Bulletin of Electrical Engineering and Informatics
Vol. 13, No. 4, August 2024, pp. 2716~2728
ISSN: 2302-9285, DOI: 10.11591/eei.v13i4.7159

Skin cancer classification using EfficientNet architecture

Mawaddah Harahap, Amir Mahmud Husein, Shane Christian Kwok, Vincent Wizley, Jocelyn Leonardi, Derrick Kenji Ong, Deskianta Ginting, Benny Art Silitonga
Faculty of Science and Technology, Universitas Prima Indonesia, Medan, Indonesia

Article Info

Article history:

Received Jul 7, 2023
Revised Oct 26, 2023
Accepted Feb 24, 2024

Keywords:

Computer vision
Convolutional neural network
EfficientNet
Image classification
Skin cancer

ABSTRACT

Skin cancer is one of the most common deadly diseases worldwide. Hence, skin cancer classification is becoming increasingly important because treatment in the early stages of skin cancer is much more effective and efficient. This study focuses on the classification of three common types of skin cancer, namely basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and melanoma using EfficientNet architecture. The dataset is preprocessed and each image in the dataset is resized to 256×256 pixels prior to incorporation in later stages. We then train all types of EfficientNet starting from EfficientNet-B0 to EfficientNet-B7 and compare their performances. Based on the test results, all trained EfficientNet models are capable of producing good accuracy, precision, recall, and F1-score in skin cancer classification. Particularly, our designed EfficientNet-B4 model achieves 79.69% accuracy, 81.67% precision, 76.56% recall, and 79.03% F1-score as the highest among others. These results confirm that EfficientNet architecture can be utilized to classify skin cancer properly.

Corresponding Author:

Mawaddah Harahap
Faculty of Science and Technology, Universitas Prima Indonesia
Medan, Indonesia
Email: mawaddah@unprimdn.ac.id

1. INTRODUCTION

Skin cancer is the corrupted cells that grew at abnormal pace in skin tissue and is often caused by overexposure to ultraviolet (UV) rays [1]–[6]. Skin cancer is without a doubt, one of the most prevalent cancers worldwide [1], [3], [7]–[9]. Based on data taken from the World Health Organization (WHO) [2], [7], more than 1.5 million people were affected worldwide in 2020, in which around 1.2 million cases of non-melanoma skin cancer and 325,000 cases of melanoma skin cancer were reported. Moreover, there were approximately 64,000 people died from non-melanoma skin cancer and 57,000 people died from melanoma in 2020 [2], [7]. The three most common skin cancer types are basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and melanoma [3], [8], [9]. Both BCC and SCC are categorized as primary non-melanoma skin cancers and are the most commonly reported skin cancers among other types [3], [4], [8], [9]. Despite that, melanoma is more deadly because it has a higher mortality rate than BCC and SCC [3], [9], [10].

If treated early, all skin cancers have a potential to be cured. Therefore, early diagnosis or detection is essential for effective and efficient treatment [5], [6], [9]–[11]. Unfortunately, diagnosing skin cancer in the early stages by using regular methods can be really challenging and prone to error [10], [11]. This issue shows that the utilization of a computer vision system in this field can be really helpful for professionals or even normal people to identify skin cancer.

Computer vision is a subset of artificial intelligence (AI) that is becoming more popular and impactful especially in dermatology [5], [6], [12]. Convolutional neural network (CNN) is a computer vision method that utilizes deep learning concept to be able to imitate visual cortex. Hence, CNN can offer a

solution for detecting and classifying skin cancer. CNN has been proven to produce excellent performance in various fields, especially object detection, image classification, and natural language processing (NLP) [13]–[15]. Currently, there are many CNN architectures, one of which is EfficientNet. The EfficientNet architecture itself also has many types starting from EfficientNet-B0 to EfficientNet-B7. EfficientNet architecture has been proven capable of producing better performance compared with several other architectures [16]–[18].

There are several previous studies related to skin cancer classification. Ali *et al.* [19] proposed a deep convolutional neural network (DCNN) model for benign and malignant skin cancers classification task. The model was evaluated on HAM10000 dataset and it produced an accuracy of 91.93%. The performance of the model was better than the performances of several transfer learning models on the same dataset. Patil and Bellary [20] used CNN approach to build a stage classification system of melanoma skin cancer. The CNN used similarity measure for text processing (SMTP) as the loss function. This method successfully achieved 96% accuracy in classifying melanoma stage 1, stage 2, and stage 3. The designed CNN model achieved 92.64% accuracy. Meanwhile, Sujaini *et al.* [6] compared linear regression model and deep learning model in melanoma detection. The results proved that the deep learning CNN model is better than the linear regression model with higher accuracy and lower false negative rate in detecting melanoma.

Gururaj *et al.* [11] also utilized CNN for skin cancer classification. The MNIST: HAM10000 dataset was filtered and preprocessed with techniques like sampling, segmentation using autoencoder and decoder, and dull razor. After preprocessing, DenseNet169 model and Resnet50 model were trained. The testing results showed that DenseNet169 with under sampling technique produced 91.2% accuracy and 91.7% F1-score, whereas Resnet50 with oversampling technique produced 83% accuracy and 84% F1-score. Xiao and Wu [21] designed Global-DNN and Global-Local models to classify skin cancer into melanoma and seborrhoeic keratosis. They integrated local binary pattern (LBP) features with deep convolutional features and utilized visual saliency detection to be able to effectively perform classification and eliminate background interference. The proposed models were tested on ISIC-2017 skin cancer dataset. The Global-DNN model achieved 85.8% accuracy in recognizing melanoma and 91.7% accuracy in recognizing seborrhoeic keratosis. The Global-Local model achieved 84.8% accuracy in recognizing melanoma and 91.3% accuracy in recognizing seborrhoeic keratosis.

Daghrir *et al.* [10] proposed a hybrid approach using deep learning and machine learning techniques for melanoma detection. Each skin cancer image was preprocessed by removing hair lines, segmentating the lesion area, and extracting features of the lesion. This study combined the prediction results of KNN, SVM, and CNN by using majority voting to identify melanoma skin cancer. The final result of the majority voting showed an accuracy of 88.4%. Thurnhofer-Hemsi and Domínguez [22] presented a deep learning framework for skin cancer classification using five state-of-art CNN models, which were GoogLeNet, InceptionV3, DenseNet201, Inception-ResNetV2, and MobileNetV2. For each model, plain classifier and hierarchical classifier were implemented. Plain classifier was used to directly classify the input within one of the 7 skin cancer classes, whereas hierarchical classifier was used to distinguish nevi class from the rest and then classify the other 6 classes. The experiments showed that DenseNet201 model with plain classifier was better than the rest of the models in classifying skin cancer.

