Journal of
Information Systems Engineering
and Business Intelligence

Vol.8, No.2, October 2022
Available online at: http://e-journal.unair.ac.id/index.php/JISEBI

Melanoma Detection using Convolutional Neural Network with Transfer Learning on Dermoscopic and Macroscopic Images

Jessica Millenia[1)], Mohammad Farid Naufal[2)*], Joko Siswantoro[3)]

[1)2)3)]*Teknik Informatika, Universitas Surabaya, Indonesia*
Jl. Raya Kalirungkut, Surabaya, Indonesia
[1)]jessicamillenia@gmail.com [2)]faridnaufal@staff.ubaya.ac.id, [3)]joko_siswantoro@staff.ubaya.ac.id

Abstract

Background: Melanoma is a skin cancer that develops when the melanocytes that produce the skin colour pigment start to grow out of control and form cancer. Detecting melanoma early is very important because it dramatically affects patients' prognosis. A skin examination of a mole indicated as melanoma can be done using dermoscopic or macroscopic images. However, manual screening takes a long time, so automatic melanoma detection is needed. Previous studies still have weaknesses because they yielded low precision or recall. The distribution of melanoma and moles datasets is imbalanced, with the number of melanomas lower than moles. In addition, previous studies have not compared several convolutional neural network (CNN) transfer learning architectures on dermoscopic and macroscopic images.
Objective: This study aims to detect melanoma using CNN with transfer learning from dermoscopic and macroscopic melanoma images. CNN with transfer learning is a popular method for classifying digital images with high performance on an imbalanced dataset.
Methods: This study compares four CNN with transfer learning architectures, namely MobileNet, Xception, VGG16, and ResNet50, in examining dermoscopic and macroscopic images. This research also uses black-hat filtering and inpainting at the pre-processing stage to remove hair from the skin image.
Results: MobileNet is the best model for classifying melanomas or moles in this experiment which has an 83.86% F1 score and 11 seconds of training time per epoch.
Conclusion: MobileNet has high average F1 scores of 84.42%, which can detect melanoma accurately even though the number of melanoma datasets is less than moles. It can be concluded that MobileNet is a suitable model for classifying melanomas and moles. In the future, the oversampling method can be implemented to balance the datasets to improve the model's performance.

Keywords: CNN, Dermoscopic, Macroscopic, Melanoma, Moles, Transfer Learning.

Article history: Received 28 July 2022, first decision 9 September 2022, accepted 30 September 2022, available online 28 October 2022

I. INTRODUCTION

Melanoma is a skin cancer that develops when the melanocyte cells that produce the pigment and colour of the skin begin to grow out of control and form cancer. Based on data from the American Cancer Society [1], 29% of melanomas typically develop from existing normal moles on human skin, and 71% grow in new places. The shape, size, colour, or thickness will change if it develops from an existing mole. If it grows in a new place, spots or dots on the skin will grow. Melanoma can grow in all parts of the human body, including the head, neck, soles of the feet, and genitals [2]. The American Cancer Society [1] estimates that by 2021 in the United States, there will be approximately 106,110 new melanoma cases (62,260 cases in men and 43,850 cases in women). The number of deaths from melanoma skin cancer was also expected to increase by 4.8 per cent in 2021. About 7,180 out of 106,110 people are expected to die from melanoma (4,600 men and 2,580 women). Based on previous data, early detection of melanoma is essential. However, often humans are late in detecting melanoma. Melanoma may cause itching or odd feeling on the skin but does not leave pain or tenderness on the skin [1]. Early detection includes a dermatologist's skin screening. However, manual screening takes time and can be costly. The easier way to detect melanoma early is by self-screening. If patients feel a change in an existing mole or feel that a new mole is oddly shaped, patients can check with a doctor for further

* Corresponding author

ISSN 2443-2555 (online) 2598-6333 (print)
doi: http://dx.doi.org/10.20473/jisebi.8.2.149-161

inspection. However, many patients do not realise if the melanoma has grown in a new place or if the mole has become a melanoma because patients do not feel pain.

Dermoscopic images can be obtained from a camera with the help of an instrument called a dermatoscope. However, since dermoscope is expensive, an alternative would be obtaining macroscopic images using a digital macro camera with a resolution greater than 1 megapixel. This camera is easier to access and use by the public.

Automatic Melanoma detection using computer vision and deep learning can make it easier for patients to detect melanoma more quickly and flexibly. CNN is a popular deep learning algorithm for classifying images. Several previous studies have used machine learning methods to detect melanoma. The previous studies only used dermoscopic or macroscopic images. They did not use pre-processed images to remove hair from the skin image, which can reduce the accuracy of melanoma detection. In addition, the distribution of melanoma and moles datasets is imbalanced, where the number of melanomas is less than moles. The previous studies still have weak precision or recall because of the imbalanced dataset.

This study solves the imbalanced classification issue in the melanoma dataset using CNN with transfer learning architecture, namely MobileNet, Xception, VGG16, and ResNet50. CNN with transfer learning is a popular method for classifying digital images with high performance on an imbalanced dataset. This research also uses black-hat filtering and inpainting at the pre-processing stage to remove hair from the skin image. The output of this research is a web application that can be accessed publicly on this link https://deteksimelanoma.herokuapp.com. People with skin problems can use the web application to do self-screening to identify their skin problem is melanoma or moles.

II. LITERATURE REVIEW

Using computer vision to differentiate melanoma from moles has been done before. For example, Murugan et al. [3] conducted a study to distinguish melanoma from moles using dermoscopic images using the K-Nearest Neighbour (KNN), the Support Vector Machine (SVM) and Random Forest. The dataset used in the study came from the *International Skin Imaging Collaboration* (ISIC) [4]. Using five stages: pre-processing using gaussian filtering, segmentation using active contour, feature extraction using moment invariant, classification process using the KNN and SVM, and performance calculation, the study produces an accuracy of 82.31% for SVM, 62.19 for KNN and 71.97% for Random Forest. The recall is 87.76% for SVM, 74.75% for KNN, and 65.72% for Random Forest. However, there is no information about the precision values.

