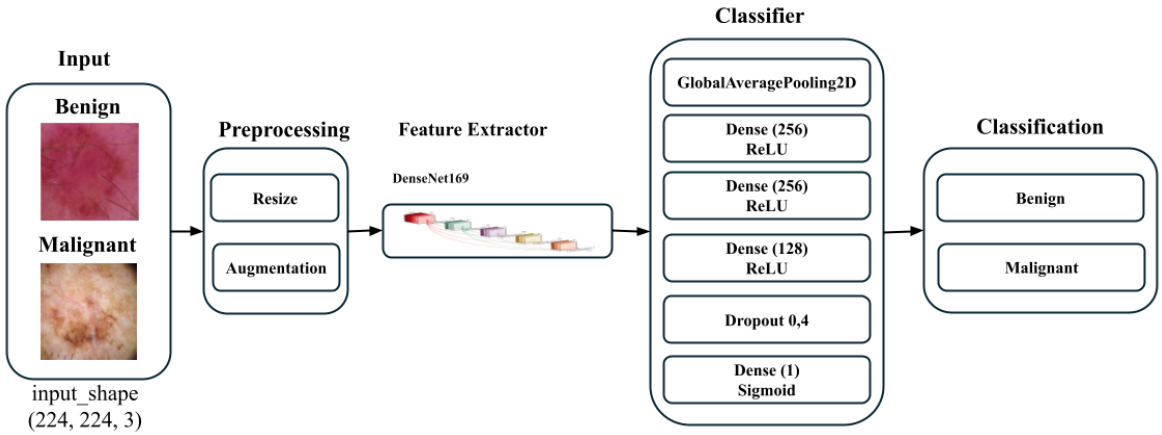


Fig. 4. Architecture of the optimized CNN model

Table 5. Performance comparison of CNN models

Method	Train Acc	Train Loss	Val Acc	Val Loss	Test Acc	Test Loss	Total inference time	Size Model
Mobilenet	88.53%	26.95%	87.52%	29.41%	87.80%	12.20%	8.17’	11 KB
DenseNet169	91.13%	21.61%	89.35%	24.74%	89.30%	10.70%	33.25’	53 KB
InceptionV3	87.20%	30.24%	86.49%	32.38%	86.70%	13.30%	13.69’	89 KB
Xception	89.09%	26.44%	87.49%	30.63%	86.90%	13.10%	23.03’	85 KB
ResNet50	65.18%	61.71%	70.17%	60.06%	74.90%	25.10%	19.27’	95 KB
VGG16	83.96%	36.19%	85.20%	34.16%	85.60%	14.40%	37.93’	58 KB
VGG19	82.84%	38.85%	84.40%	36.24%	85.20%	14.80%	47.13’	79 KB

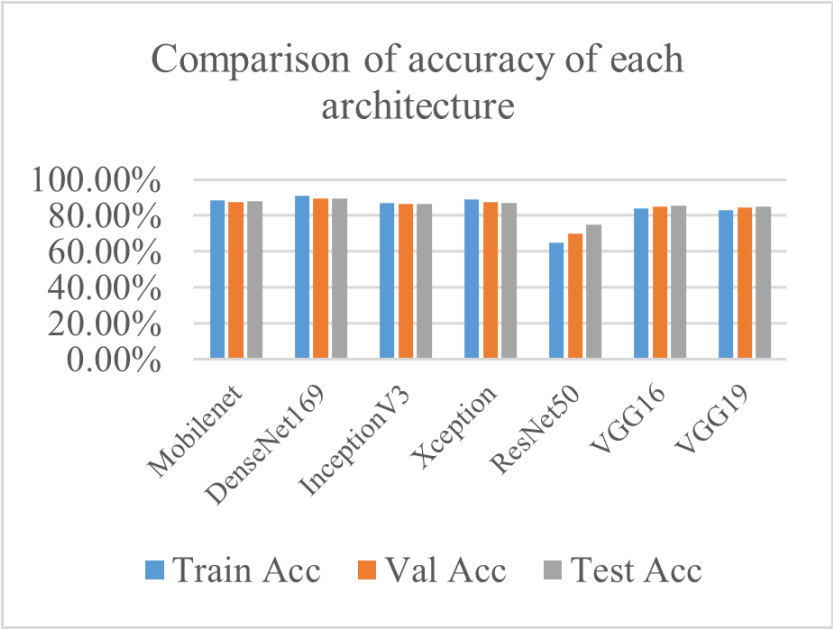


Fig. 5. Accuracy comparison across CNN architectures

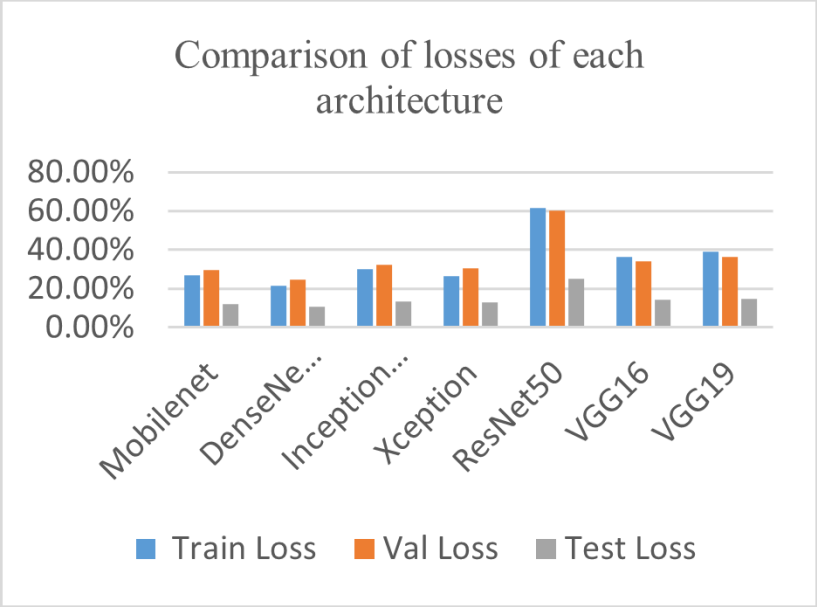


Fig. 6. Loss comparison across CNN architectures

In the training vs validation accuracy graph in Fig. 7, Fig. 8, the model accuracy increases steadily during the early epochs until it drops or stagnates near the end of training, indicating the optimal learning point. Meanwhile, the training vs validation loss graph shows a significant decrease in loss in the early phase of training, but the difference between training loss and validation loss at the end of training indicates mild overfitting. Therefore, additional regularization strategies such as early stopping or dropout prove important to control the model performance in general. To further analyze classification errors of the DenseNet169 baseline model is shown in Fig. 9.