Although BCC and SCC are the major types of non-melanoma skin cancer [3], [4], [8], most of the previous studies only focus on classifying melanoma and minor types of skin cancer. Therefore, the main contribution of this study is twofold: first, the classification of the major types of skin cancer, which are BCC, SCC, and melanoma (MEL) by using deep learning method; and second, the designed EfficientNet models that can properly perform skin cancer classification. We also focus on comparing all eight types of EfficientNet architecture in classifying BCC, SCC, and MEL. All EfficientNet models from EfficientNet-B0 to EfficientNet-B7 are designed and trained using preprocessed skin cancer dataset. Each model is then given test images and asked to classify them. The performance metrics of each model are then used for comparison. Our goal is to design and find the best EfficientNet model for skin cancer classification task.

2. METHOD

We use Python programming language on Google Colaboratory to conduct our experiments. Google Colaboratory is used because it offers free cloud-based access to a Jupyter Notebook-style interface which allows us to write, run, and share Python code seamlessly. There are several libraries to support our experiments, such as Keras for building and training neural networks; Pandas for data manipulation and analysis; Numpy which provides support for multi-dimensional arrays and mathematical functions; Matplotlib for visualizing data, model performance, and results; and Scikit-Learn for tasks such as data preprocessing, model selection, and evaluation. We also enable GPU in Google Colaboratory to accelerate training processes and inference processes. This comprehensive setup allows us to explore, develop, and evaluate models effectively.

2.1. Data acquisition

In this study, images of cancer on human skin are used as the dataset. The dataset is sourced from Kaggle and is provided by Maharana [23]. The original dataset was used for ISIC 2019 challenge. Larxel provided the original dataset on Kaggle and Maharana retransformed that dataset so that it is easier to use.

The dataset contains 25,331 images belonging to 8 classes for multiclass classification problem. However, in this study, only 3 classes are used: BCC, SCC, and MEL. Samples of each selected class are shown in Figure 1. BCC appears as a small, translucent bump or nodule on the skin. It may have a pearly, waxy texture and often develops on sun-exposed areas like face, ears, and neck. SCC often presents as a scaly red patch, open sore, or a raised growth with a central depression. MEL often appears as an irregularly shaped mole or dark spot with uneven coloration.

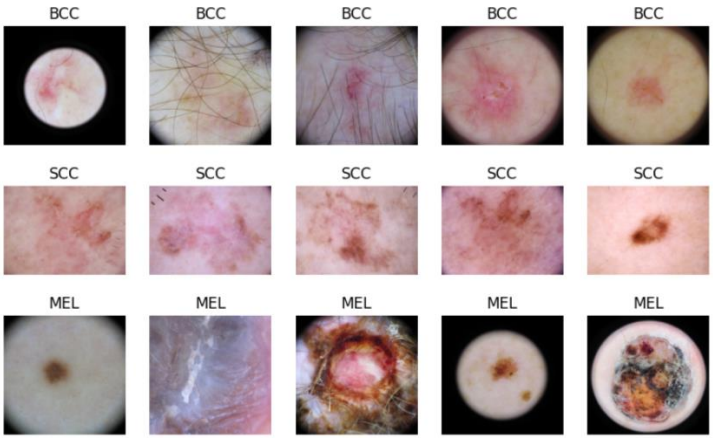

Figure 1. Samples of each selected class from the dataset

2.2. Data preprocessing

Before using the dataset in model training stage, the dataset must undergo the preprocessing stage first. Data preprocessing aims to organize, modify, and clean data so that the data is feasible and easy to use at later stages. From the 8 classes in the dataset, only images of BCC, SCC, and MEL are taken. The frequency distribution of these three selected classes can be seen in Figure 2.

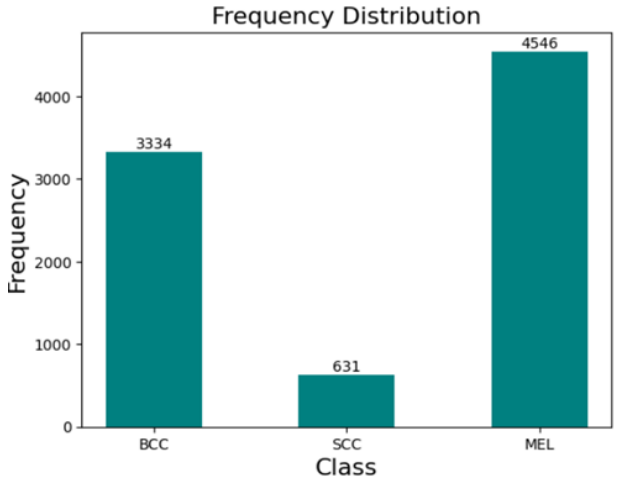

Figure 2. Frequency distribution of selected classes (BCC, SCC, MEL)

Figure 2 shows uneven distribution of samples that could cause model to prioritize majority class and ignore minority class in classification [24]. To overcome the imbalanced dataset problem, we perform

undersampling to the dataset by taking 631 samples from each class. Thus, the dataset for this study has as many as 1,893 skin cancer images. The dataset is then split into three parts by assigning 70% of the images to training set, 20% to validation set, and 10% to testing set. The distribution of the images is shown in Table 1.

Table 1. Dataset splitting

	Training set	Validation set	Testing set	Total
BCC	441	126	64	631
SCC	441	126	64	631
MEL	441	126	64	631
Total	1323	378	192	1893

All images in the dataset must be resized to 256×256 pixels before being used. The main purpose of resizing images is to make sure all the images are not taking too much time and resources when being processed or used in later stages. We specifically choose 256×256 pixels as the size because it is relatively small, efficient, and consistent (does not eliminate most of the important features of the original image). Additionally, image augmentation is also carried out by applying random flip and random rotation to the images to increase diversity and size of the training set. Data or image augmentation can improve accuracy of model and reduce overfitting effect because the model can learn with a wider variety of images [25]–[27].

2.3. Modeling

CNN is capable of extracting and learning features of image that is going to be classified. The CNN models used for skin cancer classification in this study are based on EfficientNet architecture. EfficientNet architecture is designed by Tan and Le [16] as a demonstration of their proposed compound scaling method. Scaling CNN model is usually done by increasing depth, width, and resolution which could become tedious and computationally expensive [18]. However, the proposed compound scaling method by Tan and Le [16] uses a compound coefficient to balance the network depth, width, and resolution so that a baseline CNN can be effectively scaled up. The baseline EfficientNet architecture that is known as EfficientNet-B0 is shown in Figure 3. By using the compound scaling method, baseline EfficientNet-B0 is scaled up to obtain EfficientNet-B1 to EfficientNet-B7. All of them are able to perform well with relatively smaller and faster networks [16], [18]. All EfficientNet models achieved high accuracy and efficiency on several widely used datasets, such as ImageNet, CIFAR-100, and Flowers [16].

