Skin Lesion Classification towards Melanoma Detection Using EfficientNetB3

Saumya Salian*, Sudhir Sawarkar

Department of Computer Engineering, Datta Meghe College of Engineering, Mumbai University, Mumbai, India

Received 16 February 2022; received in revised form 27 April 2022; accepted 01 May 2022

DOI: https://doi.org/10.46604/aiti.2022.9488

Abstract

The rise of incidences of melanoma skin cancer is a global health problem. Skin cancer, if diagnosed at an early stage, enhances the chances of a patient's survival. Building an automated and effective melanoma classification system is the need of the hour. In this paper, an automated computer-based diagnostic system for melanoma skin lesion classification is presented using fine-tuned EfficientNetB3 model over ISIC 2017 dataset. To improve classification results, an automated image pre-processing phase is incorporated in this study, it can effectively remove noise artifacts such as hair structures and ink markers from dermoscopic images. Comparative analyses of various advanced models like ResNet50, InceptionV3, InceptionResNetV2, and EfficientNetB0-B2 are conducted to corroborate the performance of the proposed model. The proposed system also addressed the issue of model overfitting and achieved a precision of 88.00%, an accuracy of 88.13%, recall of 88%, and F_1-score of 88%.

Keywords: malignant, skin lesion, deep learning, classification

1. Introduction

Malignant skin cancer is wreaked due to anomalous expansion of melanocyte skin cells and causing tumors to form. Tumors can be malignant or benign in nature. Malignant tumors are a threat to human life. Skin cancer generally occurs in skin that is exposed to sunlight. The high threat factor causing any type of skin cancer is exposure to natural or artificial ultraviolet light. Out of 100 different types of cancer, skin cancer is considered the most prevalent and lethal category of cancer worldwide. In America, more than 9500 people are detected with skin cancer every day [1]. The number of detected skin cancer cases gradually increased to 44 percent from 2011 to 2021. According to National Cancer Institute (NIH), around 106110 skin melanoma cases are estimated in 2022 [2].

Medical experts like dermatologists examine skin lesions using a special magnifying lens known as dermatoscopy [2]. Other imaging tests like CT scans, X-Ray, and MRI are also used to understand the metastases of pigmented skin cells. Visual examination of skin lesions using dermatoscopy is a method followed by medical experts, and its prognosis usually relies on their experience. Skin cancer, if discovered at a preliminary stage, will increase the survival rate among the patients. Hence, it is vital to build a diagnostic system based on a deep learning network to detect malignant categories of skin cancer.

Building a computer-based diagnostic system will support medical practitioners to take advantage of technological overtures and help them to have a second opinion. Since 2015, several deep learning architectures have been explored to build an automated diagnostic system that is forced to play a fundamental contribution in the timely discovery of malignant cancer [3]. Convolutional Neural Network (CNN) model serves a significant part in medical image analysis. With diversified CNN architectures, it becomes arduous to select the apposite model for melanoma classification. Choosing the right model will aid in developing an accurate melanoma skin lesion classification model.

* Corresponding author. E-mail address: srs.cm.dmce@gmail.com

Transfer learning is an approach to utilizing the learning acquired by a model that is trained and built on a peculiar target and constructs a solution for a similar target. Most of the pre-trained models are trained over the ImageNet dataset. ImageNet consists of over 15 million diverse labeled images with 1000 classes. Fine-tuning pre-trained models is a prerequisite to adjusting these models to the target domain of malignant and benign lesion classification. The fundamental weights of the pre-trained models are fine-tuned to adapt to the two-class classification task.

Many of the research papers addressed the problem of noise artifacts like the presence of hair, low contrast images, etc. Very few articles addressed the issue of ink markers in lesion images. When building deep learning models, these ink markings may be mistaken for skin lesions and result in incorrect interpretations. In the proposed model, an automated image preprocessing method is employed that effectively eliminates both surgical ink markers and hair artifacts from the ISIC 2017 dataset. In this work, a deep learning-based, fine-tuned EfficientNetB3 skin lesion classification model is proposed that classifies lesion images into malignant and benign classes. To improve model performance and reduce model overfitting, various data augmentation methods along with global average pooling GAP) and a fully connected classification layer using softmax are incorporated into the proposed model.

