**Accredited SINTA 2 Ranking**
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online at: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 8 No. 4 (2024) 506 - 515 e-ISSN: 2580-0760

# Skin Cancer Classification Using a Hybrid Pre-trained CNN with Random Forest Classifier

Okky Darmawan Kostidjan[1], Yudhi Purwananto[2], Anny Yuniarti[3]
[1,2,3]Teknik Informatika, Fakultas Teknologi Elektro dan Informatika Cerdas, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[1]OkkyDarmawanK@gmail.com, [2]Yudhi@its.ac.id, [3]Anny@its.ac.id

***Abstract***

*Skin cancer, which causes approximately 10 million deaths annually worldwide, is projected to see a rapid increase in cases if early diagnosis is not achieved. Traditional diagnostic methods, relying visual examination and histopathology, are often subjective and time-consuming. Recent advancements in Convolutional Neural Networks (CNN) have shown promise in automating and enhancing the accuracy of image analysis for the early detection of skin cancer. Current CNN approaches have leveraged transfer learning and hybrid models to improve performance. Nonetheless, the potential for overfitting remains, and there is still room for enhancing model accuracy. This study investigates the potential of pre-trained CNN models—such as DenseNet-201, InceptionV3, MobileNet, ResNet50, and VGG16—by modifying these models to improve their ability to differentiate between malignant and benign skin lesions. Additionally, a hybrid model approach is introduced, concatenating extracted features from various modified pre-trained CNN architectures and processing them through machine learning classifiers. The modifications and evaluations revealed that the proposed models surpassed the performance of state-of-the-art CNN models on ISBI 2016 datasets. The enhanced models achieved an impressive accuracy rate of 94.20%, marking a significant improvement over traditional CNN models and underscoring the potential of advanced CNN techniques in improving skin cancer diagnosis outcomes.*

Keywords: *skin cancer; classification; pre-trained CNN; hybrid model*

*How to Cite:* O. D. Kostidjan, Y. Purwanto, and A. Yuniarti, "Skin Cancer Classification Using a Hybrid Pre-trained CNN with Random Forest Classifier", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 8, no. 4, pp. 506 - 515, Aug. 2024
*DOI*: https://doi.org/10.29207/resti.v8i4.5857

## 1. Introduction

Skin cancer is one of the main causes of mortality on the global level, and up to 10 million people die worldwide because of it, as the WHO states. If a preliminary diagnosis is not made, skin cancer is expected to increase rapidly in the next 20 years [1]. It is fifth among the most diagnosed types of cancer [2]. The unusual growth of skin cells causes tumors that may be benign (the cells divide in an orderly manner, but do not disturb surrounding cells) or malignant (as their growth becomes abnormal and uncontrollable, these cells spread beyond the tumor to other parts of the body).

Every year, melanoma, one of the deadliest skin cancer classes, is responsible for the deaths of more than 55,500 people worldwide [3]. Melanoma poses the highest risk, although it accounts for only 5% of all cases of skin cancer. It is responsible for most of these skin cancer deaths, around 80% [4]. For a patient with melanoma with an incurable stage of the disease, the 5-year survival rate is very low, less than 20% [5]. Melanoma is a cancer in which early identification can really make a difference, and the survival rate can be as high as 95% if the diagnosis is made early [6].

In the past, the diagnosis of skin cancer was largely based on visual assessment of the lesions by experienced dermatologists and histopathological examinations. This process is not only subjective to individual bias and variations, but is also quite exhaustive, leading to hours of diagnosis and then treatment [7]. However, with recent advancements in deep learning, in particular Convolutional Neural Networks (CNN), which is a subset of deep learning, the possibility of changing the way skin cancer is diagnosed using automated image analysis has become a reality.

Received: 13-06-2024 | Accepted: 20-08-2024 | Published Online: 24-08-2024

Over time, artificial intelligence (AI) algorithms can now pick features and patterns in skin images that are malignant or benign with performance that is at least on par with that of human dermatologists. Clinicians can use dermoscopy combined with AI-based image analysis and standard histopathology to identify skin cancer at the preliminary stage, immediately start the appropriate treatment, and significantly improve the results [8]. Furthermore, intelligent imaging analytics complement clinical expertise, shortening and refining the analytical stages of diagnosis, and decreasing the work of a clinician. This revolutionary innovation that practically opened the door for the integration of AI in the field brings us to the dawn of a new era where life-changing technology is set to reduce the oncogenic rates of skin cancer across the world.

Many medical image databases are small and labeled, making it difficult to obtain large labeled datasets. The demand for a greater computing force is another drawback in effectively training CNN structures for medical image categorization [9]. Considering these challenges, incorporating pre-trained CNN models has been deemed a reasonable solution. These networks were previously trained and used huge images databases such as ImageNet, which contain ubiquitous properties that can be transferred to defined tasks through transfer learning (TL) [8]. When these prior constructs are aligned along with domain knowledge, investigators can build accurate classifiers without the need for relatively large observed samples. It is highly recommended to combine generalization ability with discriminative features estimated during the initial training stage to improve the performance of medical visual recognition systems that are characterized by relatively small amounts of training data and limited computational resources.

In their scoping review, Morid et al. identified 102 relevant studies on the implementation of TL in medical image examination. Emphasis on the general implementation of TL in medical image processing and the particular CNN configurations in various medical imaging applications. The study also found gaps that include the lack of a common valuation metric model and methodologies to increase the generalizability of TL modes in a variety of studies. The study also highlighted some of the limitations; for example, there were no clear measures to standardize the assessment criteria and methods that would help increase the consistency of TL strategies in previous studies [8].