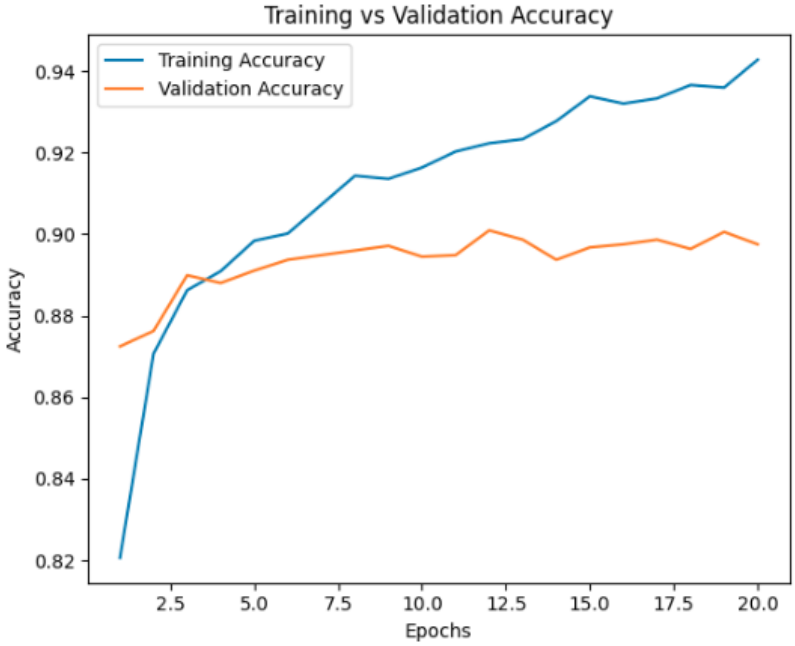


Fig. 7. Training and validation accuracy curves of the DenseNet169

Fig. 9 indicates that the baseline DenseNet169 correctly classified 1,177 benign and 1,175 malignant images, but still produced 165 false-negative cases in the benign class and 118 false-negative cases in the malignant class.

The results demonstrate that the model effectively differentiates between benign and malignant cases, as indicated by high True Positive and True Negative values. However, the occurrence of 118 false-negative cases remains clinically significant, as missed malignant lesions may lead to delayed diagnosis. This highlights the need for further optimization to improve sensitivity toward malignant skin cancer.