In this paper, an experimental evaluation of the proposed approach with other advanced models is carried out to review the potency of the proposed model. All the experimental analyses are carried out on ISIC 2017 dataset [4]. Evaluation metrics like F_1-score, accuracy, recall, and precision are computed. Empirical findings testify to the efficiency of the proposed model in comparison to other pre-trained models and also deliver favorable outcomes which address the problem of model overfitting. The contribution of the work is listed below:

(1) An automated image preprocessing model that removes noise artifacts like thin and thick hair structures and surgical ink markers from lesion images is presented in this study.

(2) Fine-tuned EfficientNetB3 deep learning model is proposed to build an efficient computer-based diagnostic system for improved melanoma classification.

(3) To achieve better accuracy and overcome the drawback of model overfitting, data augmentation techniques are employed, and a custom layer of GAP is exerted over the training and testing phase.

(4) To review the efficacy of the proposed design, comparative experimental analyses with other pre-trained deep learning models are carried out.

The paper is illustrated in the following way, related recent works are stated in Section 2, and a detailed explanation of the proposed methodology is described in Section 3. Experimental results and analysis are outlined in Section 4, and the paper is inferred in Section 5.

2. Related Work

Naronglerdrit et al. [5], offered an experimental study of diverse pre-trained transfer learning models for the classification of malignant skin lesions. Using various pre-trained models, the authors carried out tasks like pre-processing (hair removal), lesion segmentation, batch normalization, and melanoma classification. The experimental analyses noted that ResNet-101 achieved better sensitivity (recall) of 85.18% and accuracy of 97.12%.

Siddique et al. [6], furnished an image segmentation model attributed to the deep learning U-Net framework along with a pre-trained EfficientNet model. To enhance gradient learning and build a deeper U-Net model, residual connection and recurrent feedback with EfficientNet as an encoder was proposed. The proposed model achieved higher segmentation performance with a Jaccard index of 95.34% and a Dice coefficient of 88.62%.

Chaturvedi et al. [7], furnished a skin cancer classification method using MobileNet. Experiments were carried out on the HAM10000 dataset, and pre-processing approaches like image rescaling and data augmentation were applied to the dataset. The MobileNet model achieved an overall accuracy of 83.15%, top2 accuracy of 91.36%, and top3 accuracy of 95.84%. The precision, recall, and F_1-score of the model were 89%, 83%, and 83%, respectively.

Zhang [8], presented the EfficientNet-B6 model for melanoma detection on the ISIC dataset. To assess model performance, the proposed model was compared with other standard models like VGG16 and VGG19. Training and testing of the model were done for 22 epochs with a batch size of 32. EfficientNet-B6 obtained an AUC-ROC score of 91.7, whereas VGG16 and VGG19 achieved a score of 89.1% and 90.2%, respectively.

Zhang and Wang [9], proposed a DenseNet201-based melanoma recognition model for lesion images. All the investigations were carried out on the ISIC dataset from the Kaggle challenge. Training of the proposed model was carried out for 20 epochs with a batch size of 8 using Adam optimizer and a learning rate of le-4. DenseNet201 model performance was compared with VGG16 and ResNet50 over the AUC-ROC score. DenseNet201 achieved a better AUC-ROC score of 92.5 as compared to VGG16 and ResNet50.

Ashim et al. [10], reviewed diverse pre-trained models such as VGG16, ResNet50, EfficientNet, DenseNet, and Xception for lesion classification over the Kaggle dataset. The analyses were carried out on only 660 images of skin lesions, including 360 of class benign and 300 of type malignant. To handle the low precision problem, data augmentation techniques like rotation, crop, compression, brightness, and contrast were applied during the training of the model. From the analyses, ResNet50 furnished better results with a training accuracy of 88.61%, whereas EfficientNetB0 furnished a training accuracy of 78.41%.

Chen et al. [11], proposed an EfficientNetB1-based deep learning-based model using the CycleGAN data augmentation technique to boost skin lesion classification accuracy. CycleGAN approach aided in creating additional training images with labeled information and helped in saving costs in manual labeling. EfficientNet-B1 with CycleGAN data augmentation accomplished an accuracy of 94.5%.

