TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 22, No. 5, October 2024, pp. 1209~1220
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v22i5.26106

Enhancing melanoma skin cancer classification through data augmentation

Mohammed M'hamedi[1], Mohammed Merzoug[1], Mourad Hadjila[2], Amina Bekkouche[1]
[1]LRIT Laboratory, Department of Computer Science, Faculty of Science, University of Abou-Bakr Belkaid, Tlemcen, Algeria
[2]STIC Laboratory, Department of Telecommunications, Faculty of Technology, University of Abou-Bakr Belkaid, Tlemcen, Algeria

Article Info

Article history:

Received Feb 23, 2024
Revised Jul 9, 2024
Accepted Jul 12, 2024

Keywords:

Convolutional neural networks
Data augmentation
Melanoma
MobileNetV2
Skin disease
Transfer learning
Visual geometry group-19

ABSTRACT

Skin cancer is a dangerous and prevalent cancer illness. It is the abnormal growth of cells in the outermost of the skin. Currently, it has received tremendous attention, highlighting an urgent need to address this worldwide public health crisis. The purpose of this study is to propose a convolutional neural network (CNN) to help dermatology physicians in the inspection, identification, and diagnosis of skin cancer. More precisely, we offer an automated method that leverages deep learning techniques to categorize binary categories of skin lesions. Our technique enlarges skin cancer by utilizing data pre-processing and augmentation to address the imbalanced class problem. Subsequently, fine-tuning is conducted on the pre-trained models visual geometry group (VGG-19) and MobileNetV2 to extract and classify the image features using transfer learning. The model is tested on the society for imaging informatics in medicine international skin imaging collaboration (SIIM-ISIC) 2020 dataset and achieved an accuracy of 95.16%, sensitivity of 90.83%, specificity of 99.2%, area under curve (AUC) of 97.57%, and precision of 99.06%. The proposed model based on MobileNetV2 outperforms the other techniques.

Corresponding Author:

Mohammed M'Hamedi
LRIT Laboratory, Department of Computer Science, Faculty of Science, University of Abou-Bakr Belkaid Tlemcen, Algeria
Email: mohamed.mhamedi@univ-tlemcen.dz

1. INTRODUCTION

Skin cancer is one of the most difficult types of cancer to detect and is currently the fifth most common type of cancer [1]. This type of cancer is a serious disease in which abnormal skin cells grow out of control. In general, skin cancer is divided into various types, the most common of which are basal cell carcinoma, squamous cell carcinoma, and melanoma [2].

These are the most frequent, accounting for 90% of all skin malignancies, with basal cells accounting for 75% and squamous cells accounting for 20%. Carcinoma often develops beyond the age of fifty and is easily curable in the majority of instances. Basal cell carcinomas never metastasize, but squamous cell carcinomas do, primarily in lymph nodes around the tumor. Cancers often go through various stages. Squamous cell carcinomas may begin as a localized lesion in the epidermis. A crust (actinic keratosis) or eczema (Bowen's disease) appears on the skin's surface. The deeper infiltration of the dermis characterizes the stage of invasive carcinoma [3].

Melanoma accounts for just 4% of all skin cancer types but is responsible for around 75% of all skin cancer deaths [4]. Melanomas, the deadliest kind of skin cancer, arise from melanocyte cells, which produce melanin, which is responsible for the brown or red coloring of the skin. There are two skin pigmentation types: brown pigments, which provide a tan and some UV protection, and red pigments (red skin), which do not. Subjects who primarily generate red pigments do not tan and are consequently at a higher risk of developing skin cancer. People who live in mountainous areas are more likely to get skin cancer because of their exposure to sunshine and the nature of the color of their skin [5]. Malignant melanoma must be discovered and treated

as soon as possible because it may spread throughout the body and generate metastases that are extremely difficult to cure [6]. The "beauty mark" or mole is a benign lesion caused by an accumulation of melanocytes in the dermis, which explains its brown or red hue. It is critical to do self-examination and consult when there is a change in an existing lesion moles or a new lesion that does not subside after 1 to 2 months. Changes in an existing mole including the size, shape, or color or by the appearance of a new mole are usually the first indicators of melanoma [7]. If identified early enough, melanoma can be successfully treated with a simple surgical procedure, significantly improving the chances of survival. With treatment, patients at this stage have a five- and ten-year overall survival rate of more than 90% [8]. That is why early detection and treatment can dramatically improve the outcome of the disease.

A dermatologist can either confirm or check out a skin cancer diagnosis. The appearance of the lesion is often adequate, but in all questionable situations, a microscopic examination will be required. The clinical ABCD mnemonic (asymmetry, border irregularity, color variegation, and diameter greater than 6 mm), is one of the most frequently used techniques. Unfortunately, some forms of malignant melanoma may not be detectable by the ABCD clinical rule because the problem in detecting melanomas is that their characteristics overlap with those of benign moles [9]. A more effective method for the detection of malignant melanoma is through dermoscopy. Is an invaluable tool for evaluating skin lesions, and it is a more useful technique used by dermatologists that improves the diagnostic accuracy of melanoma [10]. However, some types of melanoma remain difficult to classify, such as those in patients with multiple atypical moles and nodular melanomas. A further technique that can improve the efficiency of skin cancer assessment is the use of AI. Developing convolutional neural networks (CNN) classification models for skin cancer, and melanoma in this context, is motivated by a goal to improve public health, improve healthcare accessibility, lower healthcare costs, and advance medical knowledge, all while employing the power of technology and innovation.

In this study, we propose a deep-learning model for the automatic classification of melanoma. In particular, we leverage CNN with the adoption of transfer learning to boost the overall performance and detect the disease in its early stage. The training CNN model is mostly based on the dataset from which it machine learns. The quality, diversity, and size of the dataset will remain a major factor in the performance and realizability of machine learning models; but in some cases, the dataset itself is sometimes the source of its limitations. The model always requires that the classes be properly balanced and that the dataset be rich enough for the model to learn these features quickly. These limitations of imbalanced classes are being increasingly highlighted in the field of statistical learning for the healthcare domain, and it is necessary to propose good classifiers in the face of these problems. The class imbalance has been the primary cause of the abnormalities in the melanoma dataset, as patients infected with melanoma had a considerably lower number of images than patients with non-melanoma, as illustrated in Table 1.