In the study conducted by Medhat et al., a detailed comparative analysis of various deep CNN, including AlexNet, MobileNet V2, and ResNet50, was undertaken to evaluate their effectiveness in identifying skin cancer from dermatological images captured using a smartphone. The research highlighted that among these models, the architecture that achieved the highest diagnostic accuracy was based on a pre-trained AlexNet network. This superior performance was attributed to the application of TL techniques, which leveraged the pre-existing knowledge of the AlexNet model, and the use of basic data augmentation methods to enhance the training dataset. This combination proved to be particularly effective in improving the model's ability to accurately classify skin cancer in the images [10].

Therefore, there is still more potential to improve performance, which could be realized through more complex model fine-tuning approaches and better methods of augmentation. However, some comparative work completed using dermoscopy and external validation with other clinical centers would provide more robust data on these findings in other samples used in the current study. Extending the analysis in these directions may help improve the effectiveness of these applications in providing a generalized, accurate and reliable diagnosis of skin cancer using AI [10], [11], [12].

Some researchers like Venugopal et al. proposed a modified model in the EfficientNets family. The extracted features are passed through a classification layer model consisting of a pair of fully connected layers, the first containing 512 neurons and the second containing 256 neurons, as well as the output class layers for the classification of skin cancer [12]. This study did not fine-tune the pretrained models or even state the augmentation functions used in the data. Consequently, fine-tuning and detailed methods of data augmentation may affect the model in a way that alters the performance of the algorithm in improving overall efficiency and generalization on new datasets.

Similarly, Faghihi et al. used the preserved ImageNet weight of the VGG16 and VGG19 architectures to classify skin cancer. This study also modified the model by adding three initial layers of the pretrained AlexNet network [11] . However, this study did not perform any data augmentation. This points to the possibilities for future research to improve gender parity in data augmentation comprehensiveness.

Another study evaluated the influence of ML affecting CNN architecture for the distinction of benign and malignant breast lesions. CNN was used as feature extractor by Keerthana et al., who further put forward the following forms of hybrid CNN as comparative study. Their findings indicated that when the specified hybrid models from pre-trained CNN and ML are appropriately determined, the classification accuracy can be markedly increased. Unfortunately, due to the lack of detailed description of the training methodology used by the authors, the replicability of the study and the further improvement of the modeling used by the authors are somewhat constrained [12].

DenseNet-201, InceptionV3, MobileNet, ResNet50, and VGG16 to their superior performance in research conducted were selected in this study due to their superior performance as hybrid model with ML including Support vector machine (SVM), k-nearest neighbor (KNN), and decision tree [12]. In this research Random forest (RF) is chosen over decision tree due to its greater robustness and reliability [13]. All those pre-trained model also proven realiable for identifying skin cancer [8], [10], [11], [12], [14].

The primary contributions of this paper to the field of skin cancer identification using pre-trained CNN models are as follows. First, it modifies traditional pretrained CNN architectures, including DenseNet-201, InceptionV3, MobileNet, ResNet50, and VGG16, by adding additional layers to improve their performance. Second, this study examines whether these modified traditional pre-trained CNN models can also refine their performance as feature extractors, with the extracted features then given into ML classifiers such as SVM, KNN, and RF. To ensure the difference model performance are due to the models themselves and not due to variations in other factors, this study utilized consistent hyperparameters for all models because hyperparameters determine model's architecture and have a direct impact on model’s ability to learn [15].

Finally, this research undertakes a series of experiments to explore whether modifications to pre-trained CNN models can further enhance their performance when combined, particularly through the process of concatenating extracted features and feeding them into ML classifiers. Specifically, the study examines the potential improvements that can be achieved by hybridizing traditional pre-trained CNN models with additional techniques. By concatenating the features extracted by these CNN models and then using these combined features as inputs for various ML classifiers, the researchers aim to refine and optimize the classification process. The results of these experiments demonstrate that hybrid modified models, particularly the combination of MobileNet and VGG16 with a RF classifier, achieve a notably high accuracy rate of 94.20%. This finding indicates that the hybrid models perform optimally, significantly enhancing the diagnostic accuracy compared to the traditional pre-trained CNN models used alone.

## 2. Research Methods

Pre-trained CNN along with traditional ML classifiers fulfill a pivotal role in refining the accuracy of skin cancer classification. Initially, we implemented a comparative evaluation of various traditional pre-trained CNN with their modified models using the ISBI 2016 dataset to classify skin cancer lesions. Second, we used the extracted features from these pre-trained CNN with their modified models to train ML classifiers such as SVM, KNN, and RF, and then evaluated their performance. Finally, this study compared these hybrid models as feature extractors for ML classifiers.

### 2.1 Pre-trained CNN models

The proposed modified CNN model is implemented by adding a set of fully connected layers to a pre-trained CNN model using a TL process for identifying skin cancer in dermoscopic images. CNN training performance is improved with the TL technique, and image features are extracted more efficiently than CNN training from scratch [16]. The proposed CNN is created by adding layers to DenseNet201, InceptionV3, MobileNet, ResNet50, and VGG16. DenseNet-201, proposed by Huang et al. [17], is used because of its densely connected convolutional layers, which facilitate feature reuse, leading to efficient parameter usage and improved feature propagation in deep networks.

MobileNet, as introduced by Howard et al. [18], is chosen due to its lightweight architecture, making it suitable for mobile and embedded applications, while still achieving competitive accuracy in image classification tasks. InceptionV3, developed by Szegedy et al. [19], is used for its inception modules, which allow for the extraction of multiscale features through parallel convolutions, improving the model's capability to capture intricate features in the input data. ResNet-50, proposed by He et al. [20], is used due to its residual learning framework, which facilitates deep neural network training by reducing the problem of vanishing gradient, resulting in improved performance in image recognition tasks. VGG-16, introduced by Simonyan and Zisserman [21], employs its simplicity and effectiveness, with its uniform architecture consisting of small convolutional filters that facilitate deeper network architectures and yield strong performance in various computer vision tasks.