Kavitha et al. [5] proposed the detection of melanoma from dermoscopic images using a feature extraction model by employing global and local methods. Using a simple adaptive thresholding approach, the image is pre-processed and segmented. The filtered picture is then put via feature extraction. Utilising GLCM and SURF, these texture characteristics are employed to assess skin lesion differentiation. SVM-RBF and KNN are used in the classification procedure. The experimental finding demonstrates that the local texture feature descriptor SURF, in conjunction with SVM-RBF, produces improved classification outcomes, with sensitivity, specificity, and accuracy of 86.2%, 88.4%, and 87.3%, respectively.

Ramezani et al. [6] create a novel method for distinguishing benign pigmented lesions from malignant melanoma using macroscopic pictures from DermAtlas [7] database. The photographs were captured with standard digital cameras with a spatial resolution of more than one megapixel, without restrictions, and under no special imaging settings. The suggested process includes novel approaches for reducing the impact of nonuniform lighting, addressing the impact of thick hairs and big glow on the lesion, and a new threshold-based segmentation algorithm. After 187 features representing asymmetry, border irregularity, colour variation, diameter, and texture were extracted from the lesion area, principal component analysis (PCA) was used to reduce the number of features. After that, a support vector machine classifier was used to categorise the lesions as malignant or benign. The results show that the accuracy, sensitivity, and specificity were 82.2%, 77%, and 86.93%, respectively.

Hasanah et al. [8] classified melanoma in dermoscopic images depending on whether they show Common Nevus, Atypical Nevus, or Melanoma using Mendonca et al. database [9]. The RGB is converted to grayscale. The image is smoothed using median filtering and segmented using binary images of skin lesions. In the following procedure, classification using the multi-SVM algorithm uses the values of contrast, correlation, energy, and homogeneity acquired from the texture feature extraction of the GLCM approach. The results in identifying skin cancer are very accurate.

Pillay et al. [10] examine the commonly utilised ABCD rule on macroscopic images, and the Graph-Cut segmentation approach in the Department of Dermatology of the University Medical Centre Groningen (UMCG) called the *MED-NODE* [11], which can deal with excessively texture, noise, and colours seen in macroscopic images. This study produces an accuracy of 73.52% for SVM and 75.29% for KNN. However, this study only uses regular machine learning with low accuracy, and there is no information about the precision and recall of this study.

Oliveira et al. [11] conducted a study to distinguish melanoma from moles using macroscopic images with SVM. The dataset used in this study came from Loyola University Dermatology Medical Education [12], YSP Dermatology Image Database [13], and DermAtlas [14]. Unfortunately, the dataset used in this study is not yet publicly available. At the pre-processing stage, artefacts that could affect accuracy were removed. An anisotropic diffusion filter was used to smooth the image and remove annoying noise, especially hair, without removing the necessary boundaries. Before using this filter, the image was converted to grayscale. The mole or melanomas were separated at the segmentation stage using Chan-Vese's [15], i.e., based on the average pixel intensity. After that, post-processing was carried out using morphological filtering, i.e., opening and closing. The results of this stage were combined or overlapped with the original image. After that, classification was carried out using the SVM, producing an accuracy of 74.36% and precision of 75.24%.

Hurtado et al. [16] examined the effects of the smoothing bootstrapping, which was used to expand the initial dataset, on the outcomes of a unique skin cancer classification system that operates on photos captured using a regular camera. This study used a dataset from DermAtlas [14]. Unfortunately, this dataset is not publicly available either. Eight alternative topologies (KNN, ANN, and SVM) classifiers were tested, both with and without data augmentation. It was found that the ANN with data augmentation had the greatest performance and the most balanced results. This study produces 86.3% accuracy, 86.9% of specificity, and 87.8 sensitivity.

When comparing Murugan et al. [3], Kavitha et al. [5], and Hasanah et al. [8] for dermoscopic images with the study from Ramezani et al. [6], Pillay et al. [10], Oliveira et al. [11], and Hurtado et al. [16] for macroscopic images, it is known whether the classification performance of macroscopic images is better than dermoscopic images. Murugan et al. [3] using the SVM for melanoma detection in dermoscopic images produced the best accuracy, but it was not implemented on macroscopic images. The study from Hurtado et al. [16] performed well but only classified the macroscopic images. The imbalanced dataset was also a problem in melanoma detection. In addition, in the previous study, comparisons of several CNN transfer learning architectures have not been carried out on dermoscopic and macroscopic images. They have not been developed in a real-time application. In this study, melanoma is detected from dermoscopic and macroscopic images using CNN and several transfer learning architectures, which are ResNet50, VGG16, MobileNet, and Xception, to solve the imbalanced dataset issue. This research produces a real-time melanoma detection application.

III. METHODS

The research methodology for this work includes data collection, data pre-processing, model training, model testing, and performance calculation. Fig. 1 depicts the research methodology used in this study.