Table 1. Metadata description

	Not melanoma (benign)	Melanoma (malignant)
No. of images	32542	584
No. of images/sex		
Male	16716	364
Female	15761	220
No. of images/age		
<20	798	8
<40 and >20	10239	89
>40	21437	487
No. of images/anatomical position		
Upper extremity	4872	111
Lower extremity	8293	124
Torso	16588	257
Palms/soles	370	5
Head/neck	1782	74
Oral/genital	120	4
No. of patient	2055	428
No. of patient/sex		
Male	977	168
Female	1076	260
No. of patient/anatomical position		
Upper extremity	1400	93
Lower extremity	1623	108
Torso	1942	207
Palms/soles	195	5
Head/neck	816	67
Oral/genital	77	3

Our proposed solution to this problem is to augment the dataset with new images derived from the original one using flipping, rotations, zoom, and shearing. However, several researchers have conducted various related approaches that focus on these issues: The number of images taken by [5] is equal to 1800 instances, which is 3 times less than the number of images taken in our proposal work. Rashid *et al.* [11] used the same dataset as our work, but they included 4522 malignant melanoma images taken from the ISIC 2019 dataset to remedy the problem of imbalanced classes.

Our contributions can be summarized as follows:

- We propose two CNNs based on pre-trained models, which are VGG-19 and MobileNetV2.
- The last classification layer (sigmoid) contains two neurons: the benign class and the malignant class.
- The issue of imbalanced classes is handled through data augmentation.
- The two CNN models are evaluated using the international skin imaging collaboration (ISIC) dataset.

The remainder of this paper is structured as follows: section 2 summarises the available research studies. Section 3 presents the methodology where we first provide an overview of the SIIM-ISIC 2020 dataset, followed by its preprocessing, then, we explain data augmentation, classification model and network deployment, and training and model assessment. Section 4 covers the suggested model's experimental outcomes including evaluation and comparative research. The last section concludes this paper.

2. RELATED WORKS

Various machine learning techniques [12], [13] and deep learning models have emerged for the analysis of medical images most especially skin lesion images [14]. Hurtado and Reales [5] proposed a new system for classifying skin cancer using images from a standard camera. They explored the impact of smoothed bootstrapping, a technique for augmenting datasets, on classification outcomes. The study compared eight classifiers, including support vector machines (SVM), k-nearest neighbor (KNN), and ANN, with and without data augmentation. Results showed that the ANN with data augmentation achieved the best performance and balance, reaching an area under curve (AUC) of 87.1%. Rashid *et al.* [11] introduced a new deep transfer learning model utilizing MobileNetV2 for melanoma classification, distinguishing between malignant and benign skin lesions. They evaluated the model's effectiveness using the ISIC 2020 dataset and supplemented it with images from the ISIC 2019 dataset, which had a significant class imbalance with less than 2% malignant samples. To address this, they employed various data augmentation techniques to enhance dataset diversity. Experimental results indicated that their deep learning approach outperformed existing techniques in terms of accuracy and computational efficiency. The proposed architecture achieved an impressive diagnostic accuracy of 98.2%, highlighting its effectiveness in melanoma classification. Li *et al.* [15] reviewed 45 research efforts on identifying skin diseases using deep learning technology since 2016. They assess this research in terms of illness kind, dataset, data processing technology, data augmentation technology, skin disease images identification model, deep learning framework, evaluation indicators, and efficiency of the model. Li *et al.* [16] presented an overview of deep learning methods and their applications in diagnosing skin diseases. They began with an introduction to skin problems and the methodologies for gathering dermatological photographs. They examined popular deep learning architectures as well as common frameworks for the implementation of deep learning algorithms. Srinivasu *et al.* [17] proposed a computerized skin disease classification process using deep learning based on long short-term memory (LSTM). The HAM10000 dataset has been used and the proposed method has outperformed other methods with an accuracy of more than 85%.

A novel model has been constructed by Goceri [18] using MobileNet, where a novel hybrid loss function has been used, the suggested technique achieves an accuracy of 94.76%. Adegun and Viriri [19] proposed a deep learning-based method that overcomes limitations for automatic melanoma lesion detection and segmentation. An improved encoder-decoder network with encoder and decoder subnetworks linked by a series of skip paths that propose bringing the semantic level of the encoder feature maps closer to the semantic level of the decoder feature maps. The DL model achieves an accuracy and dice coefficient of 95% and 92% on the ISIC 2017 dataset, and an accuracy and dice coefficient of 95% and 93% on the PH_2 datasets. Hussien and Alasadi [20] proposed a deep learning approach for classifying melanoma skin cancer using a CNN model with 27 layers, which shows promising results. The CNN model is meticulously crafted to extract features from skin lesion images and categorize them into melanoma and non-melanoma classes. This model incorporates multiple convolution layers, batch normalization layers, max-pooling layers, fully connected layers, dropout layers, and data augmentation techniques, all of which contribute to the accuracy and generalization of the model. Experimental findings on publicly accessible benchmark datasets for skin lesion classification demonstrate that the proposed CNN model surpasses existing state-of-the-art approaches. A good accuracy, equal to 99.99%, was achieved in the model's detection. Codella *et al.* [21] proposed a system that combines recent developments in deep learning with established machine learning approaches to create ensembles of methods that are capable of segmenting skin lesions, as well as analyzing the detected area and

surrounding tissue to detect melanoma. When compared to the average of eight expert dermatologists on a subset of 100 test images, the proposed system produces a higher accuracy (76% versus 70.5%), and specificity (62% versus 59%) evaluated at a sensitivity (82%). Li and Shen [22] proposed two deep learning methods to address three main tasks emerging in the area of skin lesion image processing, namely lesion segmentation (task 1), lesion dermoscopic feature extraction (task 2), and lesion classification (task 3). The proposed deep learning frameworks are evaluated on the ISIC 2017 dataset. The experimental results show the promising accuracies of the proposed frameworks, i.e., 0.753 for task 1, 0.848 for task 2, and 0.912 for task 3 were achieved. Kassani and Kassani [23] evaluate the performance of several state-of-the-art CNN. To improve the quality of images, the authors use several pre-processing steps. Both pre-processing and data augmentation could have a positive impact on the final accuracy. Naeem *et al*. [24] collected state-of-the-art research identifying recent research trends, challenges, and opportunities for melanoma diagnosis and explored existing solutions for the diagnosis of melanoma detection using deep learning, to help researchers in the field of melanoma detection, a proposed model, challenges and opportunities have been presented.