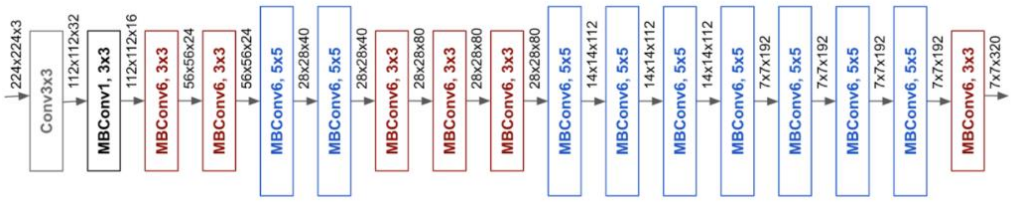

Figure 3. Baseline EfficientNet architecture (EfficientNet-B0)

In this study, we make use of transfer learning method to shorten the training time and improve performance [13], [28]. We apply transfer learning by using EfficientNet models (EfficientNet-B0 to EfficientNet-B7) that are previously trained on ImageNet to detect three categories of skin cancer, which are BCC, SCC, and MEL. All layers from each EfficientNet model are freezed to prevent the pre-trained weights from being modified. Additionally, the top layer of each EfficientNet model is replaced with global average pooling (GAP) layer for reducing the spatial dimensions of feature maps; Dropout layer for randomly setting a fraction of the neurons (units) in a layer to zero during each forward and backward pass in the training process to prevent overfitting; and an output layer to categorize input data into predefined classes. Since we are dealing with multiclass classification of skin cancer, softmax is used as the activation function in the output layer. Softmax transforms a vector of real numbers into a probability distribution which means assigning probabilities to predefined classes [29]. Categorical cross-entropy is used as the loss function to measure the dissimilarity between predicted class probabilities and true class labels. As for the optimizer, Adam is used because it is generally able to give more significant improvement in model performance than other optimizers [29]. Learning rate is set to 0.001.

A prolonged model training process usually results in overfitting which can be indicated by the validation loss value that keeps increasing. To overcome this problem, the early stopping technique is utilized. Early stopping basically means to stop the model training process when the model has become overfitting or has met certain condition [30]. Early stopping tries to remove the influence of excessive noisy labels so that the model can generalize better [31]. We set the patience of early stopping to 10 epochs which will make the training process to be stopped after 10 epochs of no improvement in validation loss.

For each EfficientNet model, the training process is carried out within 150 epochs. A batch size of 32 images is used for every iteration. Training process is directly followed by validation process to ensure generalization.

3. RESULTS AND DISCUSSION

3.1. Training and validation results

Due to the implementation of early stopping, the training and validation processes might not reach 150 epochs. The training and validation processes will stop when there is no significant improvement of loss value anymore. Each EfficientNet model with the lowest validation loss value will be used later to evaluate its performance in the testing stage.

Our EfficientNet-B0 model for skin cancer classification task only goes through 62 epochs of training and validation. The EfficientNet-B0 model at the 52nd epoch marked as the lowest validation loss value obtained. At that epoch, the training accuracy is 73.47%, the validation accuracy is 71.69%, and the validation loss is 0.5882. The graphs of training accuracy vs validation accuracy and training loss vs validation loss over the number of epochs are shown in Figure 4.

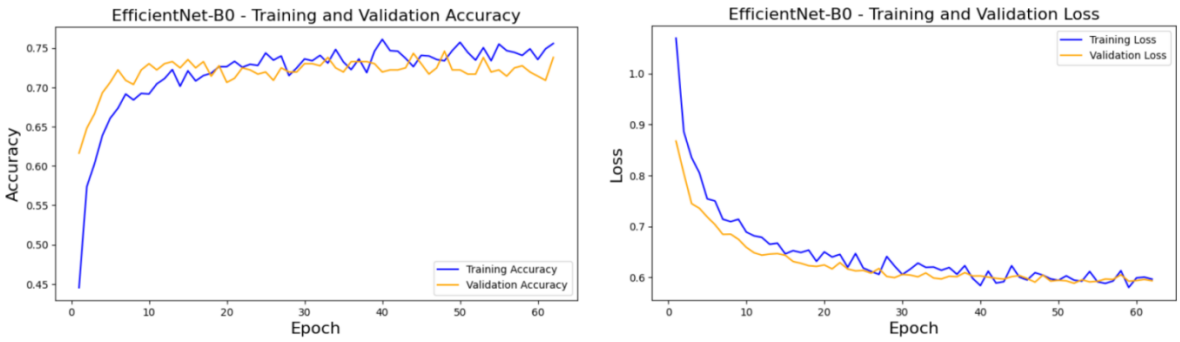

Figure 4. Training accuracy vs validation accuracy and training loss vs validation loss of EfficientNet-B0

The EfficientNet-B1 model for skin cancer classification goes through 69 epochs of training and validation. The best validation loss value achieved by the EfficientNet-B1 is 0.6057 obtained at the 59th epoch. The training accuracy is 76.72% and the validation accuracy is 73.28% at that epoch. The graphs of training accuracy vs validation accuracy and training loss vs validation loss over the number of epochs are shown in Figure 5.

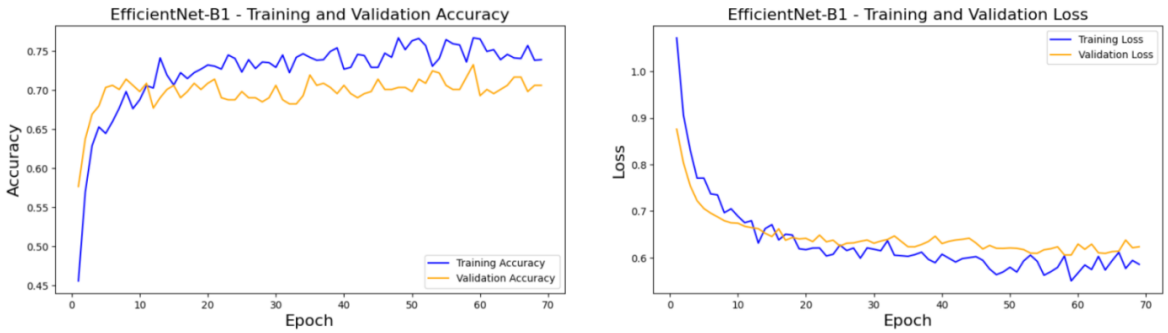

Figure 5. Training accuracy vs validation accuracy and training loss vs validation loss of EfficientNet-B1

As for the EfficientNet-B2 model, training and validation processes reach 50 epochs. The EfficientNet-B2 achieves 0.5903 as the best validation loss at the 40th epoch. At that epoch, the training accuracy is 75.66% and the validation accuracy is 72.22%. Figure 6 shows the graphs of training accuracy vs validation accuracy and training loss vs validation loss over the number of epochs.