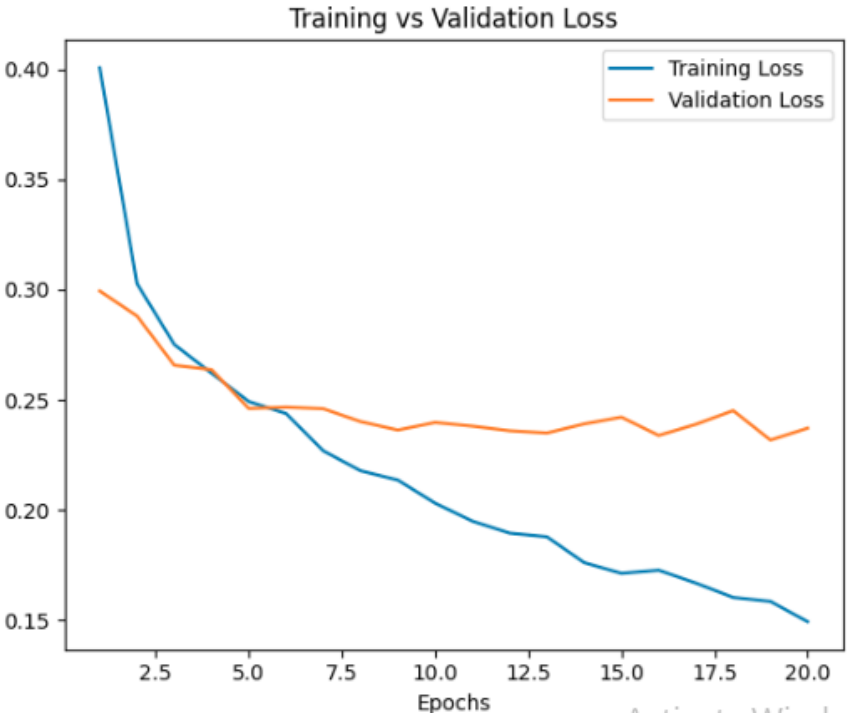


Fig. 8. Training and validation loss curves of the DenseNet169

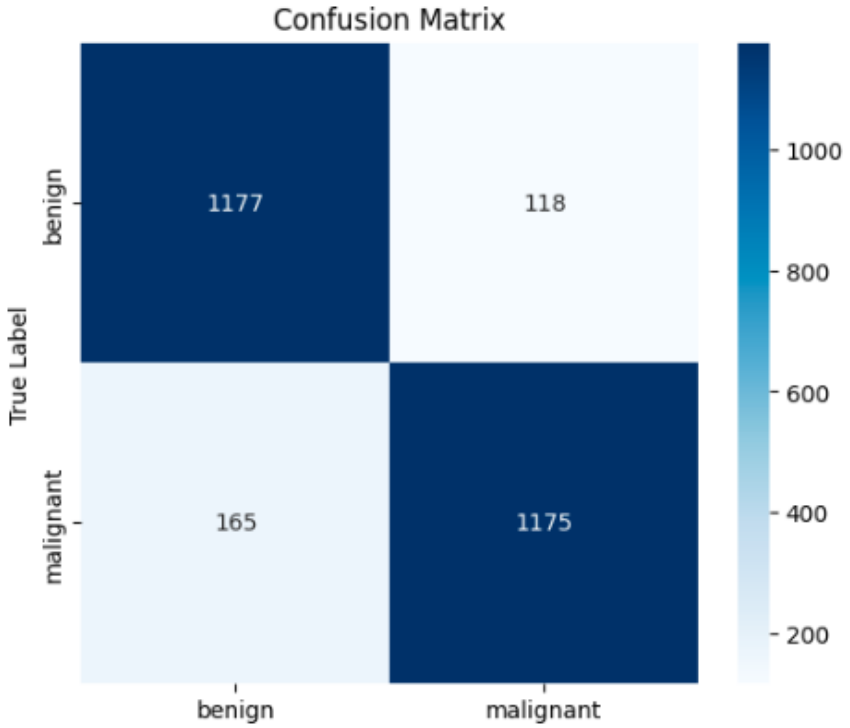


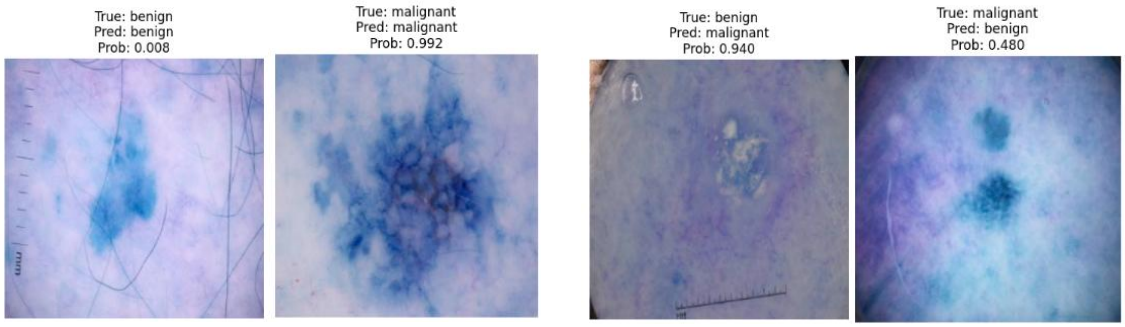
Fig. 9. Confusion matrix of DenseNet169

Table 6 summarizes the performance evaluation. Overall, the model attained an accuracy of 89.30%, indicating that the majority of images were classified correctly. In addition, the macro-averaged and weighted-averaged precision, recall, and F1-score values ranged from 0.877 to 0.909, respectively, indicating a balanced and consistent model performance across both classes, without dominance by one class. These evaluation results indicate that the model performs well in classifying skin cancer images. Although the recall value for the malignant class is slightly lower, overall metrics indicate that the model is quite reliable and can be used as a basis for an early detection system for skin cancer based on dermatoscopic images. Here is the implementation of the DenseNet169 model for image classification with testing data:

Table 6. Classification report Densenet169

	Precision	Recall	F1-Score	Support
Benign	0.877	0.909	0.893	1295
Malignant	0.909	0.877	0.893	1340
Accuracy			0.893	2635
Macro Avg	0.893	0.893	0.893	2635
Weighted Avg	0.893	0.893	0.893	2635

Fig. 10 shows the prediction results of the DenseNet169 model on dermatoscopic images of skin cancer, which are divided into two categories, namely true predictions and false predictions. Fig. 10 a shows an example of an image that was successfully classified correctly by the model. In the upper left image, a skin lesion that is actually in the benign class is successfully predicted with the same label (benign) with a probability of 0.008. Meanwhile, in the upper right image, the model also identifies the lesion as benign according to the actual label, with a higher prediction probability of 0.992 (which means the model is confident that the image is not malignant). Meanwhile, Fig. 10 b shows an example of an incorrect prediction. In the lower left image, a lesion that is actually benign is predicted as malignant with a probability of 0.940, indicating that the model has high false confidence in the image. On the other hand, the lower right image shows a misclassification of a malignant lesion that is predicted as benign with a probability of 0.480, indicating that the model is not confident enough in recognizing the characteristics of malignancy in the lesion. This visualization provides important insights into how the model recognizes visual patterns in each class. Prediction errors, especially in false negative cases (malignant predicted as benign), emphasize the importance of improving the sensitivity of the model in identifying malignant skin cancers, given the serious clinical implications of misdiagnosing such cases.



(a) Correct prediction (b) Wrong prediction

Fig. 10. Prediction image using DenseNet169

Among the seven evaluated deep learning architectures, DenseNet169 achieved the best performance with a test accuracy of 89.30%. This model showed good classification performance with a precision of 87.70% for the benign class and 90.90% for the malignant class. The recall value achieved was 90.90% for benign and 87.70% for malignant. Meanwhile, the F1-score showed a fairly good balance between the two classes, namely 89.30% for benign and 89.30% for malignant. These

results indicate that DenseNet169 has reliable and relatively stable classification performance in distinguishing the two categories of skin cancer. To explore performance improvements, modifications were made to the DenseNet169 architecture by adding dense layers, adjusting the number of neurons, applying a dropout technique with a level of 0.4 as regularization, and a fine-tuning strategy for some layers in the base model. The optimized training process still uses the same hyperparameters. Densenet169 architecture comparison shown in Table 7.

Table 8 reports the performance of the modified DenseNet169 model. The modified model achieved an average training accuracy of 95.60% with a training loss of 11.09%. On the validation set, it obtained an average accuracy of 91.04% and a loss value of 31.60%. Furthermore, testing results indicate a test accuracy of 91.20% and 8.80% test loss.

Table 7. Densenet169 architecture comparison

DenseNet169	DenseNet169 Modifikasi
Dense(128) 1 layer Dense + Dropout	Dense(256) → Dropout(0.4) → Dense(128) → Dense(1) 3 layer Dense + 1 Dropout

Table 8. Densenet169 modified model results

Method	Train Acc	Train Loss	Val Acc	Val Loss	Test Acc	Test Loss	Total inference time	Size Model
DenseNet169 Modified	95.60%	11.09%	91.04%	31.60%	91.20%	8.80%	33.25	53 KB

Fig. 11 presents the confusion matrix of the modified DenseNet169 model. Overall, the confusion matrix results reflect that the model has good generalization ability to both classes. The reliability of the model in detecting true malignant images further supports its effectiveness to be applied as a non-invasive early diagnosis tool in dermatology. However, the number of False Positives still needs to be considered so as not to cause unnecessary clinical alarms. Further optimization efforts, such as adjusting the classification threshold or integrating with additional clinical scoring systems, can be considered to achieve a better balance of precision.

Table 9 presents the classification report of the modified DenseNet169 model classification evaluation. For the benign class, the model recorded a precision value of 0.893, indicating that 89.30% of all benign predictions were correct. The recall reached 0.933, meaning the model was able to correctly identify 93.30% of all benign cases. The F1-score value reached 0.912, indicating an excellent balance between precision and sensitivity.

Table 9. Classification report Densenet169 modified

	Precision	Recall	F1-Score	Support
Benign	0.893	0.933	0.912	1295
Malignant	0.932	0.892	0.912	1340
Accuracy			0.912	2635
Macro Avg	0.912	0.912	0.912	2635
Weighted Avg	0.913	0.912	0.912	2635

Meanwhile, for the malignant class, the precision was 0.932, the recall was 0.893, and the F1-score was 0.912. These values indicate that the model is not only accurate in predicting malignant skin cancer, but also quite sensitive in detecting its presence, which is very important in the context of medical applications. Overall, the model produced a classification accuracy of 91.20% on a total of 2,635 test images. The classification report were all in the range of 0.893–0.933, indicating that the model performance was stable and balanced against both classes, without any dominance or bias towards one particular class.