Furthermore, Zhang [25] made use of an upgraded neural network framework to achieve prompt feature learning and ideal melanoma image segmentation. A batch normalization layer is used between the convolution layer and the activation layer (such as ReLU or ELU) to solve the problem of gradient disappearance and explosion. Reis *et al.* [26] proposed a deep learning-based CNN named InSiNet where the performance of the model is based on multiple pre-trained models used in this study such as GoogleNet, DenseNet-201, ResNet152V2, and EfficientNetB0. The developed InSiNet architecture managed to achieve a precision of 94.59%, 91.89%, and 90.54% in ISIC 2018, 2019, and 2020 datasets, respectively. Wan *et al*. [27] proposed a multi-scale long attention network (MSLANet) for skin lesion classification, which is composed of three long attention networks (LANet) where each LANet can fuse the context information and improve discriminative representation ability through the long attention mechanism. The results show that the developed MSLANet architecture outperforms the state-of-the-art methods achieving a rank-1 average AUC of 93.7% on the ISIC 2017 dataset and an AUC of 92.4% on the SIIM-ISIC 2020 dataset. Mijwil [28] selected and trained a deep learning network to analyze more than 24,000 skin cancer images using a convolutional neural network (ConvNet) model by applying three architectures InceptionV3, ResNet, and VGG19. To identify the best architectures in the classification of these images as well as attaining exceedingly satisfactory results, many variables are taken into consideration. According to the test results, the best architecture is InceptionV3, which achieved a diagnostic accuracy of roughly 86.90%, precision of 87.47%, sensitivity of 86.14%, and specificity of 87.66%.

3. METHOD

3.1. Dataset description

For our research purposes, we looked at the use of ISIC datasets. The ISIC 2020 dataset [29], which is frequently used, was designed to identify melanoma since it is based on binary classification. The ISIC produced the official dataset of the SIIM-ISIC melanoma classification, which contains 33126 dermoscopic images of benign and malignant skin lesions from over 2056 individuals. These images were then sorted out into two classes as follows: 32542 images representing a percentage of 98.2% for class 0 (benign) and 584 images with a percentage of 1.8% for class 1 (confirmed melanomas). The metadata that is presented with the dataset contains the additional information associated with the dermoscopic images. This metadata provides important context and information about the dataset and can include various details described in Table 1.

3.2. Data pre-processing

The biggest obstacle to accuracy when using deep learning algorithms in the dermatology domain is the lack of datasets. As shown in the previous section, there is an observable imbalance in the dataset that should be taken into consideration. We have found that the distribution of skin lesion classes in the ISIC 2020 dataset is very imbalanced, as can be seen in Table 1. The imbalanced classes can make the model learn bias in favor of the majority sample class more than the minority sample class. To avoid this problem, we have used a data augmentation approach, to increase class 1 (malignant) to be roughly equivalent to class 0 (benign). This is detailed in Figure 1.

After filtering duplicate image names based on unique patient code and anatomical position, the previously obtained 32542 images for the benign class were reduced to no more than 6271. We constructed the input dataset from 6271 images for class 0 (benign class) and 584 for class 1(malignant class), as shown in the chart represented in Figure 1. If we operate with such an initial imbalanced dataset, then bad metrics are guaranteed as results.

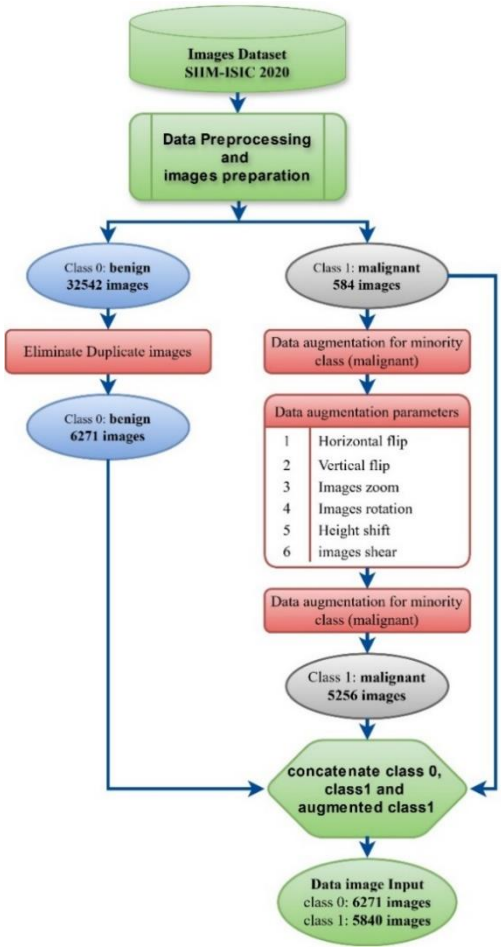
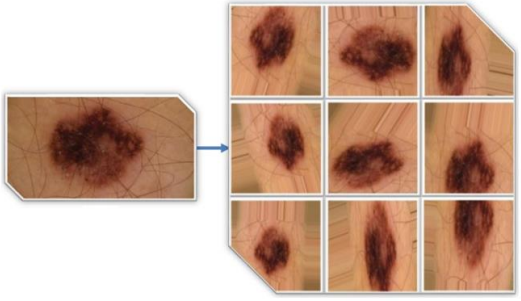


Figure 1. Data augmentation approach

Before applying data augmentation techniques, we have a larger number of samples for the training set, which is often required for CNN models. As a first step, an image generator has been created. The image data generator applies augmentation to each image where six different augmentation parameters are used to the original images: horizontal flip, vertical flip, image zoom range=0.4, image rotation range=17, shear range=0.4, height shift range=0.6

Before image augmentation, the total number of original images was 584, whereas after, it became 5840 augmented images for class 1 (malignant). In total, we have 12111 images, both of class 0 (benign) and class 1 (malignant). From these 584 original class 1 images, the execution of the image data generator is repeated 9 times, which is the number of 5256 (9 times 584) augmented images that are generated. A sample image of the malignant class is shown in Figure 2(a). Figure 2(b) illustrates the outcome of augmenting this image, producing nine variations.