Several pre-trained CNN base models are believed to be the best feasible architectures to achieve accuracy on ImageNet [8]. These CNN architectures are used to construct a refined system for extracting features in the proposed model. The extracted feature is then added to a neural classifier model that includes a global average pooling and a batch normalization layer to resolve the vanishing gradient [22], [23] and a dense layer of 128 neurons to resolve overfitting [7], [24], followed by a binary classification output layer for benign and malignant, as presented in the process stages of Figure 1.

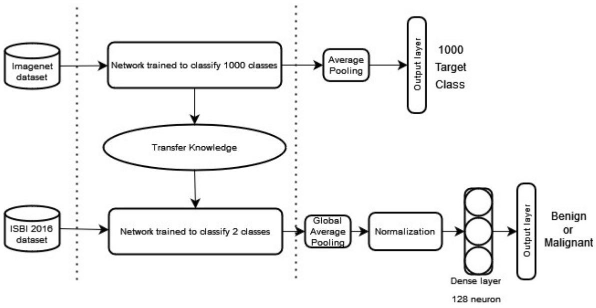


Figure 1. TL and the proposed model framework

In TL, the learning weights and biases acquired from the model trained with the ImageNet dataset are used to extract the required image features [25]. The model is then fine-tuned and trained on skin lesions of two classes, namely benign and malignant. Using individual images of skin lesions aids in the capture of more detailed features of the dermoscopy images by the training model. The training and validation method in this research uses cross-validation and early stop to avoid overfitting the model [11]. The hyperparameters used in the proposed models are presented in Table 1.

Table 1. Hyperparameter setup used in the proposed CNN

| Hyperparameter | Values |
|---|---|
| Batch size | 30 [12] |
| Image input size | 224 x 224 [12] |
| Learning rate | 0.0001 [12] |
| Max epoch | 25 [12] |
| Optimizer | Adam [7] |
| Training validation split | 80:20 [7] |

### *2.2 Pre-trained CNN with ML classifier*

The features of the dataset are extracted through pre-trained CNN and then are given to ML classifier [17]. The skin cancer classification framework is presented in Figure 2 with a pre-trained CNN and a ML classifier. The pre-trained CNN utilized in this paper are DenseNet201, InceptionV3, MobileNet, ResNet50, and VGG16. SVM, KNN, and RF are several ML classifiers used in this study[12], [26]. SVM, pioneered by Cortes and Vapnik [27], is selected for its robustness in managing high-dimensional data and its ability to identify the best hyperplane that maximizes the margin between classes, making it suitable for image classification tasks with nonlinear decision boundaries. KNN, originally proposed by Cover and Hart [28], is utilized for its simplicity and robustness in classification tasks. The classification of an input instance is decided by the most class among its k nearest neighbors in the feature space. RF, developed by Breiman [29], is used for its ensemble learning technique, which combines multiple decision trees to mitigate overfitting and improve generalization performance, making it suitable for robust image classification tasks. The evaluation parameters applied include accuracy, precision, sensitivity, specificity and the F1 score in the different models.

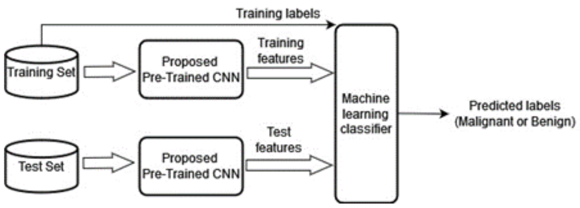


Figure 2. ML classifier framework

### 2.3. Hybrid models

Using multiple pre-trained networks for skin classification, as much as the pre-trained CNN can be used individually, there is the possibility of using multiple pre-trained CNN to arrive at a better classification [12], [30]. In this method, it is possible to combine various pre-trained models and select important features that are further passed onto the ML classifier, where the final desired outcome in terms of classification is derived. The features acquired from the pre-trained CNN are then concatenated and provided to the ML classifier, such as SVM, KNN, and RF. Finally, the output of the trained ML classifier gives the classes of the dermoscopy image input. The framework of the hybrid model is presented in Figure 3.

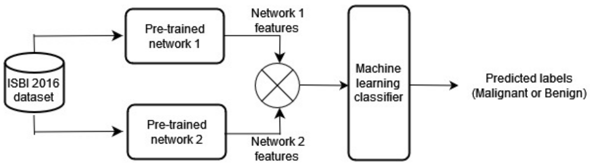


Figure 3. Hybrid model framework

The proposed models are evaluated in terms of their performance using the ISBI 2016 dataset. The hybrid models will be applied to traditional and modified pre-trained CNN structures, including MobileNet and DenseNet201, MobileNet and ResNet50, as well as MobileNet and VGG16. The performance of the model is measured by precision, precision, sensitivity, specificity and F1 score.

### *2.4 Datasets*

A subset of the ISIC dataset, the ISBI 2016 dataset [31], is used to train and test the pre-trained CNN models. The training set has 900 images, and the test set contains 379 images with image sizes ranging from 1022 x 767 to 4288 x 284 pixels. For training the models, we used training images consisting of 727 benign lesions, the remaining 173 malignant lesions. For the test, we used test images consisting of 304 benign lesions; the remaining 75 images were malignant lesions. Furthermore, for computational efficiency purposes, the training and testing images were rescaled to 224 x 224 pixels.