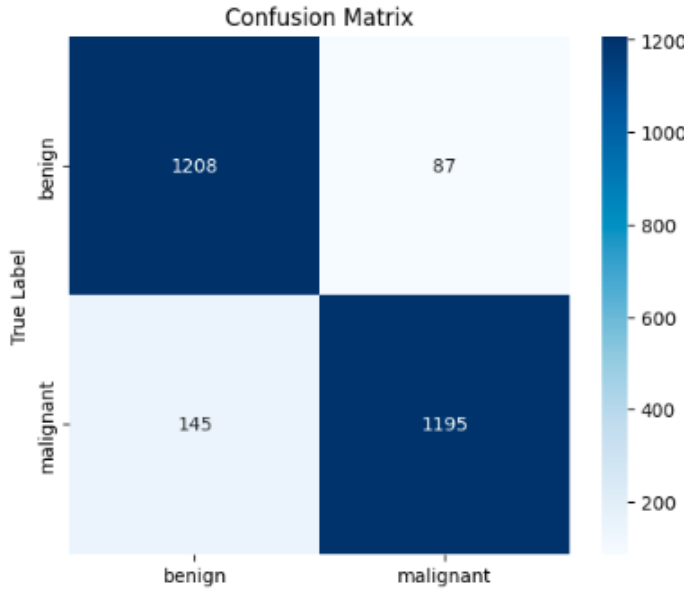
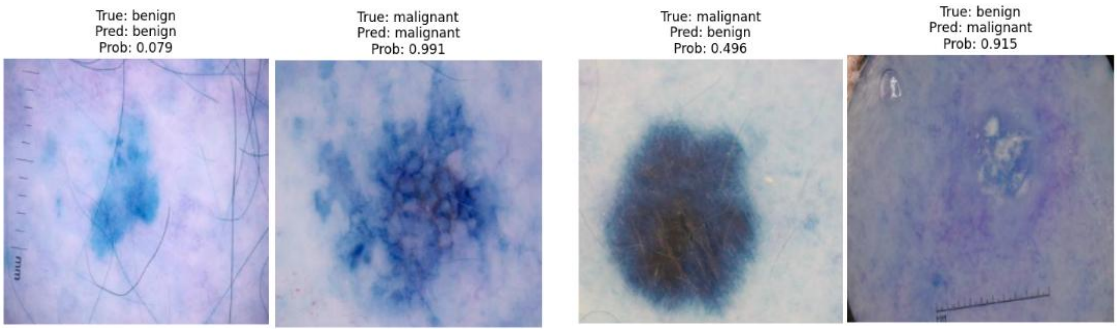


Fig. 11. Confusion matrix of DenseNet169 modified

Fig. 12 presents a visual illustration of the prediction results of the modified DenseNet169 model, which includes examples of correct and incorrect predictions on benign skin cancer images. Fig. 12 a shows an image that was successfully classified correctly, namely a skin lesion that was actually labeled benign was correctly predicted by the model as benign, with a prediction probability of 0.079. The image shows a light blue lesion morphology with relatively smooth and inconspicuous edges, which are visually common in benign lesions. The low probability value for the malignant class reflects the high level of confidence of the model in the resulting classification. In contrast, Fig. 12 b shows an image that was misclassified. A skin lesion that was actually labeled benign was incorrectly predicted as malignant, with a fairly high probability of 0.915. This shows that even though the model has been modified, there are still challenges in distinguishing certain visual features, especially in images that have a darker appearance, irregular edges, or textures that resemble malignant lesions.



(a) Correct prediction (b) Wrong prediction

Fig. 12. Prediction image using DenseNet169 modified

To add variety to the research results, a performance comparison was conducted between the standard DenseNet169 model and the modified DenseNet169. Table 10 shows that the DenseNet169 Modified model shows an increase in performance compared to the standard DenseNet169 in all evaluation metrics. Accuracy increased from 89.30% to 91.20%, accompanied by a decrease in loss value from 10.70% to 8.80%, indicating a more stable training process. In addition, the precision value increased from 87.70% to 89.30%, recall from 90.09% to 93.30%, and F1-score from 89.30% to 91.20%. In addition to accuracy, the evaluation also uses AUC-ROC as a more robust diagnostic

metric. The standard DenseNet169 model obtained an AUC of 0.9634, while the DenseNet169 modified model obtained an AUC of 0.9714. These results indicate that the architectural modification of DenseNet169 has succeeded in improving the overall performance.

Table 10. Comparison of Densenet169 results with Densenet169 modified model

Method	Train Acc	Train Loss	Val Acc	Val Loss	Test Acc	Test Loss	AUC	Total inference time	Size Model
DenseNet 169	91.13%	21.61%	89.35%	24.74%	89.30%	10.70%	0.9634	33.25’	53 KB
DenseNet16 9 Modified	95.60%	11.09%	91.04%	31.60%	91.20%	8.80%	0.9714	33.25’	53 KB

These results strengthen the conclusion that the modification of the DenseNet169 architecture provides significant performance improvements and produces a robust model for the task of classifying two classes of skin cancer, and is worthy of consideration for implementation as a digital image-based diagnostic support system.

Fig. 13 illustrates the ROC curve of the proposed model, which achieves an AUC value of 0.9714. This result indicates a strong discriminative capability in separating benign and malignant skin lesions across different decision thresholds. The ROC curve closely approaches the upper-left region, demonstrating a high true positive rate alongside a low false positive rate an essential requirement in medical diagnostic applications where reducing false-negative outcomes is crucial. Overall, these findings confirm the robustness and reliability of the proposed model.

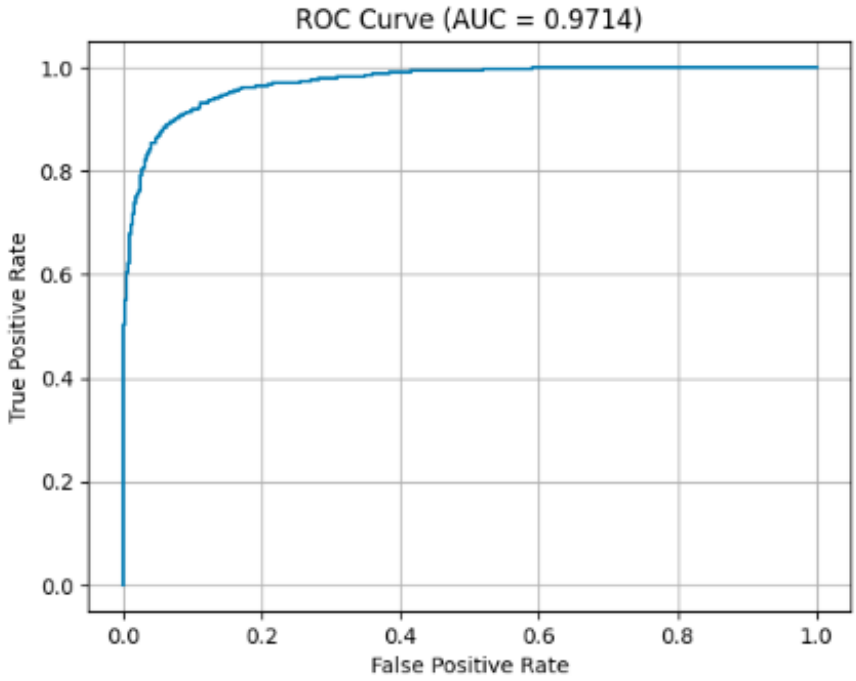


Fig. 13. ROC Curve

McNemar’s test was applied to assess whether the performance gain of the optimized DenseNet169 model over the baseline version was statistically significant. The test produced a statistic value of 14.00 with a p-value of 0.00018, confirming a significant difference ($p < 0.05$). These results indicate that the observed improvement in classification performance is statistically meaningful and not attributable to random variation.