(a) (b)

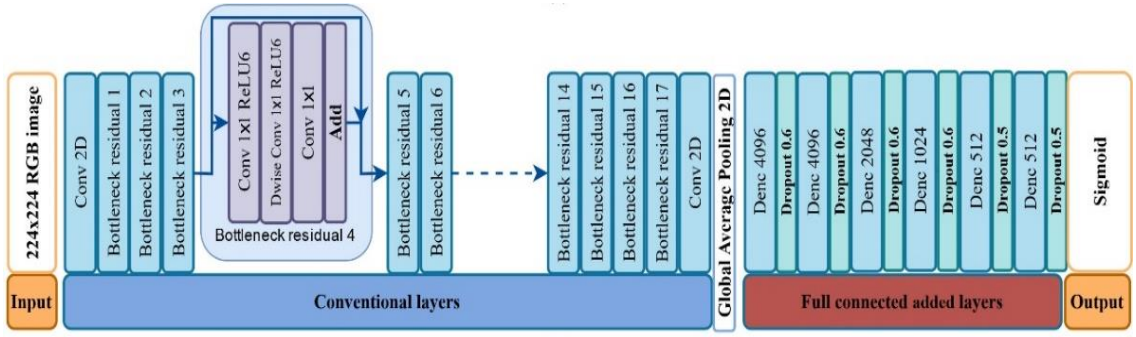
Figure 2. Images from data generator; (a) original image and (b) 9 augmented images

3.3. Classification model and network deployment

Naeem *et al.* [24] CNN is a deep learning algorithm designed for images and video. A CNN has multiple layers, an input layer, hidden layers (feature extraction), a fully connected layer (classification), and an output layer. After the rapid development of deep learning models, in practice, various pre-trained classification models are used on different datasets, such as AlexNet [30], VGG [31], GoogLeNet Tom to diseases classification [32], Inception [33], ResNet [34], and MobileNet [35].

The pre-trained models VGG-19 and MobileNetV2 are developed by Oxford University and Google AI, respectively. The layers and architecture of the VGG-19 network are shown in Figure 3(a), while Figure 3(b) presents the pre-trained architecture of MobileNetV2.

(a)



(b)

Figure 3. Network architecture description; (a) VGG-19-based model and (b) MobilerNetV2-based model

About VGG-19, the points that need to be noted are that it has an input shape (None, 224, 224, 3) and we have four last layers of the output, three fully connected layers include the prediction layer with 1000 values because the database has 1000 classes of images. On the other hand, the MobileNetV2 model has the same input shape as VGG-19 and has two connected layers including 1000 classes of prediction layers.

A softmax of 1000 categories is used as the output layer for both VGG-19 and MobileNetV2 models, we remove this layer and replace it with a binary sigmoid activation output layer since we have a binary classification problem. At first, after freezing the network, we remove the last fully connected layers and replace them with our six defined layers as shown in Figure 3. We also applied the dropout regularization to each of the added layers to avoid overfitting problems.

3.4. Training and model evaluation

The proposed model architecture for skin melanoma classification is shown in Figure 4. It is composed of six main building blocks: image preprocessing (that generates the image data input), Train-Validation-Test dataset split, feature extraction using VGG-19 and MobileNetV2 pre-trained models, added dense output layers, and finally training classification and performance evaluation.

The data collected after the data augmentation technique needs to be divided into three different ensembles: training, validation, and test [36], as shown in Figure 4. In addition, the division image between the two classes was conducted as shown in Table 2. The dataset was split into 80% training data and 20% test data. In addition, we split the training data into 75% of the training set data and 25% of the validation set data. (i.e., 60% of the training set and 20% of validation from the total dataset, Figure 4 demonstrates this setting).

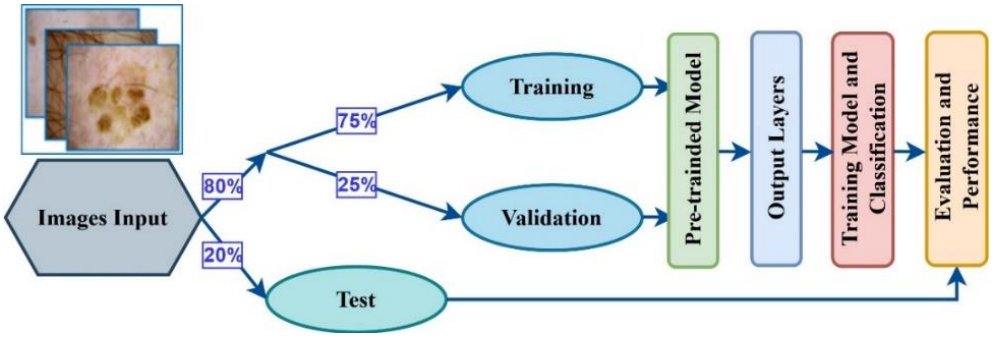


Figure 4. Proposed model for melanoma prediction

Table 2. Splitting dataset into Train, validation, and test sets

Train and validation		Test	
Class (0) benign	Class (1) malignant	Class (0) benign	Class (1) malignant
5017	4672	1254	1168

The classification evaluation is based on a confusion matrix and the following instances:

- True negative (TN): the classifier correctly predicts the originally negative class (benign class).
- False positive (FP): the originally negative class (benign class) is incorrectly predicted, as well as the prediction is positive class.
- True positive (TP): the classifier correctly predicts the originally positive class (melanoma class).
- False negative (FN): the originally positive class (melanoma class) is incorrectly predicted, as well as the prediction is negative class.

A predictive model's performance may be better understood by looking at the confusion matrix, which also shows which classes are properly and mistakenly predicted as well as the kinds of mistakes that occur. In this type of confusion matrix, each cell in Table 3 has a specific and well-understood name, summarized as:

Table 3. Confusion matrix

	Predicted class 0	Predicted class 1
True class 0	TN	FP
True class 1	FN	TP

After training the model, the subsequent stage involves assessing performance using metrics such as accuracy, precision, specificity, and sensitivity. These performance metrics have been detailed in Table 4. There is another useful metric for evaluating the performance of binary classification, especially when the dataset is imbalanced. This metric is called AUC. It quantifies the ability of the model to discriminate between positive and negative classes across all possible thresholds. A higher AUC value is an indication of better model discrimination.