To overcome the problems related to unbalanced data, several image enhancement methods, including flipping, cropping, and rotation techniques, were used in the training images. Training images are flipped around the vertical and horizontal axes. Then an angle of 20 ° is used to rotate the images. In this study, 727 benign and 173 malignant images were each augmented to 1000 images, resulting in balanced training data for the proposed models.

### *2.5 Performance Evaluation*

This research aims to identify the appropriate evaluation metrics to determine how effective the proposed models are in categorizing skin cancer into benign and malignant. The assessment includes the application of a confusion matrix along with other essential metrics, including accuracy [32], precision [32], sensitivity [32], specificity [32], and the F1-score [33]. The formula to calculate all these metrics can be seen in Formulas 1, 2, 3, 4, and 5.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity = \frac{TN}{FP + TN} \quad (4)$$

$$F1 - score = \frac{2 \times Pre \times Recall}{Pre + Recall} \quad (5)$$

The confusion matrix gives a clear understanding of classified data, distinguishing between correct and incorrect classifications [34]. Since accuracy is crucial, it measures the extent of correct classification out of the total instances. Precision quantifies the size of accurately identified benign cases among all genuine benign cases classified by the system. Sensitivity, on the other hand, is modeled as the size of genuine benign cases to the total instances that are accurately classified as benign by the algorithm. Specificity is modeled as the size of genuine malignant cases to the count of instances that the algorithm accurately identified as malignant.

The F1 score metric, the result of combining precision and sensitivity, offers a unique way to measure the performance of a model. In this case, positives mean that the skin had benign cancer and negatives indicate that it was malignant. The number of accurately classified instances is termed true positives (TP) and true negatives (TN), while false negatives (FN) and false positives (FP) are those that are misclassified.

## 3. Results and Discussions

This section delves into the exploration of various methodologies employed for skin cancer classification, detailing the datasets utilized and key performance metrics assessed. The findings are categorized into three distinct sections for clarity and depth of analysis.

Section 3 focuses on the outcomes derived from several pre-trained CNN. Within this section, Section 3.1 provides an overview of the results obtained from skin cancer classification utilizing individual pre-trained CNN models. These models were evaluated based on their ability to accurately differentiate between malignant and benign skin lesions, using established benchmark datasets.

Section 3.2 delves into the outcomes of employing pre-trained CNN models in conjunction with ML classifiers for skin cancer classification. This hybrid approach aims to capitalize on the feature extraction capabilities of CNN while leveraging the robust classification capabilities of ML algorithms such as SVM, KNN, and RF. The results presented in this section highlight the combined efficacy of CNN and ML classifiers in enhancing diagnostic accuracy.

In Section 3.3 of this study, the spotlight turns to hybrid models that capitalize on the synergy between features extracted from two distinct pre-trained CNN architectures, which are then processed through ML classifiers. This innovative approach aims to elevate classification accuracy by harnessing the complementary strengths of diverse CNN and refining their performance through strategic model fusion strategies. The thorough exploration conducted in this section sheds light on the superiority of these hybrid models over conventional methods. By effectively integrating the feature extraction capabilities of multiple CNN and leveraging the discriminative power of ML classifiers like SVM, KNN, and RF, these hybrid models demonstrate significant advancements in the field of skin cancer diagnosis. The detailed analysis presented underscores their potential to enhance diagnostic precision and reliability in medical image analysis, paving the way for more effective and efficient clinical decision-making processes.

### *3.1 Results of skin cancer classification using pre-trained CNN*

The training of pre-trained CNN models in this study adheres to a standardized approach, employing uniform settings across batch size, image input size, learning rate, epochs, and optimizer configurations as outlined in Table 1. The ISBI 2016 training dataset undergoes an 80-20 split for training and validation, respectively, within each of the five folds utilized for cross-validation. Given the constraints of data scarcity and class imbalance between benign and malignant skin lesions, augmentation techniques such as rotation, scaling, and translation are applied to augment all training images. The ISBI 2016 test dataset serves as the benchmark for evaluating model performance.

Results for traditional pre-trained CNN models are compiled and presented in Table 2, highlighting their respective accuracies. Concurrently, modified versions of these pre-trained CNN are evaluated and their performance metrics are detailed in Table 3. Notably, nearly all modified pre-trained models exhibit improved accuracy rates, with the exception of the modified InceptionV3 model, which achieves a slightly lower accuracy of 79.16% compared to the traditional InceptionV3 model's accuracy of 81.53%. Among the modified models, VGG16 stands out with the highest accuracy recorded at 85.22%.

To provide a visual representation of the performance metrics, confusion matrices are depicted. Figures 4 and 5 illustrate the confusion matrices for traditional and modified pre-trained CNN models, respectively. These matrices offer a comprehensive breakdown of classification outcomes, shedding light on the strengths and weaknesses of each model configuration in accurately distinguishing between benign and malignant skin lesions.

### *3.2 Results of skin cancer classification using pre-trained CNN models with ML classifier*

In this study, a thorough and systematic approach was adopted to train and evaluate various models using the ISBI 2016 dataset. Initially, 80% of the images from the ISBI 2016 training dataset were employed for model training. The remaining 20% of the images were set aside for testing, utilizing cross-validation techniques to ensure a robust assessment of the models' performance. Beyond this, the ISBI 2016 test dataset was also used to further gauge the effectiveness and reliability of the models.