Le et al. [12], exhibited a deep learning framework that classifies skin lesions into seven different classes. The authors carried out the training and testing of the network on the HAM10000 dataset by removing duplicate images from the dataset. To yield better performance, the ResNet50 classifier model architecture was modified by adding an average pooling and a dropout layer of 0.5, along with fine-tuning their weights. The proposed ResNet model achieved an average accuracy of 93%, precision of 81%, recall, and F_1-score of 80%, which outperformed other base models like VGG16, EfficientNetB1, and MobileNet.

Manzo and Pellino [13], presented an ensemble deep learning architecture with a transfer learning approach to extract features from images. Imbalance class datasets were addressed for the task of classification of melanoma. Pre-trained models, namely ResNet-50, AlexNet, and GoogleNet were adopted to extract features from the MED-NODE dataset, which achieved an accuracy of 0.90. Multiple image representations were designed to extract features built on a deep neural network for the correct classification of a melanoma lesion.

Kadampur and Riyaee [14], gave a cloud deep neural learning framework to predict skin cancer with improved accuracy. Deep learning studio (DLS) provided a menu-driven option to construct suitable higher convolutional neural networks with deep layers such as normalization, pooling, dropout, and flattening. A comparative assessment of the proposed model with diverse pre-trained models like ResNet, DenseNet, SqueezeNet, and InceptionNet was performed on the HAM10000 dataset. The proposed model performed better with an area under the curve value of 0.99.

Acosta et al. [15], reviewed a diverse list of state-of-art methods for melanoma classification over the ISIC challenge 2017 dataset. The proposed model incorporated the ResNet152 model with various data augmentation techniques like rotation,

random flip, random zoom, and contrast enhancement. The ResNet152 model was compared with 20 other methods proposed by other researchers over ISIC 2017 dataset and achieved the highest accuracy of 87.2%, sensitivity of 82%, and F_1-score of 84.8%.

Rezaoana et al. [16], proposed a convolution neural network (CNN) model using the transfer learning technique to classify the lesion images into benign and malignant classes. The proposed model was trained on the Kaggle ISIC dataset and various augmentation techniques such as shear range, horizontal flip, rotation, and image zooming. The proposed model based on parallel convolution feature blocks achieved a weighted average accuracy of 79.45%.

3. Proposed Methodology

In this section, a detailed explanation of the proposed methodology for malignant skin lesion classification is provided. The design methodology consists of the following subsections: (1) Data Preprocessing, (2) Data Augmentation, and (3) Fine-Tuned EfficientNetB3 Model Architecture.

3.1. Dataset preprocessing

All the experimental research was carried out on the ISIC dataset [4] available from "2017 ISBI Challenge on Skin lesion Analysis Towards Melanoma Detection". The dataset comprised 3297 images of benign and malignant skin lesions. Lesion images in the ISIC dataset are in RGB color space with varied pixel sizes in the range 540×722 and 4499×6748. Fig. 1 shows the malignant skin lesion from the ISIC dataset. The dataset images consisted of noise artifacts like surgical ink markers, and the presence of hair that impedes accurate lesion classification. The images are resized into 224×224 pixels for compatibility with pre-trained neural networks during the model training and testing phase. The images were down-sampled since most of the pre-trained deep learning networks take input images of fixed resolution. By training raw images of larger sizes, neural networks will require more computing power to handle higher parameters that may lead to model overfitting. To train images faster, improve the performance of neural networks and reduce model overfitting, it is important to resize images into smaller resolutions based on the architecture of the pre-trained network. According to the study by Talebi and Milanfar [17], the perceptual quality of resized images is not lost while building computer vision models; instead aids in boosting the performance of the network.