Table 4. Performance metrics

Metric	Formula	Description
Accuracy	$\frac{TP+TN}{\text{total number of samples}}$	The number of correct predictions divided by the total number of predictions.
Precision	$\frac{\text{TP}}{\text{TP}+\text{FP}}$	Evaluate the accuracy of the positive predictions made by a model.
True positive rate (TPR)/sensitivity	$\frac{\text{TP}}{\text{TP+FN}}$ or $1-FNR$	Refer to the rate at which the classifier correctly identifies instances from the positive class.
True negative rate (TNR)/specificity	$\frac{\text{TN}}{\text{TN+FP}}$ or $1-FPR$	Refer to the rate at which the classifier correctly identifies instances from the negative class.
False positive rate (FPR)	$\frac{\text{FP}}{\text{FP+TN}}$ or $1-TPR$	The model incorrectly predicts negative classes as positive classes.
False negative rate (FNR)	$\frac{\text{FN}}{\text{FN+TP}}$ or $1-TNR$	The model incorrectly predicts positive classes as negative classes.

4. RESULTS AND DISCUSSION

This section explains the results of our research and at the same time gives a comprehensive discussion. The experiment was carried out on Kaggle. The model has been implemented on the TensorFlow platform using the open-source Keras packages and the python programming language. In addition, it used the Adam optimizer with a learning rate equal to 0.0001 and the Binary Focal Crossentropy loss function. Our classification model is based on VGG-19 and MobileNetV2 architectures (see Figure 3), and the input images dataset distribution (training and validation see Figure 4) are described in the classification model and network deployment section.

4.1. Validation and training result

Figure 5 represents the accuracy and loss of our proposed CNN model based on VGG-19 and MobileNetV2 on the training and validation sets as training progresses through each epoch. The training accuracy increases steadily throughout the training process, while the validation accuracy may plateau or even decrease after a certain number of epochs.

Initially, both training and validation accuracies have low values. After epoch 5, the accuracies improve during this phase and the loss will continue to decrease as the model adjusts its parameters to provide a better fit to the data. For the later epochs, the rate of improvement in both accuracy and loss is slowed down and finally stabilizes at the end. According to Figure 6 (accuracy part), we obtained a validation rate of 93.11% using the model based on VGG-19 and a validation rate of 94.58% using the model based on MobileNetV2.

We also show in Figure 5 that the MobileNetV2 model is less sensitive to overfitting with respect to the VGG-19 model; this is mainly demonstrated by the weak zig-zag of the accuracy-validation curve of MobileNetV2 (while the VGG-19 has a curve with a larger zig-zag, as compared with that of MobileNetV2). Furthermore, the optimal stopping iteration for the MobileNetV2 model belongs to the interval (60, 80); this is mainly supported by the fact that both the validation and the training curves for accuracy are almost stagnant.

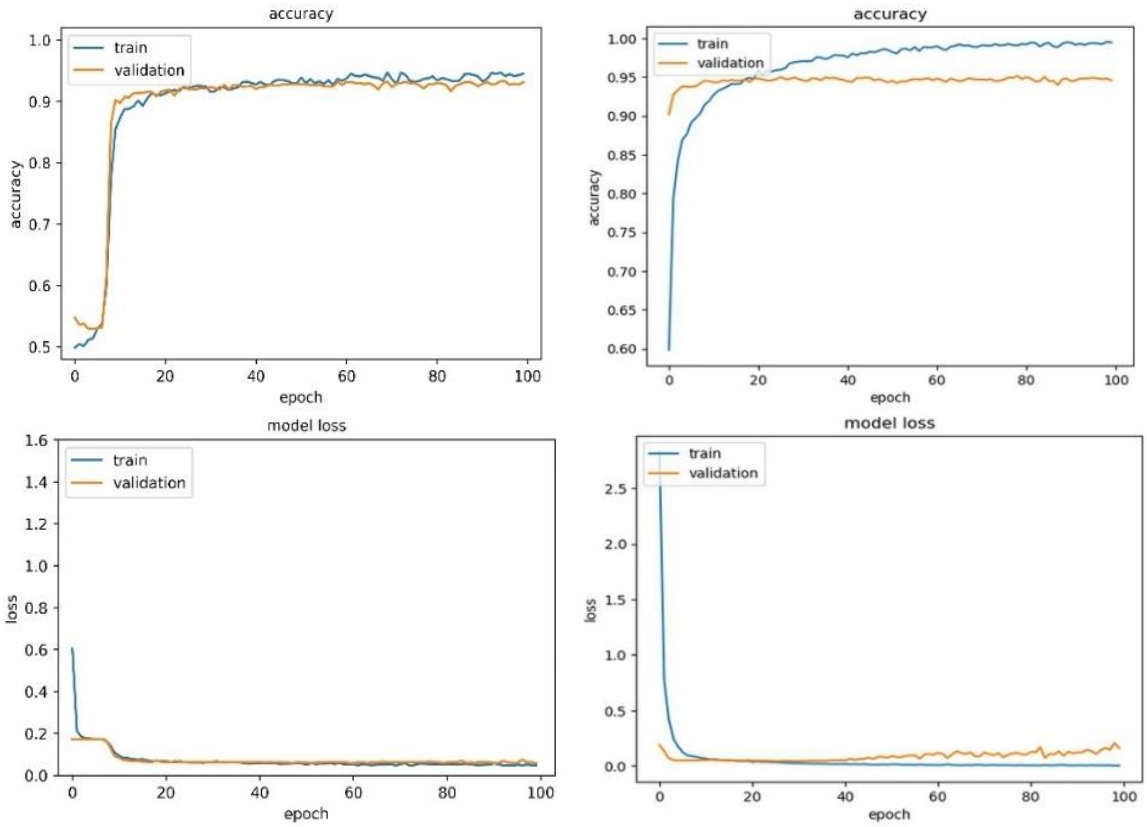
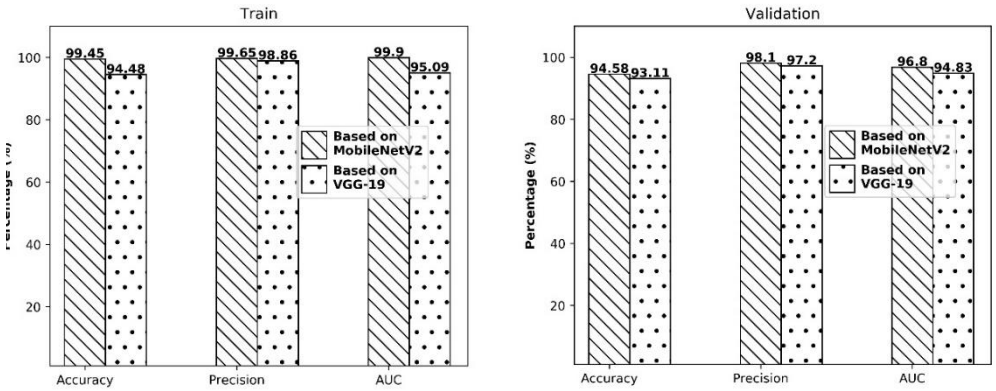