Table 2. Performance of the pre-trained CNN

| Pre-trained CNN | Accuracy | Precision | Sensitivity | Specificity | F1-score |
|---|---|---|---|---|---|
| MobileNet | 81.79% | 86.83% | 91.12% | 44.00% | 81.11% |
| DenseNet201 | 80.74% | 86.46% | 92.43% | 41.33% | 82.28% |
| ResNet50 | 64.37% | 80.21% | 76.64% | 14.67% | 68.23% |
| InceptionV3 | 81.53% | 87.10% | 88.82% | 46.67% | 79.06% |
| VGG16 | 69.12% | 87.97% | 84.21% | 53.33% | 74.96% |

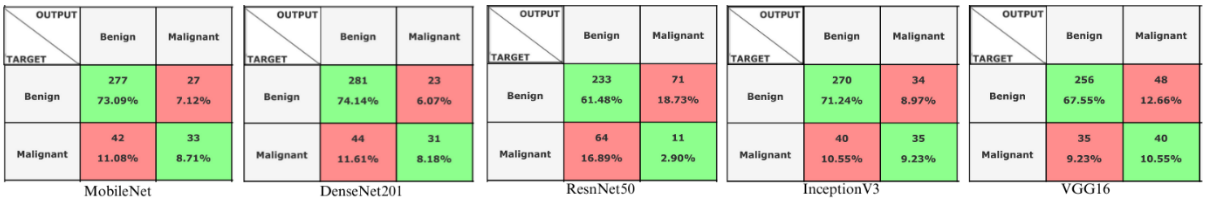


Figure 4. Confusion matrices for the traditional pre-trained CNN

Table 3. Performance of the modified pre-trained CNN

| Pre-trained CNN | Accuracy | Precision | Sensitivity | Specificity | F1-score |
|---|---|---|---|---|---|
| MobileNet | 84.90% | 89.46% | 92.11% | 56.00% | 81.99% |
| DenseNet201 | 82.85% | 86.54% | 93.09% | 41.33% | 82.87% |
| ResNet50 | 82.56% | 87.42% | 91.45% | 46.67% | 81.41% |
| InceptionV3 | 79.16% | 86.41% | 87.83% | 44.00% | 78.18% |
| VGG16 | 85.22% | 84.44% | 100.00% | 25.33% | 89.02% |

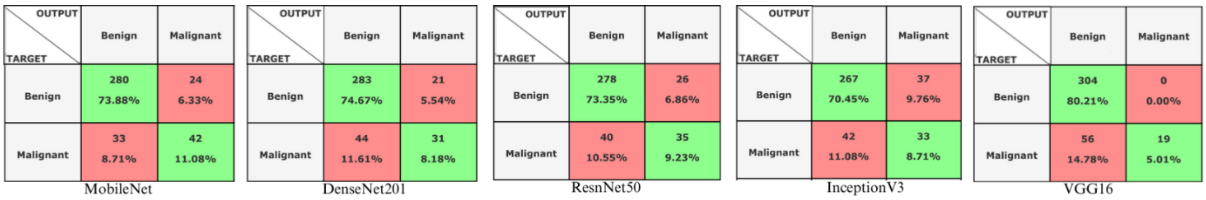


Figure 5. Confusion matrices for the modified pre-trained CNN

To provide a detailed analysis of the modified pre-trained models, confusion matrices were generated. These matrices offer an in-depth examination of the models' performance metrics, presenting the results for traditional pre-trained CNN in Figure 6 and those for the modified pre-trained CNN in Figure 7. The comprehensive performance metrics for all pre-trained CNN are summarized in Tables 4 and 5, providing a clear comparison of their effectiveness.Among the various models tested, the modified MobileNet, when combined with an SVM classifier, achieved the highest accuracy. This notable finding underscores the superior performance of the modified MobileNet model in comparison to the other models evaluated. It highlights the significant benefits of modifying and optimizing pre-trained CNN to enhance accuracy in skin lesion classification. This study demonstrates the potential for advanced CNN techniques to substantially improve diagnostic accuracy in medical image analysis, paving the way for more reliable and efficient skin cancer detection methods

Table 4. Performance of the traditional pre-trained CNN with ML classifier

| Pre-trained CNN | Classifier | Accuracy | Precision | Sensitivity | Specificity | F1-score |
|---|---|---|---|---|---|---|
| MobileNet | SVM | 83,11% | 88,22% | 91,12% | 50,67% | 89,64% |
| | KNN | 81,79% | 87,30% | 90,46% | 46,67% | 88,85% |
| | RF | 82,32% | 88,85% | 89,14% | 54,67% | 89,00% |
| DenseNet201 | SVM | 82,59% | 87,42% | 91,45% | 46,67% | 89,39% |
| | KNN | 80,47% | 88,33% | 87,17% | 53,33% | 87,75% |
| | RF | 81,53% | 89,00% | 87,83% | 56,00% | 88,41% |
| ResNet50 | SVM | 82,18% | 87,99% | 89,14% | 50,67% | 88,56% |
| | KNN | 76,78% | 85,06% | 86,18% | 38,67% | 85,62% |
| | RF | 82,06% | 88,82% | 88,82% | 54,67% | 88,82% |
| InceptionV3 | SVM | 83,91% | 87,16% | 93,75% | 44,00% | 90,33% |
| | KNN | 84,17% | 86,09% | 95,72% | 37,33% | 90,65% |
| | RF | 83,91% | 87,38% | 93,42% | 45,33% | 90,30% |
| VGG16 | SVM | 70,98% | 87,31% | 74,67% | 56,00% | 80,50% |
| | KNN | 68,34% | 83,58% | 75,33% | 40,00% | 79,24% |
| | RF | 69,31% | 82,58% | 78,22% | 33,33% | 80,34% |