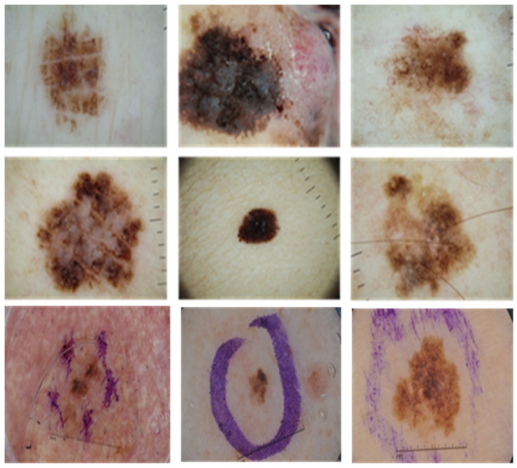
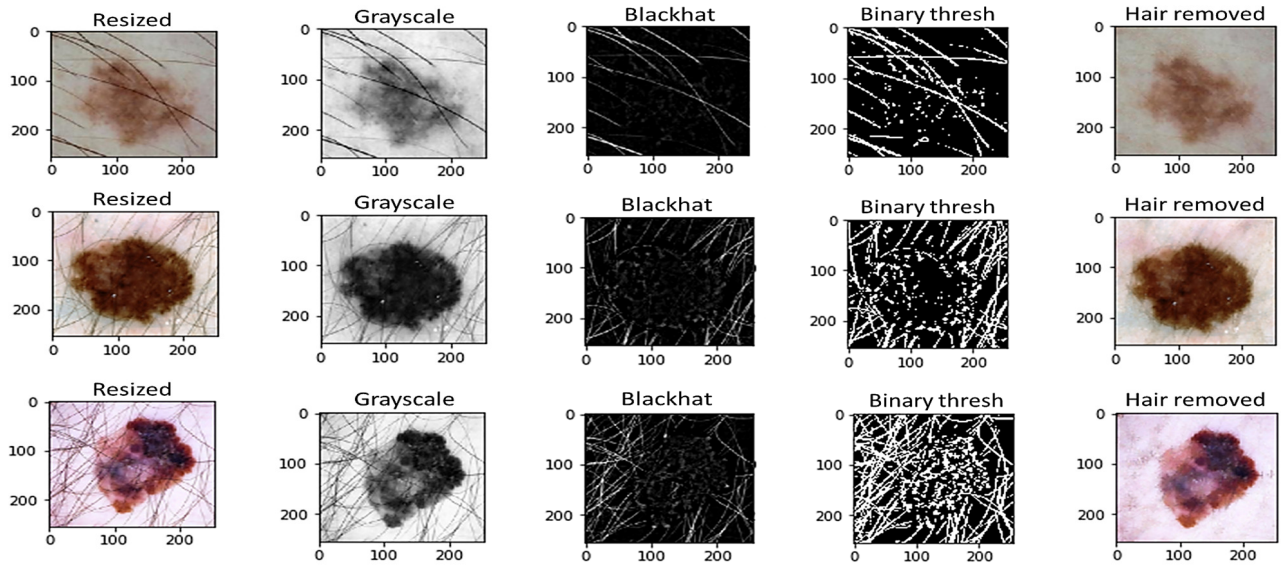


Fig. 1 Skin lesion images from the ISIC dataset

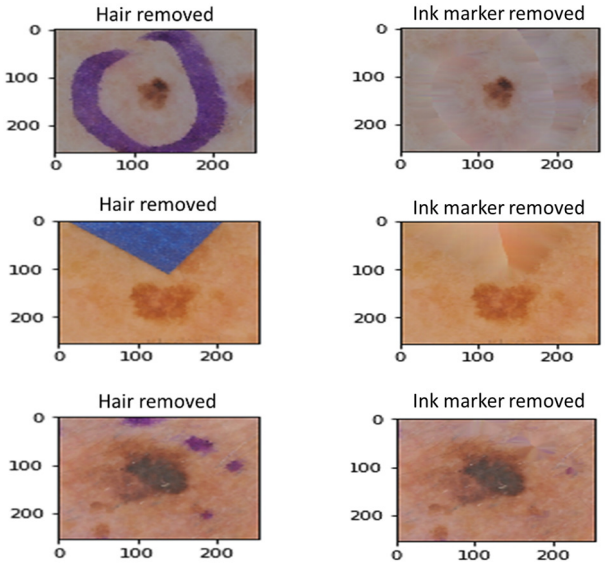
Hair artifacts in lesion images have a huge impact in building a computer-based melanoma classification system as hair structures tend to block the lesion region. Color, length, and thickness of hair are some factors that need to be considered while building an automated image preprocessing system [18]. In this study, the hair artifacts removal model is designed to efficiently remove thin and thick hair noise from dermoscopy images without impacting the quality of the image. Lesion images are converted from RGB color space to grayscale images using the weighted method. Images in grayscale aid in identifying hair