Figure 5. Accuracy/loss of VGG-19/MobileNetV2-based models vs epochs

Figure 6 also shows a detailed comparison between the VGG-19-based model and the MobileNetV2-based model. Three metrics were considered: model accuracy, AUC, and precision. We mainly observe a slight

difference between the two models for both AUC and precision performance on the training shown in Figure 6(a) and validation sets shown in Figure 6(b). More specifically, when we use the VGG-19-based model, we obtain a validation AUC of 94.83% and a precision of 97.2%. However, the performance of the MobileNetV2-based model improves even more, with a validation AUC equal to 96.8% and a precision equal to 98.1%. This result confirms the superiority of the MobileNetV2-based model over the other one in the context of melanoma diagnosis.

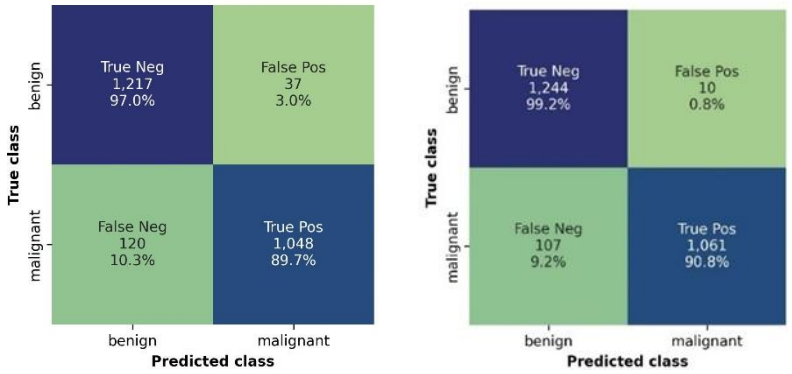


(a) (b)

Figure 6. Performance metrics for all models; (a) training and (b) validation

4.2. Evaluation and comparative study

The performance of the proposed models is reported in terms of accuracy, precision, AUC, sensitivity, and specificity, and was evaluated on a test set of 2422 images from two classes. The overall performance achieved by the model is summarized as follows, where the confusion matrix is illustrated in Figures 7, Figure 7(a) was identical to Figure 7(b) and the evaluation and performance of both state-of-the-art methods and our proposed approach are detailed in Table 5.



(a) (b)

Figure 7. Confusion matrix of; (a) VGG-19-based model and (b) MobileNetV2-based model

From the confusion matrix, we show how well the model can classify instances into benign and malignant classes. The correct predictions were shown on the diagonal in the confusion matrix. It showed that the model of the proposed approach was correctly identified. In the case of VGG-19, it correctly classified 1217 out of 1254 benign images and 1048 out of 1168 malignant images. If using MobileNetV2, we achieved the best results, accurately classifying 1244 of 1254 benign images and 1061 of 1168 malignant images.

To properly assess our suggested approach, we compared the performance metrics of our models to existing state-of-the-art melanoma ISIC 2020 image classification algorithms [26]-[28]. The contrast can be illustrated in Table 5.

Table 5. Performance of the proposed models vs. state-of-the-art techniques

	Accuracy (%)	AUC (%)	Precision (%)	Sensitivity (%)	Specificity (%)
Reis *et al*. [26]	90.54	--	--	93.33	88.64
Wan *et al*. [27]	95.6	92.4	--	59.6	97.4
Mijwil [28]	86.90	--	87.47	86.14	87.66
Based on VGG-19	93.51	95.20	96.58	89.72	97.04
Based on MobileNetV2	95.16	97.57	99.06	90.83	99.2

Hence, in terms of accuracy our model outperformed models [26], [28]. Notably, our proposed network achieved a remarkable AUC score of 95.20, surpassing the benchmark from [27] by 3%. The precision, too, played its part with a resounding 9% improvement over the model [28]. Sensitivity stood tall at 90.83%, as compared in [26]–[28] and finally, yet importantly, specificity was at 99.2%, outperforming not only [26] but also [28].

5. CONCLUSION

Skin cancer, especially melanoma, has emerged as one of the world's leading growing diseases; this will inevitably lead to death. The importance of early detection of melanoma has been to initiate treatment as early as possible to have a chance of a successful cure. This paper has dealt with the critical task of identifying and categorizing melanoma skin cancer using the SIIM-ISIC 2020 dataset and a deep CNN. We have proposed to implement the data augmentation technique to resolve the class imbalance problem in the ISIC 2020 dataset, which has been a common issue for research workers applying supervised learning techniques. It has also proposed a unique model for image classification that had been adapted to our particular application, as well as fine-tuned and applied variants to the VGG-19 and MobileNetV2 pre-trained neural networks. The suggested approach has been thoroughly tested, taking into consideration important performance variables such as accuracy, precision, specificity, and sensitivity, as well as AUC. The results of the experiment confirm the efficiency of our model since the proposed approach that has been employed by using data augmentation technique and implementing our model improvements has yielded remarkable results when using the VGG-19 architecture, with a test accuracy of 93.51%, an AUC of 95.20%, a precision of 96.58%, a sensitivity of 89.72%, and a specificity of 97.04%. Notably, for the MobileNetV2 architecture, the model has produced even better results, with an accuracy of 95.16%, an AUC of 97.57%, a precision of 99.06%, a sensitivity of 90.83%, and a specificity of 99.2%. In the proposed approach, it has been shown that the model based on MobilenetV2 is better than the model based on VGG-19, and also that it has been better than existing works in terms of accuracy, AUC, precision, sensitivity, and specificity. This study underscores the potential of deep learning in improving melanoma diagnosis accuracy, thereby enhancing patient care. Recommendations for future research focus on the adoption of a multi-model approach. This strategy harnesses various machine learning algorithms strengths to enhance precision and accuracy in melanoma classification tasks.