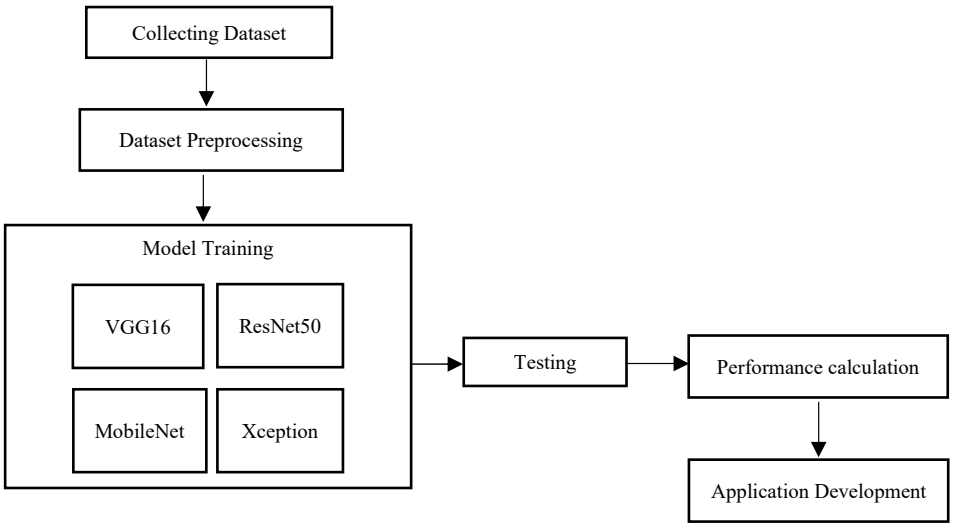

Fig. 1 Research methods

A. Collecting Dataset

In this study, three public datasets are used: the International Skin Imaging Collaboration *(ISIC)* [4], the Department of Dermatology of the University Medical Center Groningen (UMCG) or the *MED-NODE* [17], and the Dermatological and Surgical Assistance Program (PAD) at the Federal University of Espírito Santo (UFES) or the *PAD-UFES-20* [18]. *ISIC* is an academic and industrial partnership annually providing dermoscopic images of melanomas and moles. This dataset is collected by Claudio Fanconi, consisting of 200 images of melanomas and 497 images of moles. *MED-NODE* provides macroscopic images of melanomas and moles obtained with the Nikon D3, Nikon D1x body, and Nikkor 2.8/105 mm micro lens with a distance between the skin and camera of 33 cm. *MED-NODE* contains 100 images of moles and 70 images of melanoma. The *PAD-UFES-20* offers macroscopic images of melanomas and moles from patients at the Hospital Universitário Cassiano Antonio Moraes (HUCAM) obtained via a smartphone camera [18]. The *PAD-UFES-20* dataset contains 46 images of melanomas and 100 images of moles. Table 1 shows the details of the dataset. The three datasets are divided into training, testing, and validation sets. In this study, each dataset is divided into 80% for training and 20% for testing. Then the remaining 80% is divided into 80% for training and 20% for validation. In this study, classification is carried out on the merged data from those three datasets. Fig. 2 shows the dermoscopic and macroscopic images of melanoma and moles.

TABLE 1
THE SIGNIFICANCE OF THE RELATIONSHIPS IN THE MODEL

Dataset	Type	Number of images	
		Melanoma	Mole
ISIC	dermoscopic	1497	1800
MED-NODE	macroscopic	70	100
PAD-UFES-20	macroscopic	46	100

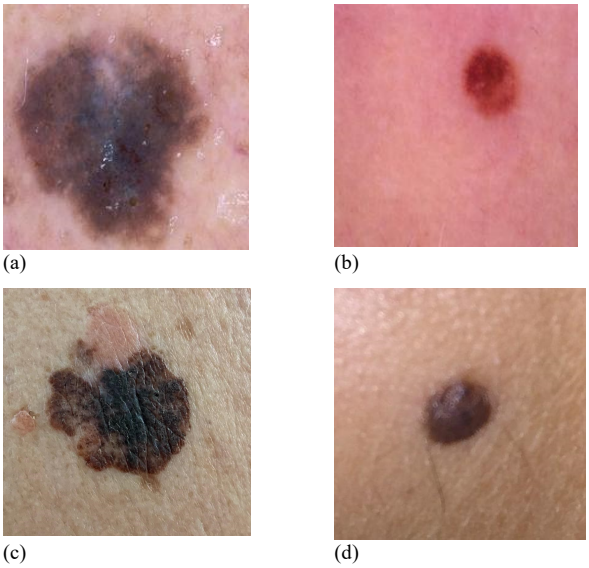

Fig. 2 (a) & (b) dermoscopic images of melanoma and moles, (c) & (d) macroscopic images of melanoma and moles

B. Dataset Pre-processing

Pre-processing is carried out to check whether there is missing data, inconsistent data, duplicate data, and noise. Data pre-processing is essential in machine learning so that the data's quality improves, improving model performance. The pre-processing methods used in this research are resizing, noise filtering, and normalisation.

1) Resize

This study uses bilinear interpolation and zero padding to resize the image, as shown in Fig. 3. Bilinear interpolation performs computations efficiently and produces good image quality [19]. Mathematically, bilinear interpolation can be written as (1), A is the pixel being scaled B, C, and D are neighbouring pixels of A, x and y are the locations of the pixel, x' and y' are the location of the new pixel after scaling, and I is the result of scaling.

$$I_{x',y'} = (1-x)\,(1-y)\,A_{x,y} + x\,(1-y)\,B_{x+1,y} + y\,(1-x)\,C_{x,y+1} + x\,y\,D_{x+1,y+1} \quad (1)$$

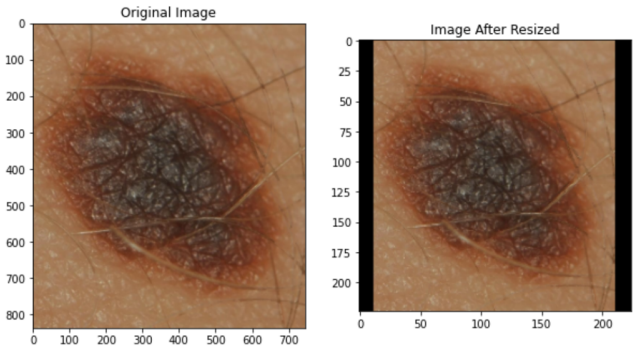

Fig. 3 Example of resizing using bilinear interpolation and zero padding

2) Noise Filtering

In this study, the noise is the hairs around the image of moles or melanomas. To remove this noise, black-hat filtering and inpainting are needed. Black-hat filtering is a technique in morphological transformations [20], which can be written as (2), where BH is the results of black-hat filtering, F is the original image filtered, B is the filter used, and operator • is a closing operation where dilation and erosion occur.

$$BH(F) = (F \bullet B) - F \quad (2)$$

The intensity of the results of black-hat filtering is increased using inpainting and binary thresholding [21]. Binary thresholding can be written as (3), where I is the result of thresholding. x and y are pixel locations that are given a threshold, which is the mean intensity. If the pixel exceeds the threshold, it is marked 1—otherwise, 0.

$$I(x,y) = \{1 \;\; if\; I(x,y) \geq threshold \; 0 \;\; if\; I(x,y) < threshold \quad (3)$$

Inpainting replaces or paints the existing noise with pixels similar to the neighbouring pixels. It removes hair in the image with its neighbouring pixels to restore the background [22]. Mathematically, inpainting can be written as (4) where I is the original image, V is the result of inpainting, i and j indicate the pixel location, and k indicates the channel.

$$V(I) = \sum_{i=0}^{a} \sum_{j=0}^{b} \sum_{k=0}^{c} \left(\left| I_{i+1,j,k} - I_{i,j,k} \right|^2 + \left| I_{i,j+1,k} - I_{i,j,k} \right|^2 + \left| I_{i,j,k+1} - I_{i,j,k} \right|^2 \right) \quad (4)$$

Fig. 4 shows the stages of the black hat filtering and inpainting process. The image in the upper left is the original image filtered. The image on the top right is the result of black-hat filtering. The image on the bottom left is the result of increasing the intensity of black-hat filtering before inpainting. Finally, the image on the bottom right is the final result of noise filtering after black-hat filtering and painting.