4.2. Comparison with Other Studies

The results obtained in this study are consistent with and extend findings reported in previous skin cancer classification research. Prior works have demonstrated the effectiveness of CNN-based architectures such as DenseNet169, VGG19, MobileNet, and Xception for dermatoscopic image analysis, with reported accuracies ranging between 88% and 96% depending on the dataset and task complexity. For instance, Gururaj et al. [5] reported a DenseNet169 accuracy of 91.20% on the HAM10000 dataset, while studies using lightweight architectures such as MobileNet and Xception achieved high accuracy on single-source datasets with limited class diversity.

Compared to these studies, the proposed approach differs in two key aspects. First, this study integrates multi-source dermatoscopic datasets, introducing higher variability in image quality, color distribution, and acquisition conditions, which better reflects real-world clinical scenarios. Second, a comprehensive benchmarking of seven CNN architectures was conducted under identical training conditions, followed by targeted architectural optimization of the best-performing model. The optimized DenseNet169 achieved a higher test accuracy and improved AUC-ROC compared to baseline implementations reported in earlier works, while also demonstrating a reduction in false-negative cases. These results indicate that architectural refinement combined with multi-source data integration can enhance model robustness beyond what is achieved using single-dataset evaluations.

4.3. Implication and Explanation of Findings

The strong performance of DenseNet169 observed in this study is largely due to its dense connectivity structure, which promotes effective feature reuse and enhances gradient propagation across network layers. This design is especially beneficial for dermatoscopic image analysis, where fine-grained variations in texture, color, and lesion boundaries are critical for differentiating benign and malignant cases. By preserving low-level and high-level features throughout the network, DenseNet169 enables more discriminative representation learning compared to architectures with isolated layer connections.

The architectural optimization applied in this study further enhances this capability by refining dense layer configurations, applying dropout regularization, and selectively fine-tuning backbone layers. The resulting reduction in false-negative predictions is especially significant from a clinical perspective, as misclassifying malignant lesions as benign may lead to delayed diagnosis and adverse patient outcomes. Therefore, the findings suggest that the proposed optimized model is not only effective in terms of overall accuracy but also more reliable for clinically sensitive decision-making, where sensitivity to malignant cases is critical.

4.4. Strengths and Limitations

A major strength of this study is the use of a large multi-source dataset, which enhances data diversity and improves model generalization. Additionally, the comprehensive benchmarking of multiple CNN architectures under consistent experimental settings provides a fair and systematic comparison, while the targeted optimization of DenseNet169 demonstrates measurable performance gains without significantly increasing model complexity. The inclusion of clinically relevant metrics such as AUC-ROC and confusion matrix analysis further strengthens the practical relevance of the findings.

Despite these strengths, several limitations must be considered. First, the study relies exclusively on dermatoscopic images, which may limit generalization to other image modalities such as smartphone-acquired clinical images. Second, external validation across independent clinical institutions was not performed, which is necessary to assess real-world deployment readiness. Third, the study does not include explainability methods such as Grad-CAM, limiting interpretability of the model's decision-making process. Addressing these limitations through broader data sources, explainable AI integration, and prospective multi-center validation is an important direction for future research.

5. Conclusion

This study demonstrates that integrating multi-source datasets with targeted architectural optimization improves the robustness and generalization of CNN-based model. DenseNet169 achieved the strongest baseline performance, and its optimized variant further reduced false-negative rates in malignant cases, which is clinically critical for early diagnosis and improved patient outcomes. Theoretically, this work contributes new evidence that architectural refinement within transfer learning frameworks enhances diagnostic reliability in heterogeneous medical imaging environments.

From a computational perspective, DenseNet169 offers superior accuracy but requires higher inference time and resources compared to lightweight models such as MobileNetV2, which are more suitable for real-time or resource-constrained clinical applications. This trade-off highlights the importance of selecting models based on deployment context.

This study has several limitations: (1) reliance on dermatoscopic images only, (2) limited external validation across clinical institutions, and (3) the absence of explainability analysis such as Grad-CAM. Future work should evaluate performance on smartphone-acquired clinical images, integrate patient metadata, apply explainable AI techniques, and conduct prospective multi-center validation studies. Overall, this research provides a practical and extensible foundation for developing reliable AI-assisted clinical decision support systems for early skin cancer detection.

Author Contribution: All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Acknowledgment: The author expresses his deepest appreciation to fellow researchers and the entire team who contributed throughout the research process. Thanks are also addressed to family and friends for the moral encouragement given, as well as to all parties who contributed to completing this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] R. Bogne Tchema, A. C. Polycarpou, and M. Nestoros, "Skin cancer classification using machine learning," *Multimedia Tools and Applications*, vol. 84, no. 6, pp. 3239-3256, 2025, https://doi.org/10.1007/s11042-025-20595-7.

[2] H. M. Balaha and A. E. S. Hassan, "Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm," *Neural Computing and Applications*, vol. 35, no. 1, pp. 815-853, 2023, https://doi.org/10.1007/s00521-022-07762-9.

[3] M. Gallazzi, S. Biavaschi, A. Bulgheroni, T. M. Gatti, S. Corchs and I. Gallo, "A Large Dataset to Enhance Skin Cancer Classification With Transformer-Based Deep Neural Networks," *IEEE Access*, vol. 12, pp. 109544-109559, 2024, https://doi.org/10.1109/ACCESS.2024.3439365.

[4] K. Mridha, M. M. Uddin, J. Shin, S. Khadka and M. F. Mridha, "An Interpretable Skin Cancer Classification Using Optimized Convolutional Neural Network for a Smart Healthcare System," *IEEE Access*, vol. 11, pp. 41003-41018, 2023, https://doi.org/10.1109/ACCESS.2023.3269694.

[5] H. L. Gururaj, N. Manju, A. Nagarjun, V. N. M. Aradhya and F. Flammini, "DeepSkin: A Deep Learning Approach for Skin Cancer Classification," *IEEE Access*, vol. 11, pp. 50205-50214, 2023, https://doi.org/10.1109/ACCESS.2023.3274848.

[6] R. Rastghalam, H. Danyali, M. S. Helfroush, M. E. Celebi and M. Mokhtari, "Skin Melanoma Detection in Microscopic Images Using HMM-Based Asymmetric Analysis and Expectation Maximization," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, pp. 3486-3497, 2021, https://doi.org/10.1109/JBHI.2021.3081185.

[7] G. Yang, S. Luo, and P. Greer, "A Novel Vision Transformer Model for Skin Cancer Classification," *Neural Processing Letters*, vol. 55, no. 7, pp. 9335-9351, 2023, https://doi.org/10.1007/s11063-023-11204-5.