ACKNOWLEDGEMENTS

We would like to acknowledge our great LRIT Laboratory and STIC Laboratory staff members, for all the support and help in this study; this paper is a part of approved Ph.D. thesis at Department of Computer Science, Faculty of Science, Tlemcen University, Algeria. The study was based on a dataset approved by the Institutional by the international skin imaging collaboration (ISIC) and the Society for Imaging Informatics in Medicine (SIIM).

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, Oct. 2023, doi: 10.1186/s12943-023-01854-3.

[2] N. H. Khan *et al.*, "Skin cancer biology and barriers to treatment: Recent applications of polymeric micro/nanostructures," *Journal of Advanced Research*, vol. 36, pp. 223–247, Feb. 2022, doi: 10.1016/j.jare.2021.06.014.

[3] C. Karimkhani, L. N. Boyers, R. P. Dellavalle, and M. A. Weinstock, "It's time for 'keratinocyte carcinoma' to replace the term 'nonmelanoma skin cancer,'" *Journal of the American Academy of Dermatology*, vol. 72, no. 1, pp. 186–187, Jan. 2015, doi: 10.1016/j.jaad.2014.09.036.

[4] A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, A. Pitiot, and C. Wang, "Fusing fine-tuned deep features for skin lesion classification," *Computerized Medical Imaging and Graphics*, vol. 71, pp. 19–29, Jan. 2019, doi: 10.1016/j.compmedimag.2018.10.007.

[5] J. Hurtado and F. Reales, "A machine learning approach for the recognition of melanoma skin cancer on macroscopic images," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 19, no. 4, pp. 1357-1368, Aug. 2021, doi: 10.12928/telkomnika.v19i4.20292.

[6] E. Okur and M. Turkan, "A survey on automated melanoma detection," *Engineering Applications of Artificial Intelligence*, vol. 73,