3) Normalisation

Normalisation is needed to change all values to be between a smaller range to produce more optimal results [23]. The new value is obtained using the normalisation formula in (5). x is the old array, x' is the range of values for the new array, and z is the result of normalisation. max is the largest value in the array, and min is the smallest value in the array.

$$z_i = \frac{x_i - (x)}{(x') - (x')} \quad (5)$$

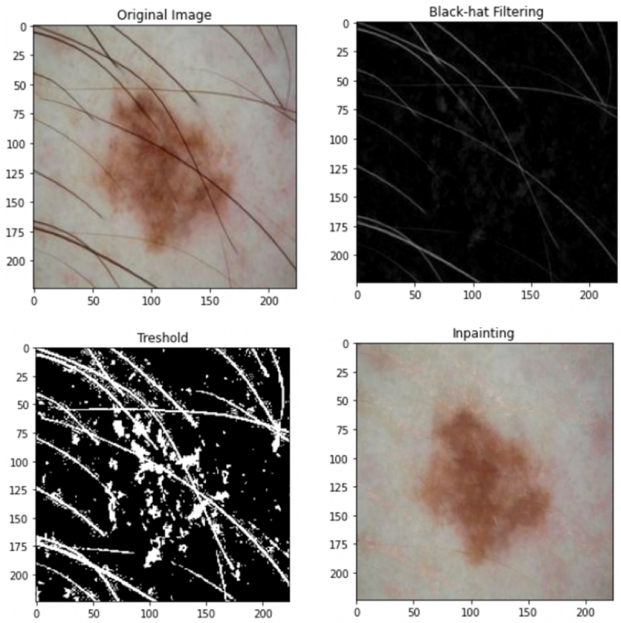

Fig. 4 the stages of the black hat filtering and inpainting process

C. Model Training

In this step, the MobileNet, VGG16, ResNet50, and Xception architectures are trained in two versions: with and without pre-processing, to see the effectiveness. MobileNet was chosen because it is a lightweight model, small, with low latency and a fast computing time [24], so it is suitable to be applied to mobile devices. Fig. 5 shows the MobileNet architecture. Meanwhile, VGG16 was chosen because it was a runner-up winner of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition in 2014. It is a deep model with a normal computation but can produce good classification results. Therefore, it is suitable to be used as a comparison [25]. Fig. 6 shows the VGG16 architecture. Likewise, Xception was chosen because it was developed from Inceptionv3 [26], and ResNet50 was chosen because it won the ILSVRC competition in 2015. ResNet50 is a development of VGG19 and equipped with a residual network. This complex model has parameters that are not too large so that computing can run faster [27]. Fig. 7 shows the ResNet50 architecture. As for Xception, it has the same number of parameters as Inceptionv3 but more superior performance than InceptionV3 on ImageNet datasets. Fig. 8 shows the Xception architecture.