artifacts from skin lesion images. Blackhat filtering technique is applied on these grayscale lesion images, which further highlights hair noise against lighter skin backgrounds. To effectively probe the hair structures, a structuring element of elliptical shape and 13-pixel size blackhat filter is used [19]. A binary thresholding function is exerted over the blackhat image to create a hair mask that further sharpens hair structures from the background skin image. The fast marching restoration INPAINT_TELEA method is applied to the masked image. INPAINT_TELEA technique builds the original image without any hair noise from the masked image. Fig. 2 shows the proposed hair removal process.



(a) Original image (b) Grayscale image (c) Blackhat filtered image (d) Threshold image (e) Hair removed

Fig. 2 Hair removal process



(a) Images with ink markers (b) Images without ink markers

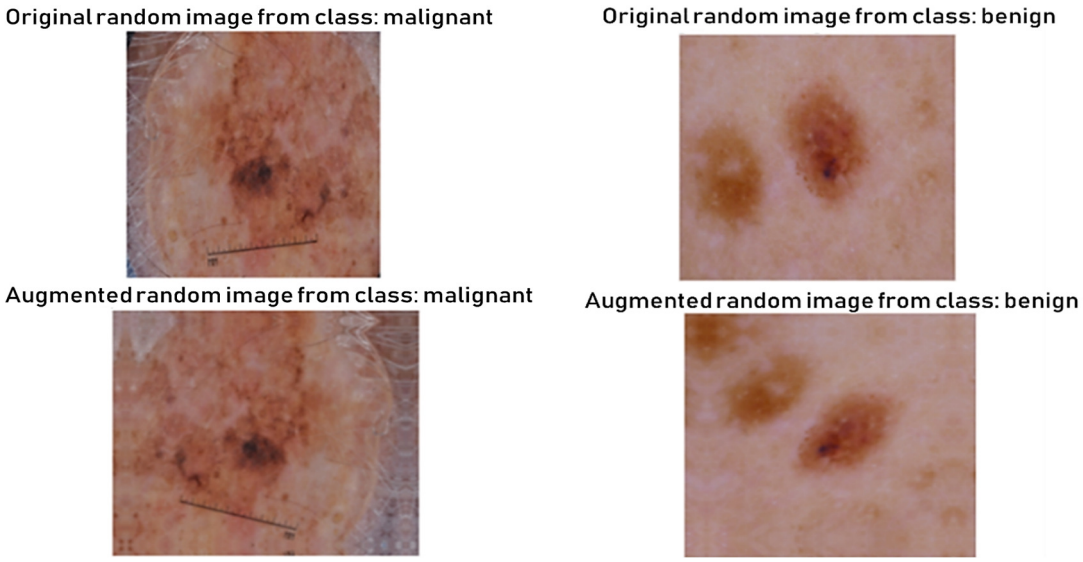
Fig. 3 Preprocessed image without markers

Clinical experts mark out suspicious skin lesions with blue or violet ink markers. These ink markings may be considered a part of skin lesions and can cause false interpretations while constructing deep learning models [20]. Therefore, it is important to eliminate these ink marker artifacts from the lesion images. To effectively remove ink markers, lesion images are transmogrified into hue-saturation-value (HSV) color space that aids in color-based segmentation to capture blue or violet ink markers. To create a masked image, the inRange function is used to set up a lower and upper band of violet color, and the

morphological dilation function is used to capture ink markers from HSV images. To restore the original image from the masked image, the inpainting method is again applied which produces lesion images without ink markers. Fig. 3 shows a comparison of the original image with ink markings and image after an automated model is applied to it.