For skin cancer classification task, the EfficientNet-B3 model undergoes 50 epochs of training and validation. The EfficientNet-B3 achieves the best validation loss at the 40th epoch that is 0.5945. The training accuracy is 74.91% and the validation accuracy is 74.87% at that epoch. Figure 7 shows training accuracy vs validation accuracy and training loss vs validation loss of the model over the number of epochs.

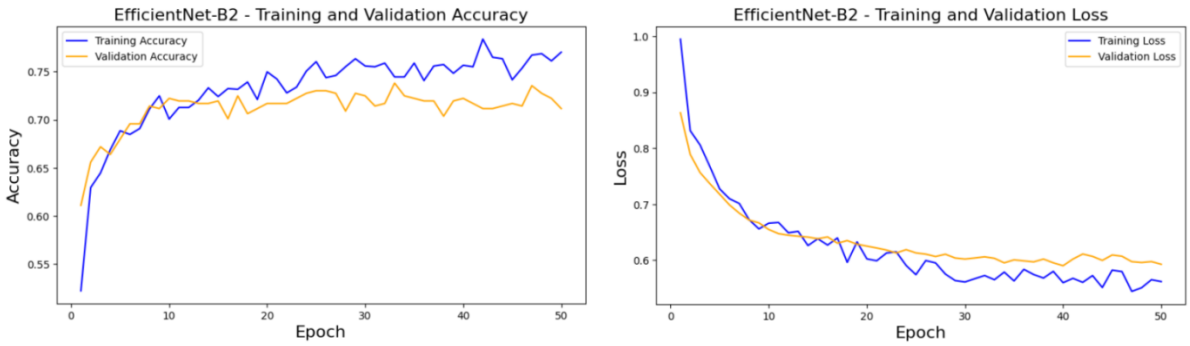

Figure 6. Training accuracy vs validation accuracy and training loss vs validation loss of EfficientNet-B2

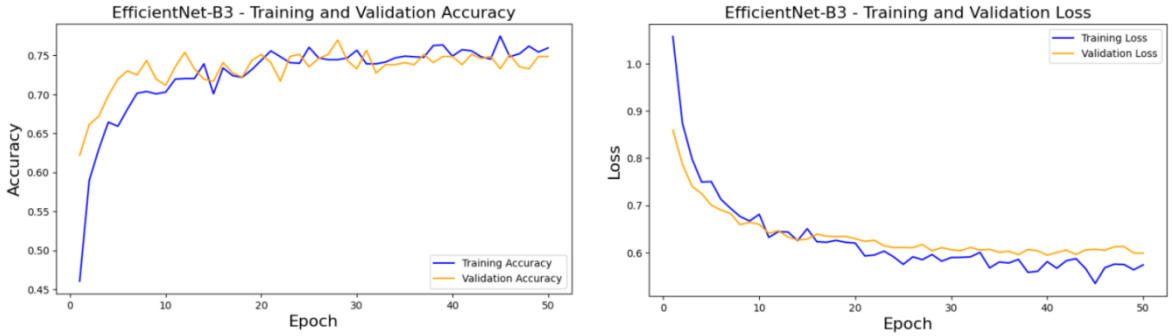

Figure 7. Training accuracy vs validation accuracy and training loss vs validation loss of EfficientNet-B3

Our EfficientNet-B4 model reaches 52 epochs of training and validation. At the 42nd epoch, the EfficientNet-B4 achieves the best validation loss that is 0.5885. The training accuracy is 74.60% and the validation accuracy is 75.66% at the 42nd epoch. Figure 8 shows training accuracy vs validation accuracy and training loss vs validation loss of the model over the number of epochs.

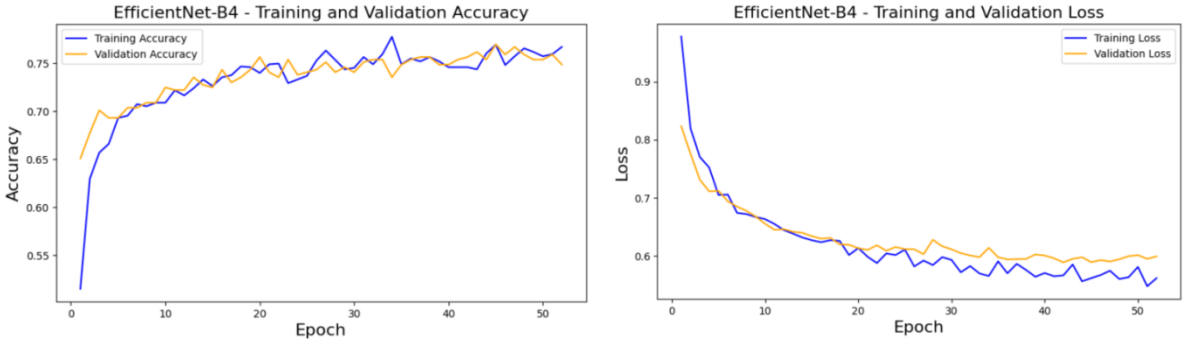

Figure 8. Training accuracy vs validation accuracy and training loss vs validation loss of EfficientNet-B4

Our EfficientNet-B5 model goes through the training and validation processes within 56 epochs. The EfficientNet-B5 successfully achieves 0.5665 at the 46th epoch as the best validation loss. The training accuracy at that epoch is 78.46%, whereas the validation accuracy is 75.93%. Figure 9 shows training accuracy vs validation accuracy and training loss vs validation loss of the model over the number of epochs.

The designed EfficientNet-B6 model undergoes 58 epochs of training and validation. The best model of EfficientNet-B6 is obtained at the 48th epoch. The best validation loss is 0.5983 along with 75.66% accuracy on training set and 73.81% accuracy on validation set. The graphs of training accuracy vs validation accuracy and training loss vs validation loss over the number of epochs are shown in Figure 10.