Figure 6. Confusion matrices for the traditional pre-trained CNN

Table 5. Performance of the modified pre-trained CNN with ML classifier

| Pre-trained CNN | Classifier | Accuracy | Precision | Sensitivity | Specificity | F1-score |
|---|---|---|---|---|---|---|
| MobileNet | SVM | 83,11% | 88,71% | 90,46% | 53,33% | 89,58% |
| | KNN | 85,49% | 88,79% | 93,75% | 52,00% | 91,20% |
| | RF | 84,43% | 89,39% | 91,45% | 56,00% | 90,41% |
| DenseNet201 | SVM | 82,06% | 88,82% | 88,82% | 54,67% | 88,82% |
| | KNN | 84,43% | 88,16% | 93,09% | 49,33% | 90,56% |
| | RF | 83,11% | 87,97% | 91,45% | 49,33% | 89,68% |
| ResNet50 | SVM | 82,32% | 88,85% | 89,14% | 54,67% | 89,00% |
| | KNN | 82,85% | 87,46% | 91,78% | 46,67% | 89,57% |
| | RF | 82,32% | 87,62% | 90,79% | 48,00% | 89,18% |
| InceptionV3 | SVM | 78,36% | 85,58% | 87,83% | 40,00% | 86,69% |
| | KNN | 80,47% | 86,86% | 89,14% | 45,33% | 87,99% |
| | RF | 81,27% | 86,07% | 91,45% | 40,00% | 88,68% |
| VGG16 | SVM | 72,30% | 90,28% | 73,36% | 68,00% | 80,94% |
| | KNN | 72,56% | 82,68% | 83,22% | 29,33% | 82,95% |
| | RF | 68,34% | 82,39% | 76,97% | 33,33% | 79,59% |


Figure 7. Confusion matrices for the modified pre-trained CNN

*3.3 Results of skin cancer classification using hybrid models*

In this study, a comprehensive approach was taken to enhance the accuracy of skin lesion classification by combining the strengths of multiple CNN architectures. The extracted features from images processed by two pre-trained CNN were concatenated and then used as input for various ML classifiers, including SVM, KNN, and RF. This method aimed to exploit the complementary features captured by different CNN to improve overall classification performance. The performance evaluations of these concatenated models are detailed in Tables 6 and 7, covering combinations such as MobileNet and DenseNet201, MobileNet and ResNet50, and MobileNet and VGG16. Each combination was assessed to determine its effectiveness in classifying skin lesions as malignant or benign. Among the various hybrid models tested, the combination of MobileNet and VGG16, paired with the RF classifier, achieved the highest accuracy of 94.20%. This optimal performance was attained by using the hyperparameters specified in Table 1, which were meticulously tuned to enhance the model's predictive capability.

To provide a thorough analysis of these hybrid models' effectiveness, confusion matrices were generated. Figure 8 presents the confusion matrices for the traditional pre-trained hybrid models, while Figure 9 displays the results for the modified pre-trained hybrid models. These matrices offer a detailed breakdown of classification accuracy and error distribution, clearly illustrating the superior performance of the modified hybrid models compared to their traditional counterparts.

This methodology underscores the significant potential of combining feature extraction from multiple CNN with advanced ML classifiers. By leveraging the unique strengths of each CNN architecture and optimizing the model through concatenation and careful hyperparameter tuning, the study demonstrates a promising approach to significantly improve the accuracy of skin lesion classification in medical image analysis

Table 6. Performance of hybrid traditional pre-trained CNN

| Pre-trained CNN | Classifier | Accuracy | Precision | Sensitivity | Specificity | F1-score |
|---|---|---|---|---|---|---|
| MobileNet and ResNet50 | SVM | 82.85% | 87.70% | 91.45% | 48.00% | 89.53% |
| | KNN | 82.06% | 88.31% | 89.47% | 52.00% | 88.89% |
| | RF | 86.54% | 89.66% | 94.08% | 56.00% | 91.81% |
| MobileNet and DenseNet201 | SVM | 85.22% | 88.04% | 94.41% | 48.00% | 91.11% |
| | KNN | 84.96% | 88.47% | 93.42% | 50.67% | 90.88% |
| | RF | 86.28% | 89.87% | 93.42% | 57.33% | 91.61% |
| MobileNet and VGG16 | SVM | 83.11% | 87.74% | 91.78% | 48.00% | 89.71% |
| | KNN | 83.11% | 88.46% | 90.79% | 52.00% | 89.61% |
| | RF | 86.28% | 89.87% | 93.42% | 57.33% | 91.61% |