pp. 50–67, Aug. 2018, doi: 10.1016/j.engappai.2018.04.028.
[7] D. Schadendorf *et al.*, "Melanoma," *Nature Reviews Disease Primers*, vol. 1, no. 1, Apr. 2015, doi: 10.1038/nrdp.2015.3.
[8] C. M. Balch *et al.*, "Final Version of the American Joint Committee on Cancer Staging System for Cutaneous Melanoma," *Journal of Clinical Oncology*, vol. 19, no. 16, pp. 3635–3648, Aug. 2001, doi: 10.1200/jco.2001.19.16.3635.
[9] C. Papageorgiou *et al.*, "Dermoscopy of difficult-to-diagnose Melanomas," *Serbian Journal of Dermatology and Venereology*, vol. 8, no. 3, pp. 121–127, Sep. 2016, doi: 10.1515/sjdv-2016-0011.
[10] C. Longo *et al.*, "Dermoscopy of melanoma according to different body sites: Head and neck, trunk, limbs, nail, mucosal and acral," *Journal of the European Academy of Dermatology and Venereology*, vol. 37, no. 9, pp. 1718–1730, May 2023, doi: 10.1111/jdv.19221.
[11] J. Rashid *et al.*, "Skin Cancer Disease Detection Using Transfer Learning Technique," *Applied Sciences*, vol. 12, no. 11, p. 5714, Jun. 2022, doi: 10.3390/app12115714.
[12] E. Vocaturo, D. Perna, and E. Zumpano, "Machine Learning Techniques for Automated Melanoma Detection," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, Nov. 2019, doi: 10.1109/bibm47256.2019.8983165.
[13] P. Johannet *et al.*, "Using Machine Learning Algorithms to Predict Immunotherapy Response in Patients with Advanced Melanoma," *Clinical Cancer Research*, vol. 27, no. 1, pp. 131–140, Jan. 2021, doi: 10.1158/1078-0432.ccr-20-2415.
[14] A. Adegun and S. Viriri, "Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art," *Artificial Intelligence Review*, vol. 54, no. 2, pp. 811–841, Jun. 2020, doi: 10.1007/s10462-020-09865-y.
[15] L.-F. Li, X. Wang, W.-J. Hu, N. N. Xiong, Y.-X. Du, and B.-S. Li, "Deep Learning in Skin Disease Image Recognition: A Review," *IEEE Access*, vol. 8, pp. 208264–208280, 2020, doi: 10.1109/access.2020.3037258.
[16] H. Li, Y. Pan, J. Zhao, and L. Zhang, "Skin disease diagnosis with deep learning: A review," *Neurocomputing*, vol. 464, pp. 364–393, Nov. 2021, doi: 10.1016/j.neucom.2021.08.096.
[17] P. N. Srinivasu, J. G. SivaSai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, "Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM," *Sensors*, vol. 21, no. 8, p. 2852, Apr. 2021, doi: 10.3390/s21082852.
[18] E. Goceri, "Diagnosis of skin diseases in the era of deep learning and mobile technology," *Computers in Biology and Medicine*, vol. 134, p. 104458, Jul. 2021, doi: 10.1016/j.compbiomed.2021.104458.
[19] A. A. Adegun and S. Viriri, "Deep Learning-Based System for Automatic Melanoma Detection," *IEEE Access*, vol. 8, pp. 7160–7172, 2020, doi: 10.1109/access.2019.2962812.
[20] M. A. Hussien and A. H. H. Alasadi, "Classification of melanoma skin cancer using deep learning approach," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 22, no. 1, pp. 129-137, 2024, doi: 10.12928/telkomnika.v22i1.25428.
[21] N. C. F. Codella *et al.*, "Deep learning ensembles for melanoma recognition in dermoscopy images," *IBM Journal of Research and Development*, vol. 61, no. 4/5, pp. 5:1-5:15, Jul. 2017, doi: 10.1147/jrd.2017.2708299.
[22] Y. Li and L. Shen, "Skin Lesion Analysis towards Melanoma Detection Using Deep Learning Network," *Sensors*, vol. 18, no. 2, p. 556, Feb. 2018, doi: 10.3390/s18020556.
[23] S. H. Kassani and P. H. Kassani, "A comparative study of deep learning architectures on melanoma detection," *Tissue and Cell*, vol. 58, pp. 76–83, Jun. 2019, doi: 10.1016/j.tice.2019.04.009.
[24] A. Naeem, M. S. Farooq, A. Khelifi, and A. Abid, "Malignant Melanoma Classification Using Deep Learning: Datasets, Performance Measurements, Challenges and Opportunities," *IEEE Access*, vol. 8, pp. 110575–110597, 2020, doi: 10.1109/access.2020.3001507.
[25] X. Zhang, "Melanoma segmentation based on deep learning," *Computer Assisted Surgery*, vol. 22, no. sup1, pp. 267–277, Oct. 2017, doi: 10.1080/24699322.2017.1389405.
[26] H. C. Reis, V. Turk, K. Khoshelham, and S. Kaya, "InSiNet: a deep convolutional approach to skin cancer detection and segmentation," *Medical & Biological Engineering & Computing*, vol. 60, no. 3, pp. 643–662, Jan. 2022, doi: 10.1007/s11517-021-02473-0.
[27] Y. Wan, Y. Cheng, and M. Shao, "MSLANet: multi-scale long attention network for skin lesion classification," *Applied Intelligence*, vol. 53, no. 10, pp. 12580–12598, Sep. 2022, doi: 10.1007/s10489-022-03320-x.
[28] M. M. Mijwil, "Skin cancer disease images classification using deep learning solutions," *Multimedia Tools and Applications*, Apr. 2021, doi: 10.1007/s11042-021-10952-7.
[29] ISIC, "The ISIC 2020 Challenge Dataset," 2020. [Online]. Available: https://challenge2020.isic-archive.com/
[30] M. Syarief and W. Setiawan, "Convolutional neural network for maize leaf disease image classification," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 3, pp. 1376-1381, Jun. 2020, doi: 10.12928/telkomnika.v18i3.14840.
[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
[32] L. Aversano, M. L. Bernardi, M. Cimitile, M. Iammarino, and S. Rondinella, "Tomato diseases Classification Based on VGG and Transfer Learning," in *2020 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*, IEEE, Nov. 2020, doi: 10.1109/metroagrifor50201.2020.9277626.
[33] Y. Ibrahim, H. Wang, and K. Adam, "Analyzing the Reliability of Convolutional Neural Networks on GPUs: GoogLeNet as a Case Study," in *2020 International Conference on Computing and Information Technology (ICCIT-1441)*, IEEE, Sep. 2020, doi: 10.1109/iccit-144147971.2020.9213804.
[34] C. Peng, Y. Liu, X. Yuan, and Q. Chen, "Research of image recognition method based on enhanced inception-ResNet-V2," *Multimedia Tools and Applications*, vol. 81, no. 24, pp. 34345–34365, May 2022, doi: 10.1007/s11042-022-12387-0.
[35] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *Computer Vision and Pattern Recognition*, vol. 14, no. 2, pp. 53–57, 2009, doi: 10.48550/arXiv.1704.04861.
[36] H. L. Dawson, O. Dubrule, and C. M. John, "Impact of dataset size and convolutional neural network architecture on transfer learning for carbonate rock classification," *Computers & Geosciences*, vol. 171, p. 105284, Feb. 2023, doi: 10.1016/j.cageo.2022.105284.

BIOGRAPHIES OF AUTHORS

Mohammed M'Hamedi received an Engineering degree in Telecommunication from Abou-Bekr Belkaid University of Tlemcen, Algeria, in 2007, and received a Magister's degree in Computer Science and Telecommunication Engineering in 2009, from the same university. Currently Ph.D. candidate in the Department of Computer Science, Faculty of Sciences, University of Abou-Bakr Belkaid, Tlemcen, in Algeria. His research interests include IoT, e-health, artificial intelligence, and image processing. He can be contacted at email: mohamed.mhamedi@univ-tlemcen.dz.

Mohammed Merzoug received an Engineering degree in Computer Science from Abou-Bekr Belkaid University of Tlemcen, Algeria, in 2003, and an M.S. degree in Information and Communication Sciences in 2009, from the same university. After that, he joined the service-oriented computing group at LRIT Laboratory, Tlemcen, Algeria, where he started his Ph.D. thesis. In Juin, 2018, he obtained a Ph.D. degree in computer science. His research interests include artificial intelligence, deep learning, and image processing. He can be contacted at email: mohammed.merzoug@univ-tlemcen.dz.

Mourad Hadjila received his Ph.D. degree in Telecommunications in 2014 at the University of Tlemcen, Algeria. Since 2002, he has worked as a lecturer and member of the STIC laboratory at the same university. He has been habilitated to supervise research since 2019 in telecommunications. His research interests are in IoT, intrusion detection systems, artificial intelligence, machine learning, deep learning, and wireless sensor networks. He can be contacted at email: mourad.hadjila@univ-tlemcen.dz.

Amina Bekkouche received an Engineering degree in Computer Science from DjillaliLiabes of Sidi Bel Abbes, Algeria, in 2004, and an M.S. degree in Information and Communication Sciences in 2012, from Abou-Bekr Belkaid University of Tlemcen, Algeria. After that, she joined the service-oriented computing group at the Computer Science Research Laboratory, Sidi Bel Abbes, Algeria, where she started her Ph.D. thesis. In July 2018, she obtained a Ph.D. degree in Computer Science from the University of Tlemcen, Algeria. Her research interests include artificial intelligence, semantic web, service-oriented computing, IoT, ontology engineering, context-aware computing, and deep learning. She can be contacted at email: bekkouche@univ-tlemcen.dz.