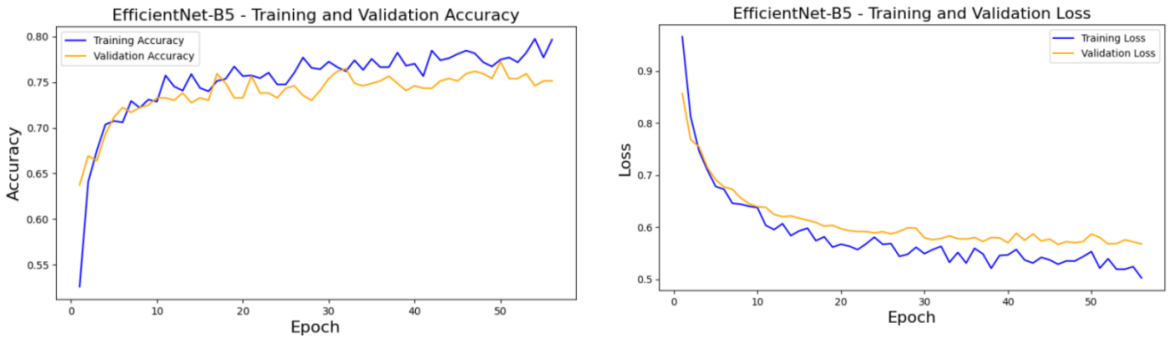

Figure 9. Training accuracy vs validation accuracy and training loss vs validation loss of EfficientNet-B5

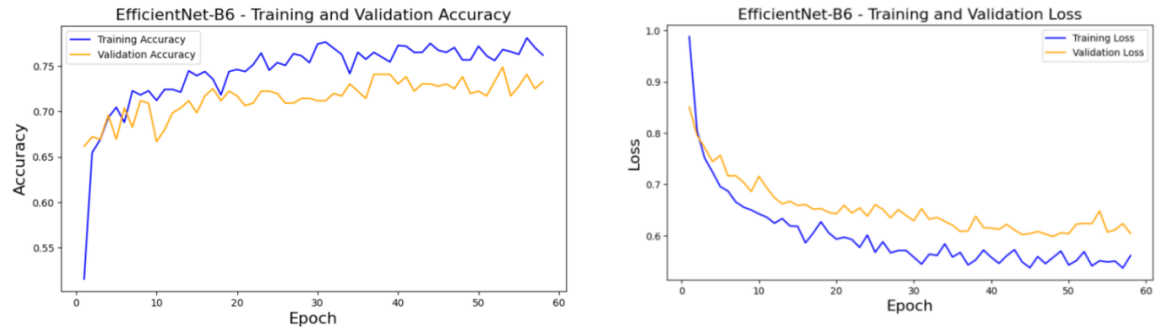

Figure 10. Training accuracy vs validation accuracy and training loss vs validation loss of EfficientNet-B6

Our designed EfficientNet-B7 model for skin cancer classification undergoes 54 epochs of training and validation. The best model of EfficientNet-B7 is obtained at the 44th epoch. The best validation loss is 0.6148 along with 77.25% accuracy on training set and 71.16% accuracy on validation set. The graphs of training accuracy vs validation accuracy and training loss vs validation loss over the number of epochs are shown in Figure 11.

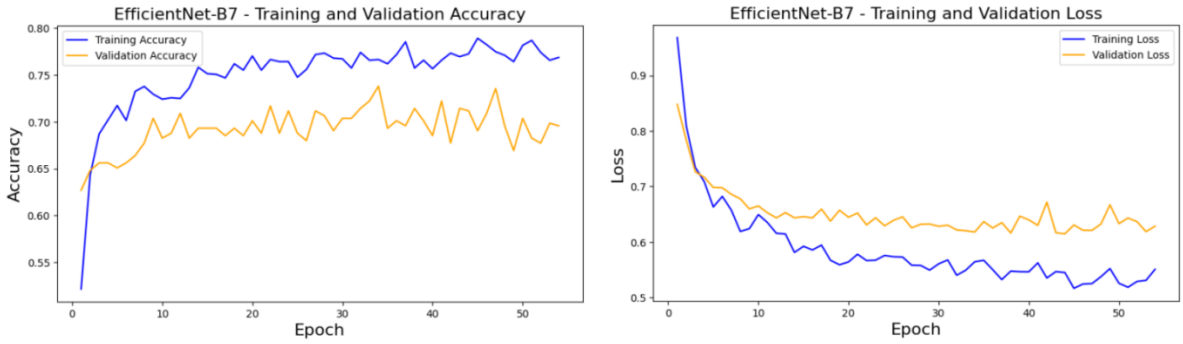

Figure 11. Training accuracy vs validation accuracy and training loss vs validation loss of EfficientNet-B7

Performances of all designed EfficientNet models on training set and validation set can be summarized as seen in Table 2. The highest accuracy on the training set is 78.46% which is achieved by EfficientNet-B5. In addition, EfficientNet-B5 also achieves the highest accuracy on the validation set that is 75.93%. In general, all of our designed EfficientNet models can perform well in classifying skin cancer on training set and validation set.

Table 2. Performances of EfficientNets on training set and validation set

Model	Training set		Validation set	
	Accuracy (%)	Loss	Accuracy (%)	Loss
EfficientNet-B0	73.47	0.5944	71.69	0.5882
EfficientNet-B1	76.72	0.5502	73.28	0.6057
EfficientNet-B2	75.66	0.5600	72.22	0.5903
EfficientNet-B3	74.91	0.5808	74.87	0.5945
EfficientNet-B4	74.60	0.5660	75.66	0.5885
EfficientNet-B5	**78.46**	**0.5286**	**75.93**	**0.5665**
EfficientNet-B6	75.66	0.5576	73.81	0.5983
EfficientNet-B7	77.25	0.5451	71.16	0.6148

3.2. Testing results

The best weights obtained in the training stage are applied to the EfficientNet models. Then, we test all of our EfficientNet models on the testing set. The testing set used in this study contains 192 skin cancer images that have been transformed to 256×256 pixels. The results of skin cancer classification on the testing set are presented as confusion matrices as shown in Figure 12.