[8] E. H. Houssein, D. A. Abdelkareem, G. Hu, M. A. Hameed, I. A. Ibrahim, and M. Younan, "An effective multiclass skin cancer classification approach based on deep convolutional neural network," *Cluster Computing*, vol. 27, no. 9, pp. 12799-12819, 2024, https://doi.org/10.1007/s10586-024-04540-1.

[9] A. A. Abdulredah, M. A. Fadhel, L. Alzubaidi, Y. Duan, M. Kherallah, and F. Charfi, "Towards unbiased skin cancer classification using deep feature fusion," *BMC Medical Informatics and Decision Making*, vol. 25, no. 1, p. 48, 2025, https://doi.org/10.1186/s12911-025-02889-w.

[10] M. H. Trager, L. J. Geskin, F. H. Samie, and L. Liu, "Biomarkers in melanoma and non-melanoma skin cancer prevention and risk stratification," *Experimental Dermatology*, vol. 31, no. 1, pp. 4-12, 2022, https://doi.org/10.1111/exd.14114.

[11] A. K. Sharma *et al*., "Dermatologist-Level Classification of Skin Cancer Using Cascaded Ensembling of Convolutional Neural Network and Handcrafted Features Based Deep Neural Network," *IEEE Access*, vol. 10, pp. 17920-17932, 2022, https://doi.org/10.1109/ACCESS.2022.3149824.

[12] A. D. Khalaf, H. Hamdan, A. Abdul Halin and N. Manshor, "Segmentation and Classification of Skin Cancer Diseases Based on Deep Learning: Challenges and Future Directions," *IEEE Access*, vol. 13, pp. 90163-90184, 2025, https://doi.org/10.1109/ACCESS.2025.3569170.

[13] M. I. Faizi and S. M. Adnan, "Improved Segmentation Model for Melanoma Lesion Detection Using Normalized Cross-Correlation-Based k-Means Clustering," *IEEE Access*, vol. 12, pp. 20753-20766, 2024, https://doi.org/10.1109/ACCESS.2024.3360223.

[14] Q. Wang *et al*., "Identification of Melanoma From Hyperspectral Pathology Image Using 3D Convolutional Networks," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 218-227, 2021, https://doi.org/10.1109/TMI.2020.3024923.

[15] A. S. Qureshi and T. Roos, "Transfer Learning with Ensembles of Deep Neural Networks for Skin Cancer Detection in Imbalanced Data Sets," *Neural Processing Letters*, vol. 55, no. 4, pp. 4461-4479, 2023, https://doi.org/10.1007/s11063-022-11049-4.

[16] Q. Wang *et al*., "Identification of Melanoma From Hyperspectral Pathology Image Using 3D Convolutional Networks," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 218-227, 2021, https://doi.org/10.1109/TMI.2020.3024923.

[17] M. N. Hamza *et al*., "Designing a High-Sensitivity Microscale Triple-Band Biosensor Based on Terahertz MTMs to Provide a Perfect Absorber for Non-Melanoma Skin Cancer Diagnostic," *IEEE Photonics Journal*, vol. 16, no. 2, pp. 1-13, 2024, https://doi.org/10.1109/JPHOT.2024.3381649.

[18] H. He *et al*., "Machine Learning Analysis of Human Skin by Optoacoustic Mesoscopy for Automated Extraction of Psoriasis and Aging Biomarkers," *IEEE Transactions on Medical Imaging*, vol. 43, no. 6, pp. 2074-2085, 2024, https://doi.org/10.1109/TMI.2024.3356180.

[19] L. Riaz *et al*., "A Comprehensive Joint Learning System to Detect Skin Cancer," *IEEE Access*, vol. 11, pp. 79434-79444, 2023, https://doi.org/10.1109/ACCESS.2023.3297644.

[20] N. A. AlSadhan, S. A. Alamri, M. M. Ben Ismail, and O. Bchir, "Skin Cancer Recognition Using Unified Deep Convolutional Neural Networks," *Cancers*, vol. 16, no. 7, p. 1246, 2024, https://doi.org/10.3390/cancers16071246.

[21] Y. S. P. MD, Z. Apalla, G. Salerni, A. Patil, S. Grabbe, and M. Goldust, "Non-invasive diagnostic techniques in pigmentary skin disorders and skin cancer," *Journal of Cosmetic Dermatology*, vol. 21, no. 2, pp. 444-450, https://doi.org/10.1111/jocd.14547.

[22] A. A. Hussein, A. M. Montaser, and H. A. Elsayed, "Skin cancer image classification using hybrid quantum deep learning model with BiLSTM and MobileNetV2," *Quantum Machine Intelligence*, vol. 7, no. 2, p. 66, 2025, https://doi.org/10.1007/s42484-025-00288-y.

[23] N. Merlina, A. Prasetio, I. Zuniarti, N. A. Mayangky, D. N. Sulistyowati, and F. Aziz, "Deep CNN Models for Detecting Cervical Cancer in Pap Smear Images," *TEM Journal*, vol. 14, no. 2, pp. 1073-1083, 2025, https://doi.org/10.18421/TEM142-09.

[24] N. Merlina, A. Prasetyo, I. Zuniarti, N. A. Mayangky, D. N. Sulistyowati, and F. Aziz, "Improving Early Detection of Cervical Cancer Through Deep Learning-Based Pap Smear Image Classification," *Journal of Applied Data Sciences*, vol. 6, no. 2, pp. 969-980, 2025, https://doi.org/10.47738/jads.v6i2.576.

[25] T. Dhar, N. Dey, S. Borra and R. S. Sherratt, "Challenges of Deep Learning in Medical Image Analysis-Improving Explainability and Trust," *IEEE Transactions on Technology and Society*, vol. 4, no. 1, pp. 68-75, 2023, https://doi.org/10.1109/TTS.2023.3234203.

[26] N. Kumar Kar, S. Jana, A. Rahman, P. Rahul Ashokrao, I. G and R. Alarmelu Mangai, "Automated Intracranial Hemorrhage Detection Using Deep Learning in Medical Image Analysis," *2024 International Conference on Data Science and Network Security (ICDSNS)*, pp. 1-6, 2024, https://doi.org/10.1109/ICDSNS62112.2024.10691276.

[27] B. Haznedar and N. Y. Simsek, "A Comparative Study on Classification Methods for Renal Cell and Lung Cancers Using RNA-Seq Data," *IEEE Access*, vol. 10, pp. 105412-105420, 2022, https://doi.org/10.1109/ACCESS.2022.3211505.

[28] R. A. Welikala *et al*., "Automated Detection and Classification of Oral Lesions Using Deep Learning for Early Detection of Oral Cancer," *IEEE Access*, vol. 8, pp. 132677-132693, 2020, https://doi.org/10.1109/ACCESS.2020.3010180.

[29] S. Malarvannan and M. Angulakshmi, "A Review on Lung Cancer Classification Using Deep Learning Techniques," *IEEE Access*, vol. 13, pp. 76161-76184, 2025, https://doi.org/10.1109/ACCESS.2025.3564633.