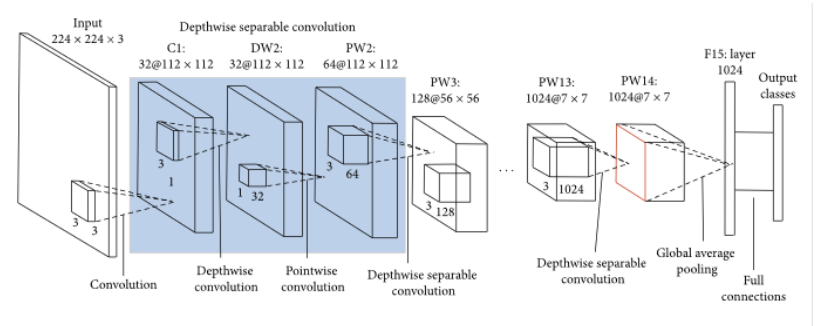

Fig. 5 MobileNet architecture

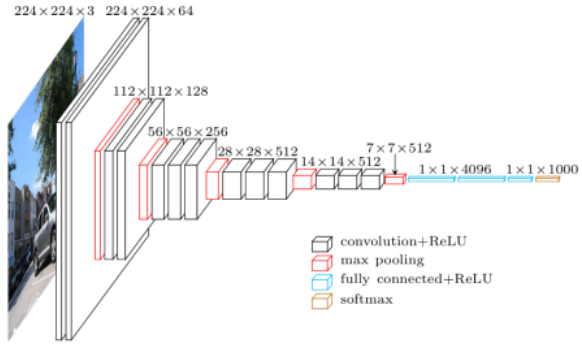

Fig. 6 VGG16 architecture

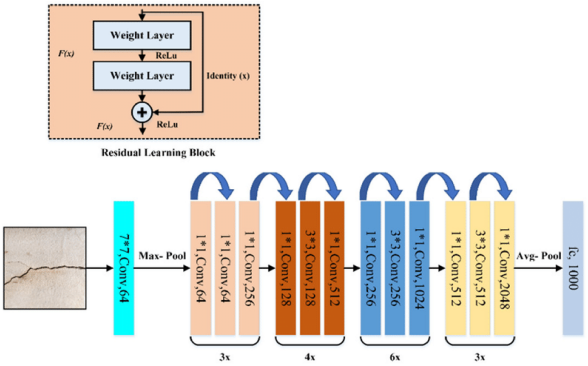

Fig. 7 ResNet50 architecture

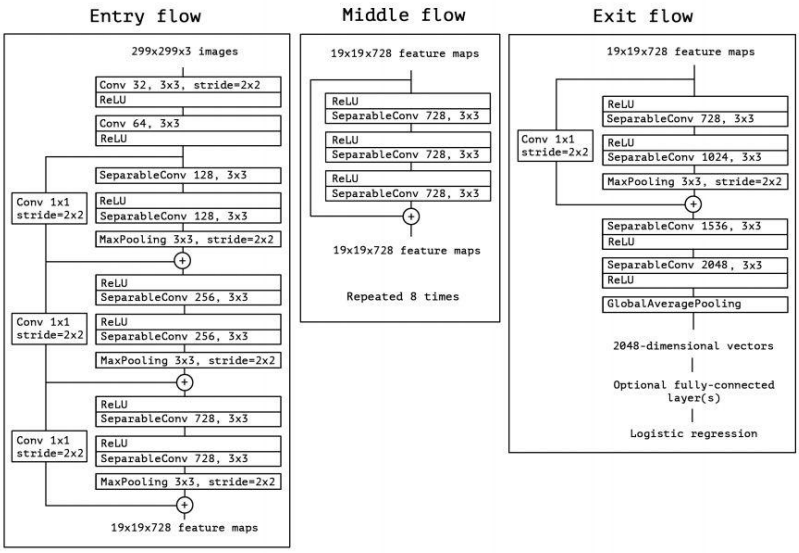

Fig. 8 Xception architecture

TABLE 2
MODEL PARAMETERS

Parameter	Values
Transfer learning architecture	VGG16, ResNet50, MobileNet, Xception
Optimizer	Adam
Learning rate	0.0004
Loss function	binary crossentropy
Early Stopping	if the validation accuracy of the next 3 epochs is not better than the current accuracy.

TABLE 3
COMPUTING MACHINE SPECIFICATION

Parameter	Specification
CPU	Intel® Xeon®, 2.30 GHz, 2 cores
RAM	12 GB
Space of Disk	358 GB
GPU Model Name	Nvidia K80, 12 GB

Pre-trained models of the transfer learning architecture are used in the training process. The top layer is replaced to match the required input, and the bottom layer is replaced according to the required prediction results, which is binary classification. Then, the model is compiled with the Adam optimiser using a learning rate of 0.0004, binary cross-entropy for the loss function and 20 epochs. In addition, a call-back is also installed to stop the training process so that there is no overfit while monitoring the accuracy of the validation set. Table 2 shows the parameter of the model. Finally, the best checkpoints are used for evaluation using a test set after completing the training process. The training process is carried out on a computing machine with specifications that can be seen in Table 3.

D. Testing and Performance Calculation

Next, all the models that have been trained are tested. The testing process uses test-set data. A confusion matrix shows class 0 for melanoma and class 1 for moles. The testing output is accuracy, precision, specificity, recall, and F1 scores of the combined dataset and each dataset. This step involves calculating how well each algorithm performed throughout the testing phase. The performance parameters are accuracy, precision, recall, F1 score, and training and testing procedure execution time. Each algorithm's performance in each cross-validation is calculated.

Equation 6 displays the accuracy calculation, which is done by dividing the sum of true positives (TP) and negatives (TN) by the sum of true positives and true negatives, false positives (FP), and false negatives (FN).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

The calculating formula for precision is shown in (7). By dividing TP by the sum of TP and FP, precision is computed.

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

The specificity calculation formula is shown in (8). A recall is determined by dividing the TP by the sum of the TP and FN.

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

The recall calculation formula is shown in (9). A recall is determined by dividing the TP by the sum of the TP and FN.

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

The F1 score computation formula is shown in (10). The sum and the multiplication of precision and recall are used to get the F1 score.

$$F1\ Score = \frac{Precision\ x\ Recall}{Precision + Recall} \tag{10}$$

E. Application and Development

The application is web-based and can be accessed from mobile or desktop devices. The main feature of this application is that users can upload the skin image, crop the image so that only moles or melanoma are visible, and the system will provide the prediction. Then, the system will resize and filter the image to remove the noise. After that, the image is displayed on the website, and the user can press the predict button. Then, the system will rescale the image. Finally, the model will make predictions on the skin images, and the user will receive the prediction results from the application. The application can be accessed through this link https://deteksimelanoma.herokuapp.com. Fig. 9 shows the user interface and the step of melanoma detection using the detector application.