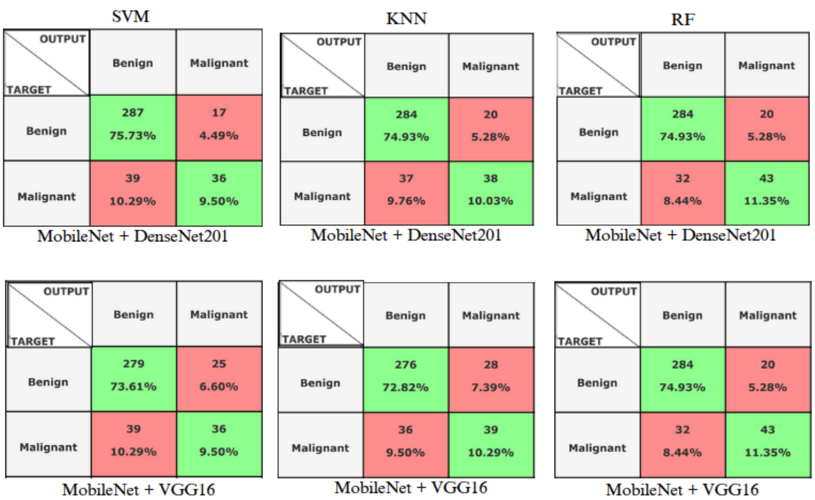


Figure 8. Confusion matrices for hybrid traditional pre-trained CNN

Table 7. Performance of a hybrid modified pre-trained CNN

| Pre-trained CNN | Classifier | Accuracy | Precision | Sensitivity | Specificity | F1-score |
|---|---|---|---|---|---|---|
| MobileNet and ResNet50 | SVM | 87.60% | 90.54% | 94.41% | 60.00% | 92.43% |
| | KNN | 87.86% | 89.81% | 95.72% | 56.00% | 92.68% |
| | RF | 92.35% | 92.05% | 99.01% | 65.33% | 95.40% |
| MobileNet and DenseNet201 | SVM | 87.34% | 88.79% | 96.38% | 50.67% | 92.43% |
| | KNN | 86.28% | 89.13% | 94.41% | 53.33% | 91.69% |
| | RF | 90.24% | 91.59% | 96.71% | 64.00% | 94.08% |
| MobileNet and VGG16 | SVM | 91.03% | 91.93% | 97.37% | 65.33% | 94.57% |
| | KNN | 85.49% | 89.78% | 92.43% | 57.33% | 91.09% |
| | RF | 94.20% | 94.06% | 99.01% | 74.67% | 96.47% |

 

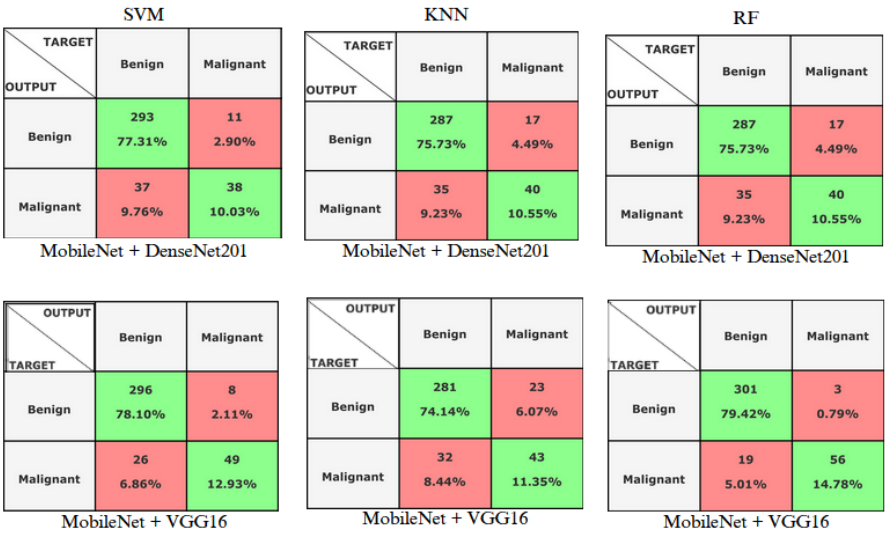


Figure 9. Confusion matrices for hybrid modified pre-trained CNN

## 4. Conclusions

This study delves deeply into the realm of skin cancer image classification, proposing novel methods to enhance the performance of traditional pre-trained CNN models for more effective classification tasks. The research systematically compares the efficacy of standard pre-trained CNN against modified versions in the context of skin cancer image classification. Notably, the modified VGG16 model achieves the highest accuracy among all models tested, reaching 85.22%, highlighting its superior performance in distinguishing between benign and malignant skin lesions. Expanding upon this initial investigation, the study explores the integration of alternative classifiers such as SVM, KNN, and RF The most precise classification result emerges from the modified MobileNet model coupled with the KNN classifier, achieving an impressive accuracy of 85.49%. Further advancing the research, a hybrid model approach is introduced, combining features extracted from various modified pre-trained CNN architectures and processing them through ML classifiers. Particularly noteworthy is the concatenation of MobileNet and VGG16 with an RF classifier, yielding a remarkable accuracy of 94.20%, underscoring the effectiveness of hybrid models in optimizing classification performance. Looking forward, future endeavors could explore additional dimensions of optimization for the proposed pre-trained CNN models, including variations in epochs, batch sizes, classifiers, and optimization strategies. Furthermore, integrating preprocessing steps into the input stream before network processing could further refine the proposed method's ability to classify a broader spectrum of skin lesion classes with enhanced accuracy. These avenues of exploration hold promise for advancing the capabilities of CNN-based diagnostic systems in dermatology, ultimately improving clinical decision-making and patient outcomes.

## Acknowledgments

We declare no conflict of interest and acknowledge the support of the Institut Teknologi Sepuluh Nopember.

## References

[1] N. Hasan *et al.*, "Skin cancer: understanding the journey of transformation from conventional to advanced treatment approaches," *Molecular Cancer*, vol. 22, no. 1. BioMed Central Ltd, Dec. 01, 2023. doi: 10.1186/s12943-023-01854-3.

[2] A. Deshpande, V.V. Estrela, and R. N, *Computational Intelligence Methods for Super-Resolution in Image Processing Applications*. Cham: Springer International Publishing, 2021. doi: 10.1007/978-3-030-67921-7.

[3] G. M. Abd-Allah *et al.*, "miRNAs as potential game-changers in melanoma: A comprehensive review," *Pathology Research and Practice*, vol. 244. Elsevier GmbH, Apr. 01, 2023. doi: 10.1016/j.prp.2023.154424.

[4] K. Urban, S. Mehrmal, P. Uppal, R. L. Giesey, and G. R. Delost, "The global burden of skin cancer: A longitudinal analysis from the Global Burden of Disease Study, 1990–2017," *JAAD Int*, vol. 2, pp. 98–108, Mar. 2021, doi: 10.1016/j.jdin.2020.10.013.

[5] K. Shi *et al.*, "Clinical characteristics of malignant melanoma in central China and predictors of metastasis," *Oncol Lett*, Dec. 2019, doi: 10.3892/ol.2019.11219.