[30] J. Guo, W. Cao, B. Nie and Q. Qin, "Unsupervised Learning Composite Network to Reduce Training Cost of Deep Learning Model for Colorectal Cancer Diagnosis," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 11, pp. 54-59, 2023, https://doi.org/10.1109/JTEHM.2022.3224021.

[31] T. Hidayat, N. Khasanah, E. Firasari, L. Kurniawati, and E. H. Hermaliani, "Comparative Evaluation of CNN Architectures for Skin Cancer Classification," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 16, no. 10, 2025, https://dx.doi.org/10.14569/IJACSA.2025.0161007.

[32] R. Kaur, H. GholamHosseini, and M. Lindén, "Advanced Deep Learning Models for Melanoma Diagnosis in Computer-Aided Skin Cancer Detection," *Sensors*, vol. 25, no. 3, p. 594, 2025, https://doi.org/10.3390/s25030594.

[33] J. Yang *et al*., "IoT-Driven Skin Cancer Detection: Active Learning and Hyperparameter Optimization for Enhanced Accuracy," *IEEE Journal of Biomedical and Health Informatics*, pp. 1-11, 2025, https://doi.org/10.1109/JBHI.2025.3578419.

[34] A. Jimi, N. Zrira, O. Guendoul, I. Benmiloud, H. A. Khan, and S. Nawaz, "ESC-UNET: A hybrid CNN and Swin Transformers for skin lesion segmentation," *Intelligence-Based Medicine*, p. 100257, 2025, https://doi.org/10.1016/j.ibmed.2025.100257.

[35] S. Selvi and A. Shalome, "Melanoma Detection by CNN Model with Optimal Rough Set Approximation-based Similarity Measure Classification," *Procedia Computer Science*, vol. 260, pp. 585-599, 2025, https://doi.org/10.1016/j.procs.2025.03.237.

[36] P. Nanda, D. Rout, and S. Kumari, "Multi-Class Skin Cancer Detection Using CNN-Architecture Based Deep Learning Models," *Procedia Computer Science*, vol. 260, pp. 226-235, 2025, https://doi.org/10.1016/j.procs.2025.03.197.

[37] R. K. Gupta, "Interpretable AI-Enabled Model for Skin Cancer Diagnosis using LIME," *Procedia Computer Science*, vol. 260, pp. 3-11, 2025, https://doi.org/10.1016/j.procs.2025.03.171.

[38] D. Keerthana *et al*., "Hybrid convolutional neural networks with SVM classifier for classification of skin cancer," *Biomedical Engineering Advances*, vol. 5, p. 100069, 2023, https://doi.org/10.1016/j.bea.2022.100069.

[39] M. Saeed, A. Naseer, H. Masood, S. U. Rehman and V. Gruhn, "The Power of Generative AI to Augment for Enhanced Skin Cancer Classification: A Deep Learning Approach," *IEEE Access*, vol. 11, pp. 130330-130344, 2023, https://doi.org/10.1109/ACCESS.2023.3332628.

[40] R. Sadik, A. Majumder, A. A. Biswas, B. Ahammad, and M. M. Rahman, "An in-depth analysis of Convolutional Neural Network architectures with transfer learning for skin disease diagnosis," *Healthcare Analytics*, vol. 3, p. 100143, 2023, https://doi.org/10.1016/j.health.2023.100143.

[41] C. Fanconi, "Skin Cancer: Malignant vs. Benign," *Kaggle*, 2025, https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign.

[42] M. H. Javid, "Melanoma Skin Cancer Dataset of 10000 Images," *Kaggle*, 2025, https://www.kaggle.com/datasets/hasnainjaved/melanoma-skin-cancer-dataset-of-10000-images.

[43] D. Husen, "Klasifikasi Citra MRI Tumor Otak Menggunakan Metode Convolutional Neural Network," *Journal bit-Tech*, vol. 7, no. 1, pp. 143-152, 2024, https://doi.org/10.32877/bt.v7i1.1576.

[44] B. Mittal, "Melanoma Cancer Image Dataset," *Kaggle*, 2025, https://www.kaggle.com/datasets/bhaveshmittal/melanoma-cancer-dataset.

[45] N. M. Rashad, N. M. Abdelnapi, A. F. Seddik, and M. A. Sayedelahl, "Automating skin cancer screening: a deep learning," *Journal of Engineering and Applied Science*, vol. 72, no. 1, p. 6, 2025, https://doi.org/10.1186/s44147-024-00573-w.

[46] A. Lembhe, P. Motarwar, R. Patil, and S. Elias, "Enhancement in skin cancer detection using image super resolution and convolutional neural network," *Procedia Computer Science*, vol. 218, pp. 164-173, 2023, https://doi.org/10.1016/j.procs.2022.12.412.

[47] E. Firasari, N. Khasanah, F. L. D. Cahyanti, D. N. Kholifah, U. Khultsum and F. Sarasati, "Performance Evaluation of ResNet50 and MobileNetV2 in Skin Cancer Image Classification with Various Optimizers," *2024 International Conference on Information Technology Research and Innovation (ICITRI)*, pp. 376-380, 2024, https://doi.org/10.1109/ICITRI62858.2024.10698943.

[48] D. U. E. Saputri, N. Khasanah, F. Aziz, and T. Hidayat, "Enhancing skin cancer classification using optimized inceptionv3 model," *Journal Medical Informatics Technology*, vol. 1, no. 3, pp. 65-69, 2023, https://doi.org/10.37034/medinftech.v1i3.14.

[49] C. Kavitha, S. Priyanka, M. P. Kumar, and V. Kusuma, "Skin cancer detection and classification using deep learning techniques," *Procedia Computer Science*, vol. 235, pp. 2793-2802, 2024, https://doi.org/10.1016/j.procs.2024.04.264.

[50] G. Yang, S. Luo, and P. Greer, "Advancements in skin cancer classification: a review of machine learning techniques in clinical image analysis," *Multimedia Tools and Applications*, vol. 84, no. 11, pp. 9837-9864, 2025, https://doi.org/10.1007/s11042-024-19298-2.

[51] R. Wang, H. Zuo, Z. Fang, and J. Lu, "Integrated Image-Text Augmentation for Few-Shot Learning in Vision-Language Models," *ACM Transactions on Intelligent Systems and Technology*, vol. 16, no. 2, pp. 1–19, 2025, https://doi.org/10.1145/3712700.

[52] C. Chakraborty, U. Achar, S. Nayek, A. Achar, and R. Mukherjee, "CAD-PsorNet: deep transfer learning for computer-assisted diagnosis of skin psoriasis," *Scientific Reports*, vol. 14, no. 1, p. 26557, 2024, https://doi.org/10.1038/s41598-024-76852-6.

[53] S. K. Mathivanan, S. Sonaimuthu, S. Murugesan, H. Rajadurai, B. D. Shivahare, and M. A. Shah, "Employing deep learning and transfer learning for accurate brain tumor detection," *Scientific Reports*, vol. 14, no. 1, p. 7232, 2024, https://doi.org/10.1038/s41598-024-57970-7.