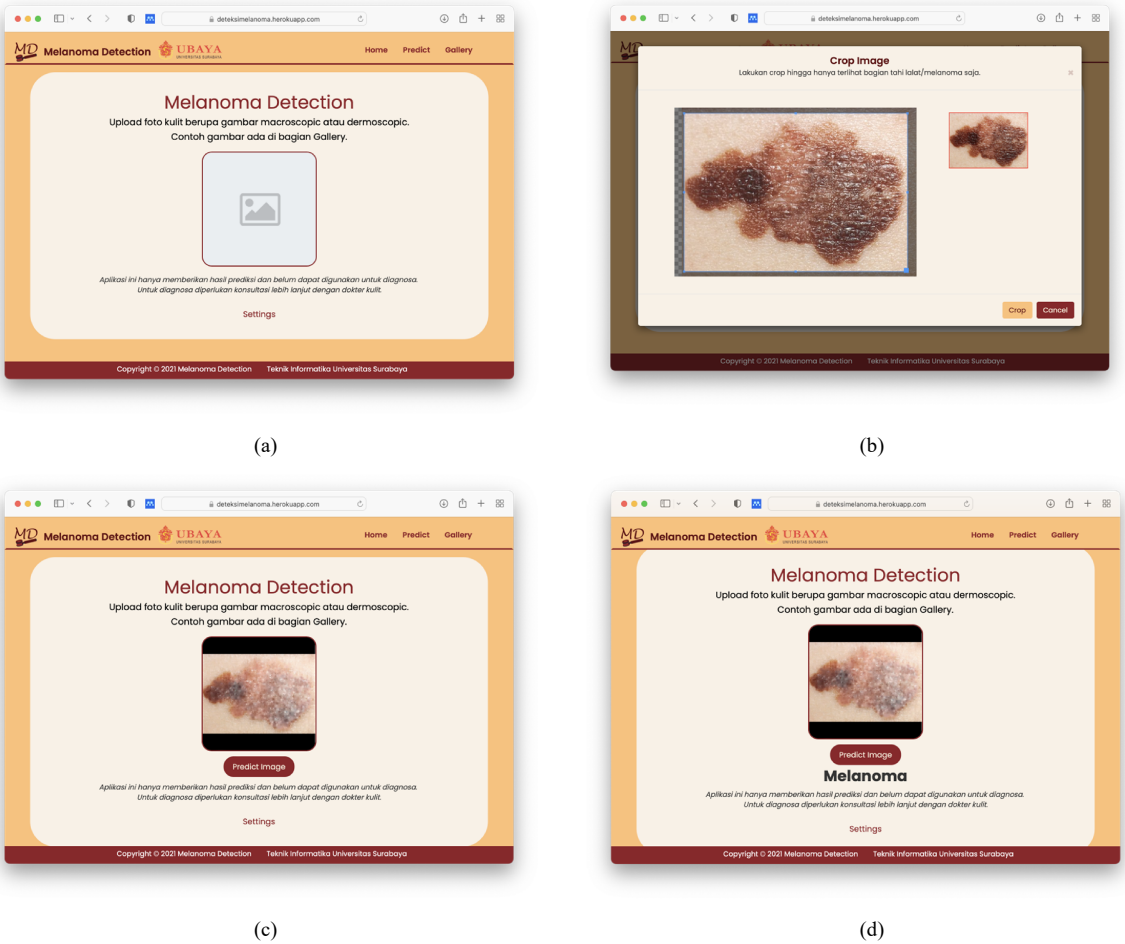

Fig. 9 User interface and the step of melanoma detection using melanoma detector application: (a) homepage, (b) crop the image, (c) predict the image, (d) the result

IV. RESULTS

A. Training

The training process with MobileNet, Xception, VGG16, and ResNet50 architecture runs for 20, 13, 6, and 7 epochs, respectively. These models managed to get the lowest loss validation in the 19th iteration with a training accuracy of 91.42% and validation accuracy of 82.07% for MobileNet, the 13th iteration with a training accuracy of 87.35% and validation accuracy of 78.62% for Xception, the 6th iteration with a training accuracy of 73.11% and validation accuracy of 86.83% for VGG16, the 7th iteration with a training accuracy of 57.47% and validation accuracy of 55.17% for ResNet50. The graph of loss and accuracy of each architecture in each training iteration can be seen in Fig. 10. Table 4 shows training performance results.

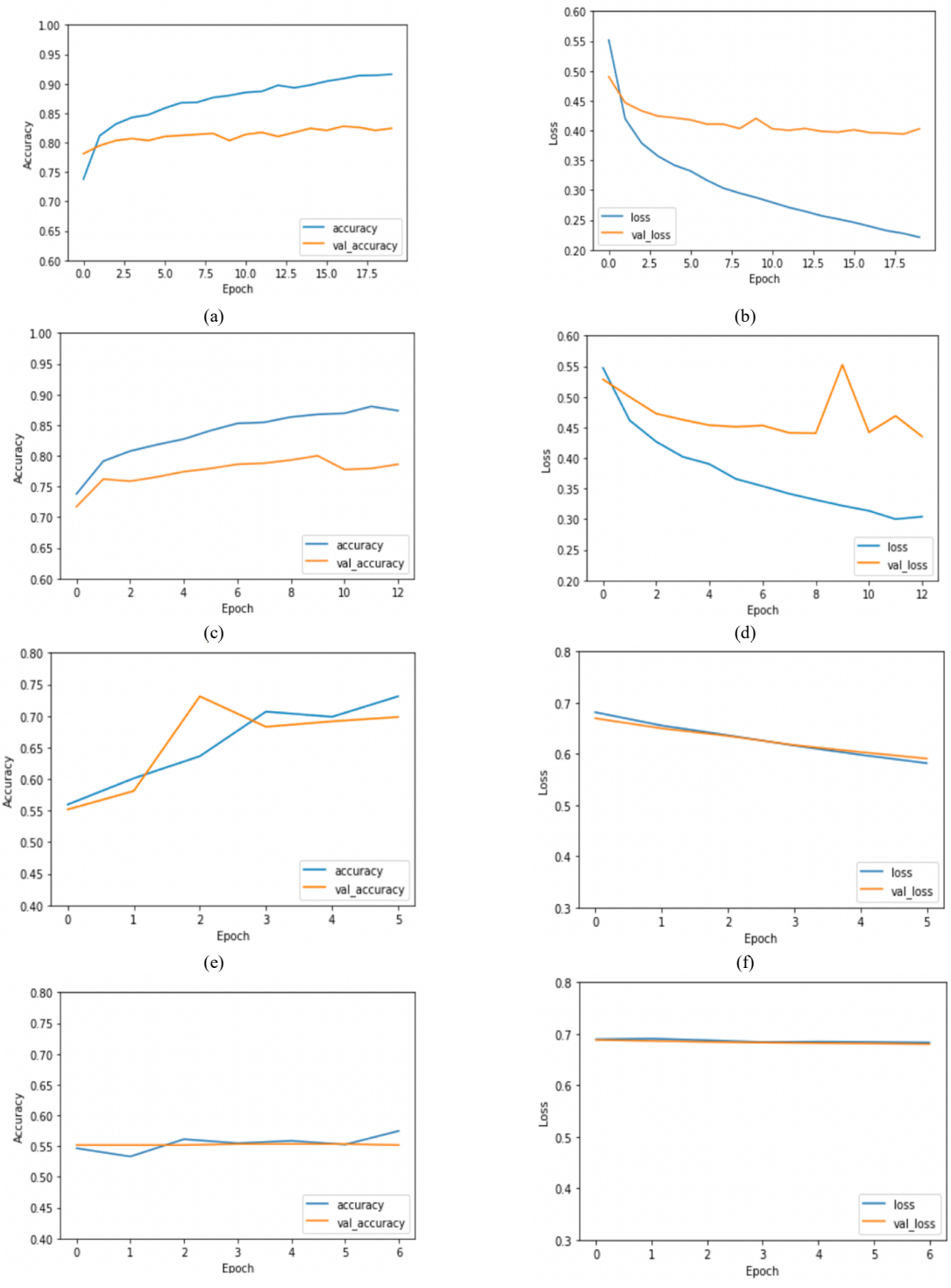

Fig. 10 The accuracy and loss graph of MobileNet (a)(b), Xception (c)(d), VGG16 (e)(f), and ResNet50 (g)(h)

TABLE 4
TRAINING PERFORMANCE

Metrics	MobileNet	Xception	VGG16	ResNet50
Training time/epoch (s)	11	29	22	22
The number of epochs	20	13	6	7
Training Time (s)	220	377	132	154
Training Accuracy	91.42%	87.35%	73.11%	57.47%
Validation Accuracy	82.07%	78.62%	86.83%	55.17%

B. Testing

Table 5 shows the results of model testing on dermoscopic + macroscopic, dermoscopic, macroscopic, and average scores of all experiments. Testing on MobileNet obtained an average F1 score of 84.43%. MobileNet is the best model in this experiment. MobileNet also obtains the best accuracy when compared to other models. However, VGG16 has better precision and specificity values compared to MobileNet, although it has the worst recall value compared to others.