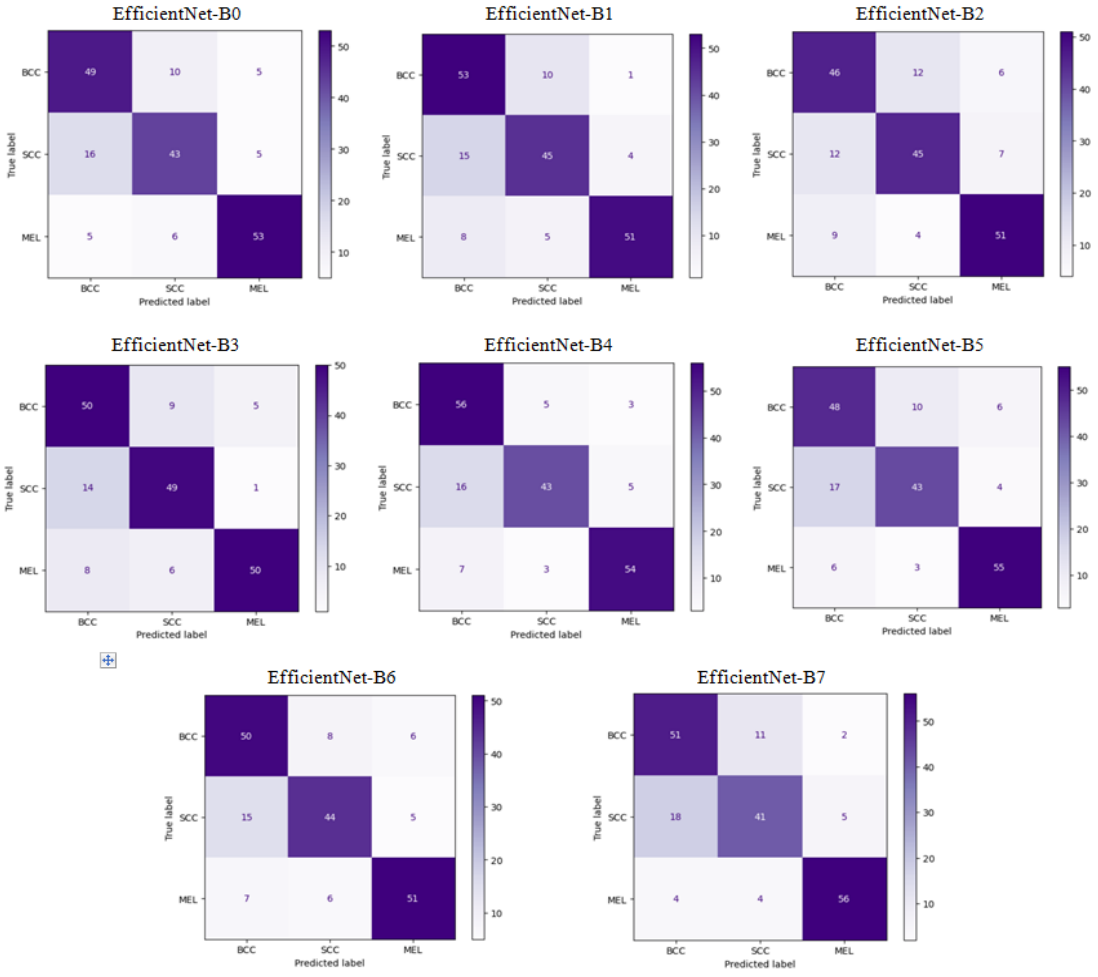

Figure 12. Classification results of all EfficientNets presented through confusion matrices

The confusion matrices in Figure 12 show that all of our EfficientNet models classify rightly most of the test images. However, there are still some misclassifications, especially between BCC and SCC. For the purpose of discovering and evaluating the performance of the models in classifying skin cancer, the metrics that we use include accuracy, precision, recall, and F1-score. The formulas of these metrics are as (1) to (4):

$$accuracy = \frac{number\ of\ correctly\ classified\ images}{total\ number\ of\ images} \quad (1)$$

$$precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2)$$

$$recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (3)$$

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall} \quad (4)$$

We use accuracy to measure the percentage of correct classification performed by a model. However, accuracy cannot be used to evaluate how well a model treats certain classes. For that reason, we also make use of precision metric and recall metric. For example, in the context of BCC class, precision calculates the number of images classified by a model as BCC that are truly BCC images and recall calculates the number of true BCC images that are classified correctly by a model. Precision and recall are able to inform us about whether a model ignores certain classes or not. Additionally, F1-score is used to measure the overall performance of a model by using harmonic mean to combine precision and recall. The values of accuracy, precision, recall, and F1-score for all designed EfficientNet models on the testing set can be seen in Table 3.

Table 3. Accuracy, precision, recall, and F1-score on testing set

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
EfficientNet-B0	75.52	80.24	69.79	74.65
EfficientNet-B1	77.60	79.53	70.83	74.93
EfficientNet-B2	73.96	78.29	71.35	74.66
EfficientNet-B3	77.60	**85.98**	73.44	**79.21**
EfficientNet-B4	**79.69**	81.67	**76.56**	79.03
EfficientNet-B5	76.04	78.57	74.48	76.47
EfficientNet-B6	75.52	77.22	72.40	74.73
EfficientNet-B7	77.08	82.53	71.35	76.54

EfficientNet-B4 has the highest accuracy that is 79.69%. The second highest accuracy is 77.60% achieved by EfficientNet-B1 and EfficientNet-B3. Highest precision is 85.98% which is obtained by using EfficientNet-B3. In addition, EfficientNet-B0, EfficientNet-B4, and EfficientNet-B7 also successfully achieve more than 80% precision. EfficientNet-B4 also achieves the best recall that is 76.56%. Regarding the F1-score, EfficientNet-B3 has the highest value that is 79.21%. Following that, EfficientNet-B4 has the second highest F1-score that is 79.03%. The confusion matrices in Figure 12 along with the values of metrics in Table 3 indicate that all designed EfficientNet models do not have bias towards certain classes. In accordance with the results, our designed EfficientNet-B4 can be regarded as the best model for skin cancer classification among other EfficientNets.

3.3. Discussion

Skin cancer is a prevalent and lethal cancer in the world. However, with a proper treatment in the early stages of skin cancer, it can be effectively cured. The role of a computer vision system is crucial in this field as it can be utilized to help in early diagnosis, detection, and classification of skin cancer. We design eight CNN models based on EfficientNet-B0 up to EfficientNet-B7 architecture in this study. Overall, all of the designed EfficientNet models successfully achieves high accuracy, precision, recall, and F1-score. What this means is, our EfficientNet models have good performance and quality in classifying BCC, SCC, and MEL skin cancer.

We consider the EfficientNet-B4 model as the best model among other EfficientNets for this task. We intend to compare this EfficientNet-B4 model with other models or methods from previous studies within the scope of skin cancer classification. The comparison can be seen in Table 4. Note that we only put the best result from each study in the table.

From Table 4, we can see that there are many different groups of classes although all of the studies are related to skin cancer classification task. This means each study listed in the table is not necessarily better than the others. Nevertheless, the performance metrics can still be good indicators for evaluating the studies on skin cancer classification task in general.