[54] B. Santoshkumar and K. Deb, "Handling Objectives With Heterogeneous Evaluation Times in Surrogate-Assisted Evolutionary Multi-Objective Optimization," *IEEE Transactions on Evolutionary Computation*, 2025, https://doi.org/10.1109/TEVC.2025.3560922.

[55] Ş. Öztürk and T. Çukur, "Deep Clustering via Center-Oriented Margin Free-Triplet Loss for Skin Lesion Detection in Highly Imbalanced Datasets," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 9, pp. 4679-4690, 2022, https://doi.org/10.1109/JBHI.2022.3187215.

[56] S. T. Tabibi, A. Nikravanshalmani and H. Saboohi, "An Ensemble Classifier Based on Diverse Convolutional Neural Networks for Skin Lesions Classification," *IEEE Access*, vol. 13, pp. 195673-195686, 2025, https://doi.org/10.1109/ACCESS.2024.3442827.

[57] A. Meliboev, J. Alikhanov, and W. Kim, "Performance Evaluation of Deep Learning Based Network Intrusion Detection System across Multiple Balanced and Imbalanced Datasets," *Electronics*, vol. 11, no. 4, p. 515, 2022, https://doi.org/10.3390/electronics11040515.

[58] T. A. Suleiman, D. T. Anyimadu, A. D. Permana, H. A. A. Ngim, and A. Scotto di Freca, "Two-step hierarchical binary classification of cancerous skin lesions using transfer learning and the random forest algorithm," *Visual Computing for Industry, Biomedicine, and Art*, vol. 7, no. 1, p. 15, 2024, https://doi.org/10.1186/s42492-024-00166-7.

[59] H. Li, G. K. Rajbahadur, D. Lin, C. -P. Bezemer and Z. M. Jiang, "Keeping Deep Learning Models in Check: A History-Based Approach to Mitigate Overfitting," *IEEE Access*, vol. 12, pp. 70676-70689, 2024, https://doi.org/10.1109/ACCESS.2024.3402543.

[60] M. K. Anam, S. Defit, H. Haviluddin, L. Efrizoni, and M. B. Firdaus, "Early stopping on CNN-LSTM development to improve classification performance," *Journal of Applied Data Sciences*, vol. 5, no. 3, pp. 1175–1188, 2024, https://doi.org/10.47738/jads.v5i3.312.

[61] H. Bechinia, D. Benmerzoug and N. Khlifa, "Approach Based Lightweight Custom Convolutional Neural Network and Fine-Tuned MobileNet-V2 for ECG Arrhythmia Signals Classification," *IEEE Access*, vol. 12, pp. 40827-40841, 2024, https://doi.org/10.1109/ACCESS.2024.3378730.

[62] D. T. Pham, T. V. Tran, X. Zhu, and H. N. Pham, "Optimizing Deep Learning for Building Extraction: Dataset Efficiency and Model Backbones Under Data Constraints," *Remote Sensing Applications: Society and Environment*, p. 101876, 2026, https://doi.org/10.1016/j.rsase.2026.101876.

[63] J. P. Dominguez-Morales, J. -C. Hernández-Rodríguez, L. Duran-Lopez, J. Conejo-Mir and J. -J. Pereyra-Rodriguez, "Melanoma Breslow Thickness Classification Using Ensemble-Based Knowledge Distillation With Semi-Supervised Convolutional Neural Networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 1, pp. 443-455, 2025, https://doi.org/10.1109/JBHI.2024.3465929.

[64] M. A. Rasel, U. H. Obaidellah and S. A. Kareem, "Convolutional Neural Network-Based Skin Lesion Classification With Variable Nonlinear Activation Functions," *IEEE Access*, vol. 10, pp. 83398-83414, 2022, https://doi.org/10.1109/ACCESS.2022.3196911.

[65] T. S. Arulananth, S. W. Prakash, R. K. Ayyasamy, V. P. Kavitha, P. G. Kuppusamy and P. Chinnasamy, "Classification of Paediatric Pneumonia Using Modified DenseNet-121 Deep-Learning Model," *IEEE Access*, vol. 12, pp. 35716-35727, 2024, https://doi.org/10.1109/ACCESS.2024.3371151.

[66] V. -D. Hoang, X. -T. Vo and K. -H. Jo, "Categorical Weighting Domination for Imbalanced Classification With Skin Cancer in Intelligent Healthcare Systems," *IEEE Access*, vol. 11, pp. 105170-105181, 2023, https://doi.org/10.1109/ACCESS.2023.3319087.

[67] M. S. Islam *et al*., "Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach," *Artificial Intelligence Review*, vol. 57, no. 3, p. 62, 2024, https://doi.org/10.1007/s10462-023-10651-9.

[68] K. M. Hosny, D. Elshoura, E. R. Mohamed, E. Vrochidou and G. A. Papakostas, "Deep Learning and Optimization-Based Methods for Skin Lesions Segmentation: A Review," *IEEE Access*, vol. 11, pp. 85467-85488, 2023, https://doi.org/10.1109/ACCESS.2023.3303961.

[69] S. E. Sorour, A. A. Hany, M. S. Elredeny, A. Sedik and R. M. Hussien, "An Automatic Dermatology Detection System Based on Deep Learning and Computer Vision," *IEEE Access*, vol. 11, pp. 137769-137778, 2023, https://doi.org/10.1109/ACCESS.2023.3340735.

[70] S. Anuyah, M. K. Singh, and H. Nyavor, "Advancing clinical trial outcomes using deep learning and predictive modelling: bridging precision medicine and patient-centered care," *World Journal of Advanced Research and Reviews*, vol. 24, no. 3, pp. 001–025, 2024, https://doi.org/10.30574/wjarr.2024.24.3.3671.

[71] P. Dey, T. Mahmud, S. R. Nahar, M. S. Hossain and K. Andersson, "Plant Disease Detection in Precision Agriculture: Deep Learning Approaches," *2024 2nd International Conference on Intelligent Data*

Communication Technologies and Internet of Things (IDCIoT), pp. 661-667, 2024, https://doi.org/10.1109/IDCIoT59759.2024.10467525.

[72] S. Ness, "Securing Networks Against Adversarial Domain Name System Tunneling Attacks Using Hybrid Neural Networks," *IEEE Access*, vol. 13, pp. 46697-46709, 2025, https://doi.org/10.1109/ACCESS.2025.3550853.

[73] D. Chicco and G. Jurman, "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," *BioData Mining*, vol. 16, no. 1, p. 4, 2023, https://doi.org/10.1186/s13040-023-00322-4.

[74] M. Heydarian, T. E. Doyle and R. Samavi, "MLCM: Multi-Label Confusion Matrix," *IEEE Access*, vol. 10, pp. 19083-19095, 2022, https://doi.org/10.1109/ACCESS.2022.3151048.