TABLE 5
TESTING PERFORMANCE COMPARISON

	Dataset	Type	Classifier	Accuracy	Precision	Specificity	Recall	F1 Score
Murugan et al. [3]	ISIC [4]	Dermoscopic	SVM	82.31%	-	78.86%	87.76%	-
			KNN	62.19%	-	69.19%	74.75%	-
			Random Forest	71.97%	-	58.66%	65.72%	-
Kavitha et al [5]	Private dataset	Dermoscopic	SVM-RBF	87.3%	78.9%	88.4%	86.2%	84.3%
			KNN	85.2%	77.4%	86.5%	84.1%	84.1%
Hasanah et al [8]	Mendonca [9]	Dermoscopic	-	86.67%	-	-	-	-
Ramezani et al [6]	DermAtlas [7]	Macroscopic	SVM + PCA	82.2%	-	86.93%	77%,	-
Pillay et al. [10]	MED-NODE [17]	Macroscopic	SVM	73.52%	-	-	-	-
			KNN	75.29%	-	-	-	-
Oliveira et al. [11]	Loyola University Dermatology Medical Education [12], YSP Dermatology Image Database [13], and DermAtlas [14]	Macroscopic	SVM	74.36%	75.24%.	-	-	-
Hurtado et al [11]	DermAtlas [14]	Macroscopic	ANN	86.3%	-	86.9%	87.8%	-
Proposed	ISIC [4], MED-NODE [17], and PAD-UFES-20 [18]	Dermoscopic	**MobileNet**	**84.84%**	**81.84%**	**84.16%**	**89%**	**83.71%**
			VGG16	72.12%	80.85%	90.00%	50.66%	62.29%
			ResNet50	55.15%	70.00%	99.16%	99.16%	82.06%
			Xception	85.00%	82.95%	85.55%	84.33%	83.63%
		Macroscopic	**MobileNet**	**89.11%**	**83.76%**	**90.00%**	**88%**	**85.71%**
			VGG16	72.16%	90.00%	97.50%	30.72%	44.31%
			ResNet50	62.94%	70.00%	47.50%	57.15%	62.93%
			Xception	79.90%	73.26%	82.50%	74.29%	73.50%
		Dermoscopic and Macroscopic	**MobileNet**	**85.22%**	**82.00%**	**84.75%**	**86%**	**83.86%**
			VGG16	72.09%	81.12%	90.75%	49.07%	61.15%
			ResNet50	55.80%	62.50%	98.50%	30.09%	40.62%
			Xception	84.53%	82.12%	85.25%	83.64%	82.87%
		Average	**MobileNet**	**86.39%**	**83.00%**	**86.30%**	**87.25%**	**84.43%**
			VGG16	72.12%	83.99%	92.75%	43.48%	55.92%
			ResNet50	57.96%	67.50%	81.72%	62.13%	61.87%
			Xception	83.14%	79.44%	84.43%	80.75%	80.00%

V. DISCUSSION

MobileNet is the best model for classifying melanomas or moles in this experiment. The dataset used in this study is imbalanced, so the F1 score is the most suitable for assessing the model's performance. The number of moles dataset is more than melanoma. VGG16 and ResNet50 cannot detect melanoma, so they have a low recall value. Therefore, the high specificity values of VGG16 and ResNet50 cannot be used as a benchmark for assessing model performance.

ResNet50 and VGG16 can accurately identify moles but not melanomas. Table 5 shows the comparison results of this experiment with previous studies.

MobileNet and Xception have high F1 scores, so they can detect melanoma accurately even though the number of melanoma datasets is less than moles. Therefore, it can be concluded that MobileNet and Xception are suitable models for classifying melanomas and moles. However, MobileNet has the fastest training time per epoch, 11 seconds. MobileNet requires 20 epochs to reach early stopping, so it requires 220 seconds of training time. Xception has a training time of 29 seconds per epoch and requires 13 epochs to achieve early stopping, so the total Xception training time is 377 seconds. It can be concluded that MobileNet is the best model to classify melanoma or mole.

This study has the most complete performance metrics compared to previous studies. This study uses dermoscopic and macroscopic datasets, whereas previous studies used only dermoscopic or macroscopic images. This study is compared to Murugan et al. [3] with the same dermoscopic dataset. The recall value and accuracy did not improve. Meanwhile, compared to Pillay et al. [10] with the same macroscopic dataset. The accuracy of this study is considerably better. It can be concluded that this study has an excellent performance in detecting melanoma in macroscopic images. The limitation of this study is that all previous studies' datasets are not tested due to their public availability. Nevertheless, this study provides information on CNN with transfer learning architectural performance. The method can detect melanoma using dermoscopic and macroscopic images and compares its performance with other methods.

VI. CONCLUSION

This study compared CNN with transfer learning architecture with black-hat pre-processing step and inpainting for detecting melanoma in dermoscopic and macroscopic images. MobileNet and Xception are the best models for classifying melanomas and moles with imbalanced dataset characteristics. MobileNet has an average F1 score of 84.43%, while Xception is 80.00%. However, MobileNet has a faster training computation time of 220 seconds compared to Xception 377 seconds. The results show that MobileNet is suitable for melanoma detection, which has the characteristics of an imbalanced dataset. This study uses MobileNet, which has the best F1 score compared with previous studies.

This study has 13.82% better accuracy compared with the previous study by Pillay et al. [10], which used the same dataset in macroscopic images. This study also has a recall of 2.24%, better than the previous study by Murugan et al. [3], which used the same dataset in dermoscopic images. This shows that this study can address the imbalanced dataset issue and improve the recall and accuracy score of the previous study. In future studies, combining other pre-processing methods with CNN transfer learning architectures can be applied to discover the most appropriate transfer learning model for detecting melanoma. In addition, the oversampling method to balance the dataset can also be applied to improve the classification model's performance.