Table 4. Comparison with previous studies on skin cancer classification

Study	Classes	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Ours	BCC, SCC, and MEL	79.69	81.67	76.56	79.03
Ali *et al.* [19]	Benign and Malignant	91.43	96.57	93.66	95.09
Patil and Bellary [20]	Stage 1, Stage 2, and Stage 3 (of melanoma)	96.00	94.44	98.04	95.96
Sujaini *et al.* [6]	Positive, negative (in context of melanoma)	70.00	71.43	75.00	73.17
Gururaj *et al.* [11]	Akiec, BCC, BKL, Df, Mel, Nv, and Vasc	91.20	-	-	91.70
Xiao and Wu [21]	MM and SK	91.70	-	72.20	-
Daghrir *et al.* [10]	Benign and Malignant	88.40	-	-	-
Thurnhofer-Hemsi and Domínguez [22]	Akiec, BCC, BKL, Df, Mel, Nv, and Vasc	87.70	-	-	-

Ali *et al.* [19] achieved a high accuracy, precision, recall, and F1-score with their custom DCNN in identifying whether a skin lesion is benign or malignant. Among studies listed in Table 4, Patil and Bellary [20] with their CNN system achieved the highest accuracy, precision, recall, and F1-score. The CNN utilized SMTP loss function to classify stage 1, 2, and 3 of melanoma. On the other hand, Sujaini *et al.* [6] designed a CNN model to classify whether a skin lesion is melanoma or not. Gururaj *et al.* [11] utilized transfer learning from DenseNet169 to produce the high accuracy and F1-score shown in Table 4. They used seven classes which included BCC, MEL, and several types of skin lesion (benign and malignant). Xiao and Wu [21] classified melanoma and seborrhoeic keratosis separately. Their Global-DNN model had 91.7% accuracy and 72.2% recall in recognizing seborrhoeic keratosis while having 85.8% accuracy and 61.5% recall in recognizing melanoma. Daghrir *et al.* [10] also classified whether a skin lesion is benign or malignant. They did the task by utilizing a majority voting method to combine results from KNN, SVM, and CNN. Lastly, Thurnhofer-Hemsi and Domínguez [22] also used the same seven classes as Gururaj *et al.* [11]. However, they used transfer learning from DenseNet201 to achieve the best result.

All of these studies successfully showed that a computer vision system is really beneficial on skin cancer diagnosis, detection, and classification. In spite of that, they did not cover the three major types of skin cancer (BCC, SCC, and MEL [3], [8], [9]) at once and mostly focused on melanoma or other minor types of skin cancer. In our study, we classify BCC, SCC, and MEL by utilizing transfer learning from EfficientNets (EfficientNet-B0 up to EfficientNet-B7). All EfficientNets can work really well with high accuracy and efficiency as stated by Tan and Le [16]. Our designed EfficientNet models agree to that and show a great potential in classifying skin cancer. However, the designed models can still misclassify skin cancer. This is most likely due to some similarities between BCC, SCC, and MEL. Those similarities can occur because of genetic aspects that alter the patterns of skin cancer and create more variants [4]. Besides that, the dataset that we used is relatively small since we performed undersampling to handle the class imbalance problem. We believe that the performance of the models can be improved if the models are trained on bigger dataset. Additionally, improving classification performance also can possibly be done by utilizing image segmentation technique.

4. CONCLUSION

In this study, we design models using EfficientNet architecture to classify three most dominant types of skin cancer, namely BCC, SCC, and MEL. We preprocessed the imbalanced dataset before utilizing it to train, validate, and test our EfficientNets. To ensure generalization, we also implemented image augmentation and early stopping in model training. The results demonstrate that the EfficientNet-B4 is the best model at classifying skin cancer among other EfficientNets. On the whole, all EfficientNet models designed in this study are capable to classify BCC, SCC, and MEL with good performance. Our recommendations for future study are: improving the classification capability so that more kinds of skin cancer can be detected; testing and adapting the EfficientNet models built in this study on various datasets for skin cancer classification; and using other models, methods, or techniques in skin cancer classification.

ACKNOWLEDGEMENTS

This study would not have been possible without the ISIC 2019 dataset provided by Prasad Maharana and Larxel that are shared publicly on Kaggle. This study is supported by Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) Universitas Prima Indonesia.