[6] R. K. Singh, R. Gorantla, S. G. R. Allada, and P. Narra, "SkiNet: A deep learning framework for skin lesion diagnosis with uncertainty estimation and explainability," *PLoS One*, vol. 17, no. 10, p. e0276836, Oct. 2022, doi: 10.1371/journal.pone.0276836.

[7] K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, "Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer," *Neuroscience Informatics*, vol. 2, no. 4, p. 100034, Dec. 2022, doi: 10.1016/j.neuri.2021.100034.

[8] M. A. Morid, A. Borjali, and G. Del Fiol, "A scoping review of transfer learning research on medical image analysis using ImageNet," *Computers in Biology and Medicine*, vol. 128. Elsevier Ltd, Jan. 01, 2021. doi: 10.1016/j.compbiomed.2020.104115.

[9] A. Takiddin, J. Schneider, Y. Yang, A. Abd-Alrazaq, and M. Househ, "Artificial intelligence for skin cancer detection: Scoping review," *Journal of Medical Internet Research*, vol. 23, no. 11. JMIR Publications Inc., Nov. 01, 2021. doi: 10.2196/22934.

[10] S. Medhat, H. Abdel-Galil, A. E. Aboutabl, and H. Saleh, "Skin cancer diagnosis using convolutional neural networks for smartphone images: A comparative study," *J Radiat Res*

*Appl Sci*, vol. 15, no. 1, pp. 262–267, Mar. 2022, doi: 10.1016/j.jrras.2022.03.008.

[11] A. Faghihi, M. Fathollahi, and R. Rajabi, "Diagnosis of skin cancer using VGG16 and VGG19 based transfer learning models," *Multimed Tools Appl*, Jun. 2023, doi: 10.1007/s11042-023-17735-2.

[12] D. Keerthana, V. Venugopal, M. K. Nath, and M. Mishra, "Hybrid convolutional neural networks with SVM classifier for classification of skin cancer," *Biomedical Engineering Advances*, vol. 5, p. 100069, Jun. 2023, doi: 10.1016/j.bea.2022.100069.

[13] X. Ma, J. Shan, F. Ning, W. Li, and H. Li, "EFFNet: A skin cancer classification model based on feature fusion and random forests," *PLoS One*, vol. 18, no. 10, p. e0293266, Oct. 2023, doi: 10.1371/journal.pone.0293266.

[14] V. Venugopal, N. I. Raj, M. K. Nath, and N. Stephen, "A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images," *Decision Analytics Journal*, vol. 8, Sep. 2023, doi: 10.1016/j.dajour.2023.100278.

[15] A. Panesar, "Evaluating Machine Learning Models," in *Machine Learning and AI for Healthcare*, Berkeley, CA: Apress, 2021, pp. 189–205. doi: 10.1007/978-1-4842-6537-6_7.

[16] A. W. Salehi *et al.*, "A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope," *Sustainability (Switzerland)*, vol. 15, no. 7. MDPI, Apr. 01, 2023. doi: 10.3390/su15075930.

[17] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 2261–2269. doi: 10.1109/CVPR.2017.243.

[18] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 2017.

[19] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 2818–2826. doi: 10.1109/CVPR.2016.308.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[21] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2014.

[22] P. V. S. P. Raghavendra, C. Charitha, K. G. Begum, and V. B. S. Prasath, "Deep Learning–Based Skin Lesion Multi-class Classification with Global Average Pooling Improvement," *J Digit Imaging*, vol. 36, no. 5, pp. 2227–2248, Jul. 2023, doi: 10.1007/s10278-023-00862-5.

[23] C. Garbin, X. Zhu, and O. Marques, "Dropout vs. batch normalization: an empirical study of their impact to deep learning," *Multimed Tools Appl*, vol. 79, no. 19–20, pp. 12777–12815, May 2020, doi: 10.1007/s11042-019-08453-9.

[24] K. Mridha, Md. 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, doi: 10.1109/ACCESS.2023.3269694.

[25] V. Venugopal, N. I. Raj, M. K. Nath, and N. Stephen, "A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images," *Decision Analytics Journal*, vol. 8, Sep. 2023, doi: 10.1016/j.dajour.2023.100278.

[26] X. Ma, J. Shan, F. Ning, W. Li, and H. Li, "EFFNet: A skin cancer classification model based on feature fusion and random forests," *PLoS One*, vol. 18, no. 10 October, Oct. 2023, doi: 10.1371/journal.pone.0293266.

[27] C. Cortes and V. Vapnik, "Support-vector networks," *Mach Learn*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[28] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans Inf Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967, doi: 10.1109/TIT.1967.1053964.

[29] L. Breiman, "Random forests," *Mach Learn*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[30] S. Abbas *et al.*, "BCD-WERT: a novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm," *PeerJ Comput Sci*, vol. 7, p. e390, Mar. 2021, doi: 10.7717/peerj-cs.390.

[31] D. Gutman *et al.*, "Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC)." [Online]. Available: https://isic-archive.com/

[32] M. M. Mijwil, "Skin cancer disease images classification using deep learning solutions," *Multimed Tools Appl*, vol. 80, no. 17, pp. 26255–26271, Jul. 2021, doi: 10.1007/s11042-021-10952-7.

[33] S. S. Chaturvedi, J. V. Tembhurne, and T. Diwan, "A multi-class skin Cancer classification using deep convolutional neural networks," *Multimed Tools Appl*, vol. 79, no. 39–40, pp. 28477–28498, Oct. 2020, doi: 10.1007/s11042-020-09388-2.

[34] J. Xu, Y. Zhang, and D. Miao, "Three-way confusion matrix for classification: A measure driven view," *Inf Sci (N Y)*, vol. 507, pp. 772–794, Jan. 2020, doi: 10.1016/j.ins.2019.06.064.