Author Contributions: *Jessica Millenia*: Methodology, Data collection, Software. *Mohammad Farid Naufal*: Writing the conceptualisation, Methodology, Validation. *Joko Siswantoro*: Supervising, Writing- Reviewing and Editing,

Funding: This research received no specific grant from any funding agency.

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

REFERENCES

[1] A. C. Society, "Key Statistics for Melanoma Skin Cancer", [Online]. Available: https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html

[2] D. K. DePalo, K. M. Elleson, M. J. Carr, P. E. Spiess, and J. S. Zager, "Genitourinary melanoma: An overview for the clinician," *Asian J. Urol.*, Jun. 2022, doi: 10.1016/J.AJUR.2022.01.003.

[3] A. Murugan, S. A. H. Nair, and K. P. S. Kumar, "Detection of Skin Cancer Using SVM, Random Forest and kNN Classifiers," *J. Med. Syst.*, vol. 43, no. 8, 2019, doi: 10.1007/s10916-019-1400-8.

[4] "ISIC Archive." https://www.isic-archive.com/#!/topWithHeader/wideContentTop/main (accessed Jul. 05, 2022).

[5] J. Kavitha, S. A, and N. D, "Melanoma Detection in Dermoscopic Images using Global and Local Feature Extraction," *Int. J. Multimed. Ubiquitous Eng.*, vol. 12, no. 5, pp. 19–28, May 2017, doi: 10.14257/IJMUE.2017.12.5.02.

[6] M. Ramezani, A. Karimian, and P. Moallem, "Automatic Detection of Malignant Melanoma using Macroscopic Images," vol. 4, no. 4, 2014.

[7] "Dermnet.com." https://dermnet.com/ (accessed Aug. 20, 2022).

[8] R. L. Hasanah and D. Riana, "Classification of Dermoscopic Image of Skin Cancer Using the GLCM Method and Multi-SVM Algorithm," *Rekayasa*, vol. 14, no. 3, pp. 407–415, Dec. 2021, doi: 10.21107/REKAYASA.V14I3.12213.

[9] T. Mendonca, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "PH2 - A dermoscopic image database for research and benchmarking," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 5437–5440, 2013, doi: 10.1109/EMBC.2013.6610779.

[10] V. Pillay and S. Viriri, "Skin Cancer Detection from Macroscopic Images," *2019 Conf. Inf. Commun. Technol. Soc.*, pp. 1–9, 2019, doi: 10.1109/ICTAS.2019.8703611.

[11] R. B. Oliveira, N. Marranghello, A. S. Pereira, J. Manuel, and R. S. Tavares, "A computational approach for detecting pigmented skin lesions in macroscopic images," vol. 61, pp. 53–63, 2016, doi: 10.1016/j.eswa.2016.05.017.

[12] "Loyola University Dermatology Medical Education Website." http://www.meddean.luc.edu/lumen/meded/medicine/dermatology/melton/hello1.htm (accessed Aug. 24, 2022).

[13] "YSP Dermatology Image Database." http://ysp.in.coocan.jp/index_eng.htm (accessed Aug. 24, 2022).

[14] "Interactive Dermatology Atlas." https://www.dermatlas.net/ (accessed Jul. 05, 2022).

[15] R. Abdelkader, N. Ramou, M. Khorchef, N. Chetih, and Y. Boutiche, "Segmentation of x-ray image for welding defects detection using an improved Chan-Vese model," *Mater. Today Proc.*, vol. 42, pp. 2963–2967, Jan. 2021, doi: 10.1016/J.MATPR.2020.12.806.

[16] J. Hurtado and F. Reales, "A machine learning approach for the recognition of melanoma skin cancer on macroscopic images," *TELKOMNIKA Telecommun. Comput. Electron. Control*, vol. 19, no. 4, pp. 1357–1368, 2021, doi: 10.12928/TELKOMNIKA.v19i4.20292.

[17] "Dermatology database used in MED-NODE." https://www.cs.rug.nl/~imaging/databases/melanoma_naevi/ (accessed Jul. 07, 2022).

[18] A. Pacheco *et al.*, "PAD-UFES-20: a skin lesion dataset composed of patient data and clinical images collected from smartphones," vol. 1, 2020, doi: 10.17632/ZR7VGBCYR2.1.

[19] M. B. Assad and R. Kiczales, "Deep Biomedical Image Classification Using Diagonal Bilinear Interpolation and residual network," *Int. J. Intell. Networks*, vol. 1, pp. 148–156, Jan. 2020, doi: 10.1016/J.IJIN.2020.11.001.

[20] M. Liao, Y. Q. Zhao, X. H. Wang, and P. S. Dai, "Retinal vessel enhancement based on multi-scale top-hat transformation and histogram fitting stretching," *Opt. Laser Technol.*, vol. 58, pp. 56–62, 2014, doi: 10.1016/J.OPTLASTEC.2013.10.018.

[21] A. H. Khan, D. N. F. A. Iskandar, J. F. Al-Asad, and S. El-Nakla, "Classification of skin lesion with hair and artifacts removal using black-hat morphology and total variation," *Int. J. Comput. Digit. Syst.*, vol. 10, no. 1, 2021, doi: 10.12785/IJCDS/100157.

[22] B. Li, B. Zheng, H. Li, and Y. Li, "Detail-enhanced image inpainting based on discrete wavelet transforms," *Signal Processing*, vol. 189, p. 108278, Dec. 2021, doi: 10.1016/J.SIGPRO.2021.108278.

[23] F. Farhangi, "Investigating the role of data preprocessing, hyperparameters tuning, and type of machine learning algorithm in the improvement of drowsy EEG signal modeling," *Intell. Syst. with Appl.*, vol. 15, p. 200100, Sep. 2022, doi: 10.1016/J.ISWA.2022.200100.

[24] A. Michele, V. Colin, and D. D. Santika, "Mobilenet convolutional neural networks and support vector machines for palmprint recognition," *Procedia Comput. Sci.*, vol. 157, pp. 110–117, 2019, doi: 10.1016/J.PROCS.2019.08.147.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Sep. 2015. Accessed: May 27, 2021. [Online]. Available: http://www.robots.ox.ac.uk/

[26] F. Chollet, "XCeption: Deep Learning with Depthwise Separable Convolutions," *Comput. Vis. Found.*, 2016, doi: 10.4271/2014-01-0975.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.