REFERENCES

[1] M. Arnold, M. Laversanne, D. Singh, J. Vignat, S. Vaccarella, and F. Meheus, "Melanoma awareness month 2022," *International Agency for Research on Cancer*, 2022. https://www.iarc.who.int/news-events/melanoma-awareness-month-2022/.
[2] World Health Organization (WHO), "Ultraviolet radiation," *World Health Forum*, 1995. .
[3] L. Zeng *et al.*, "Advancements in nanoparticle-based treatment approaches for skin cancer therapy," *Molecular Cancer*, vol. 22, no. 1, pp. 1–50, 2023, doi: 10.1186/s12943-022-01708-4.
[4] M. Zambrano-Román, J. R. Padilla-Gutiérrez, Y. Valle, J. F. Muñoz-Valle, and E. Valdés-Alvarado, "Non-melanoma skin cancer: a genetic update and future perspectives," *Cancers*, vol. 14, no. 10, 2022, doi: 10.3390/cancers14102371.
[5] N. Melarkode, K. Srinivasan, S. M. Qaisar, and P. Plawiak, "AI-powered diagnosis of skin cancer: a contemporary review, open challenges and future research directions," *Cancers*, vol. 15, no. 4, 2023, doi: 10.3390/cancers15041183.
[6] H. Sujaini, E. Y. Ramadhan, and H. Novriando, "Comparing the performance of linear regression versus deep learning on detecting melanoma skin cancer using apple core ml," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 6, pp. 3110–3120, 2021, doi: 10.11591/eei.v10i6.3178.
[7] "Skin cancer," *International Agency for Research on cancer*, 2020. https://www.iarc.who.int/cancer-type/skin-cancer/ (accessed Feb. 03, 2023).
[8] W. Hu, L. Fang, R. Ni, H. Zhang, and G. Pan, "Changing trends in the disease burden of non-melanoma skin cancer globally from 1990 to 2019 and its predicted level in 25 years," *BMC Cancer*, vol. 22, no. 1, pp. 1–11, 2022, doi: 10.1186/s12885-022-09940-3.
[9] P. Fontanillas *et al.*, "Disease risk scores for skin cancers," *Nature Communications*, vol. 12, no. 1, pp. 1–13, 2021, doi: 10.1038/s41467-020-20246-5.
[10] J. Daghrir, L. Tlig, M. Bouchouicha, and M. Sayadi, "Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach," 2020, doi: 10.1109/ATSIP49331.2020.9231544.
[11] 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, doi: 10.1109/ACCESS.2023.3274848.
[12] K. Miller, "A sharper eye: how advanced imaging technologies and artificial intelligence are transforming melanoma diagnosis," *The Melanoma Letter*, vol. 38, no. 1, 2020.
[13] D. Mohammad, I. Aljarrah, and M. Jarrah, "Searching surveillance video contents using convolutional neural network," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 2, pp. 1656–1665, Apr. 2021, doi: 10.11591/ijece.v11i2.pp1656-1665.
[14] M. A. Uthaib and M. S. Croock, "Multiclassification of license plate based on deep convolution neural networks," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 6, pp. 5266–5276, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5266-5276.
[15] A. Berroukham, K. Housni, M. Lahraichi, and I. Boulfrifi, "Deep learning-based methods for anomaly detection in video surveillance: a review," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 1, pp. 314–327, 2023, doi: 10.11591/eei.v12i1.3944.
[16] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *36th International Conference on Machine Learning, ICML 2019*, May 2019, pp. 10691–10700.
[17] T. Selim, I. Elkabani, and M. A. Abdou, "Students engagement level detection in online e-learning using hybrid EfficientNetB7 together with TCN, LSTM, and Bi-LSTM," *IEEE Access*, vol. 10, pp. 99573–99583, 2022, doi: 10.1109/ACCESS.2022.3206779.
[18] M. Moghaddam, M. Charmi, and H. Hassanpoor, "Jointly human semantic parsing and attribute recognition with feature pyramid structure in EfficientNets," *IET Image Processing*, vol. 15, no. 10, pp. 2281–2291, 2021, doi: 10.1049/ipr2.12195.
[19] M. S. Ali, M. S. Miah, J. Haque, M. M. Rahman, and M. K. Islam, "An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models," *Machine Learning with Applications*, vol. 5, 2021, doi: 10.1016/j.mlwa.2021.100036.
[20] R. Patil and S. Bellary, "Machine learning approach in melanoma cancer stage detection," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 6, pp. 3285–3293, 2022, doi: 10.1016/j.jksuci.2020.09.002.
[21] F. Xiao and Q. Wu, "Visual saliency based global-local feature representation for skin cancer classification," *IET Image Processing*, vol. 14, no. 10, pp. 2140–2148, 2020, doi: 10.1049/iet-ipr.2019.1018.
[22] K. Thurnhofer-Hemsi and E. Domínguez, "A convolutional neural network framework for accurate skin cancer detection," *Neural Processing Letters*, vol. 53, no. 5, pp. 3073–3093, 2021, doi: 10.1007/s11063-020-10364-y.
[23] P. Maharana, "ISIC 2019 skin lesion images for classification," 2022. https://www.kaggle.com/datasets/salviohexia/isic-2019-skin-lesion-images-for-classification (accessed Feb. 11, 2023).
[24] E. Rendón, R. Alejo, C. Castorena, F. J. Isidro-Ortega, and E. E. Granda-Gutiérrez, "Data sampling methods to dealwith the big data multi-class imbalance problem," *Applied Sciences (Switzerland)*, vol. 10, no. 4, 2020, doi: 10.3390/app10041276.
[25] D. Yarats, R. Fergus, and I. Kostrikov, "Image augmentation is all you need: regularizing deep reinforcement learning from pixels," *International Conference on Learning Representations*, Vienna, Austria, 2021, pp. 1-21.
[26] K. Okokpujie, S. John, C. Ndujiuba, J. A. Badejo, and E. Noma-Osaghae, "An improved age invariant face recognition using data augmentation," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 1, pp. 179–191, 2021, doi: 10.11591/eei.v10i1.2356.
[27] J. C. Setyono and A. Zahra, "Data augmentation and enhancement for multimodal speech emotion recognition," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 5, pp. 3008–3015, 2023, doi: 10.11591/eei.v12i5.5031.
[28] S. Bozinovski, "Reminder of the first paper on transfer learning in neural networks, 1976," *Informatica (Slovenia)*, vol. 44, no. 3, pp. 291–302, 2020, doi: 10.31449/INF.V44I3.2828.
[29] T. Yu and H. Zhu, "Hyper-parameter optimization: a review of algorithms and applications," *arXiv*, pp. 1–56, 2020, doi: 10.48550/arXiv.2003.05689.
[30] L. Rice, E. Wong, and J. Z. Kolter, "Overfitting in adversarially robust deep learning," in *37th International Conference on*

Machine Learning, ICML 2020, 2020, pp. 8049–8074.
[31] X. Xia *et al.*, "Robust early-learning: hindering the memorization of noisy labels," in *ICLR 2021-9th International Conference on Learning Representations*, 2021, pp. 1–15.

BIOGRAPHIES OF AUTHORS

Mawaddah Harahap received her master degree in Computer Science from Universitas Sumatera Utara (USU) in 2014. She is currently a lecturer at the Department of Computer Science, Universitas Prima Indonesia (UNPRI). Her research interests cover various fields in computer vision and data science. In this study, her main role is in conceptualization and methodology. She can be contacted at email: mawaddah@unprimdn.ac.id.

Amir Mahmud Husein received his master degree in Computer Science from Universitas Sumatera Utara (USU) in 2011. He is currently a lecturer at the Department of Computer Science, Universitas Prima Indonesia (UNPRI). His research interests include various fields in big data, computer vision, and data science. In this study, his main role is in investigation and resources. He can be contacted at email: amirmahmud@unprimdn.ac.id.

Shane Christian Kwok is currently a student at Universitas Prima Indonesia majoring in Computer Science. He is under Mrs. Mawaddah Harahap's guidance for this research. His research interests are primarily in the area of computer vision and data science. In this study, his main role is in project administration, programming, and writing. He can be contacted at email: shanechristian03@gmail.com.

Vincent Wizley is currently a student at Universitas Prima Indonesia majoring in Computer Science. He is under Mrs. Mawaddah Harahap's guidance for this research. His research interests include computer vision, data science, and machine learning. In this study, his main role is in data analysis, programming, and writing. He can be contacted at email: vincentwizley72@gmail.com.

Jocelyn Leonardi is currently a student at Universitas Prima Indonesia majoring in Computer Science. She is under Mrs. Mawaddah Harahap's guidance for this research. Her research interests are in computer vision and data science. In this study, her main role is in programming and validation. She can be contacted at email: jcelynwht@gmail.com.

Derrick Kenji Ong is currently a student at Universitas Prima Indonesia majoring in Computer Science. He is under Mrs. Mawaddah Harahap's guidance for this research. His research interests are in computer vision and data science. In this study, his main role is in validation and visualization. He can be contacted at email: derrick.kenji.ong@gmail.com.

Deskianta Ginting is currently a student at Universitas Prima Indonesia majoring in Computer Science. He is under Mrs. Mawaddah Harahap's guidance for this research. His research interests are in computer vision and data science. In this study, his main role is in validation and visualization. He can be contacted at email: deskiginting39@gmail.com.

Benny Art Silitonga is currently a student at Universitas Prima Indonesia majoring in Computer Science. He is under Mrs. Mawaddah Harahap's guidance for this research. His research interests are in computer vision and data science. In this study, his main role is in data analysis and writing. He can be contacted at email: bennyartsilitonga0@gmail.com.