3.2. Data augmentation

To build a good classification deep learning model, it is substantial to train the neural model with a huge volume of data. The majority of pre-trained networks are trained on ImageNet, which comprises a large set of data with 1000 classes. The data augmentation approach is adapted to expand the size of the training dataset to develop an effectual melanoma lesion classification model. In the data augmentation method, training data is synthetically expanded by minor alterations to existing original data [21]. To ameliorate the functioning of the melanoma classification model and prevent overfitting of the model, data augmentation is added on top of the EfficientNetB3 network. The training dataset is expanded by creating altered versions of images pertained to equivalent classes. Diverse augmentation techniques like random zoom, random rotation, random flip, random shift by height, and random shift by width are applied using Keras preprocessing layers like keras.layers.Resizing, keras.layers.RandomFlip, keras.layers.Rescaling and keras.layers.RandomRotation to build Keras Sequential model. Table 1 indicates various data augmentation methods. Fig. 4 depicts the augmentation operation applied to lesion images.



(a) Random augmented image of malignant class (b) Random augmented image of benign class

Fig. 4 Random augmented images

Table 1 Data augmentation approaches

Approach	Description
Random width	The width of images arbitrarily shifted by 20%
Random rotation	Images arbitrarily rotated by 20%
Random flip	Images arbitrarily flipped
Random zoom	Images arbitrarily zoomed by 20%
Random height	The height of images arbitrarily shifted by 20%

3.3. Model architecture

EfficientNet models include a family of 8 models from B0-B7 trained over ImageNet. EfficientNet models are deemed as the uttermost computationally effective deep learning model that acquires top accuracy gain appertaining to the compound scaling approach. In the compound scaling approach, the size of the baseline convolution network model is expanded by scaling the network uniformly across depth, width, and resolution to the target model size [22]. Fig. 5 exhibits the scaling of

the EfficienNet model. EfficientNet models consist of inverted residual convolutional blocks (MBConv) originally based on MobileNetV2 [23] with multiple kernel sizes of 3×3 and 5×5. The model architecture is broadened evenly through compound scaling coefficient ∅ by depth, width, and resolution in the following procedure:

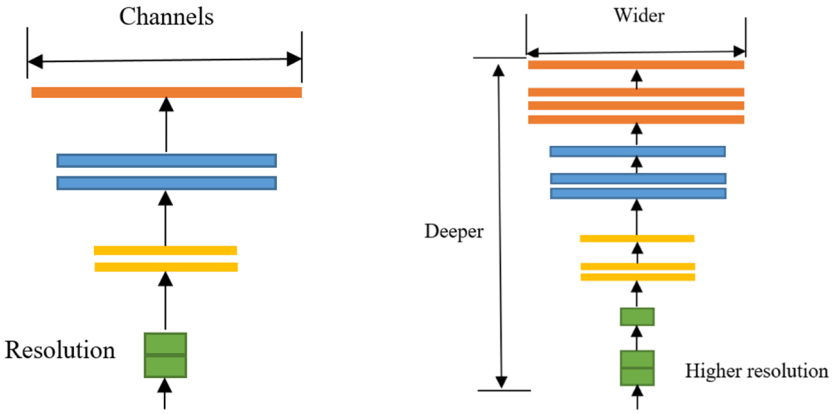
$$d = \alpha^{\phi}, r = \gamma^{\phi}, w = \beta^{\phi} \tag{1}$$

such that

$$\alpha\beta^2\gamma^2 \approx 2, \ \alpha \geq 1, \ \gamma \geq 1, \ \beta \geq 1 \tag{2}$$

where d indicates the depth of the network, r indicates the resolution of the network, w indicates the width of the network, and α, β, and γ are constants.

∅ value in Eq. (1) indicates the level at which the network can be scaled up. EfficientNetB0 baseline model is constructed on ∅ value of 0, w value of 1, and r value of 1. The EfficientNetB3 model is established on ∅ value of 3, w value of α^3, and r value of γ^3. The higher value of ∅ signifies extensive resources accessible to obtain superior results.



(a) Scaling of baseline EfficientNet model (b) Compound scaling of higher model

Fig. 5 Scaling of EfficientNet

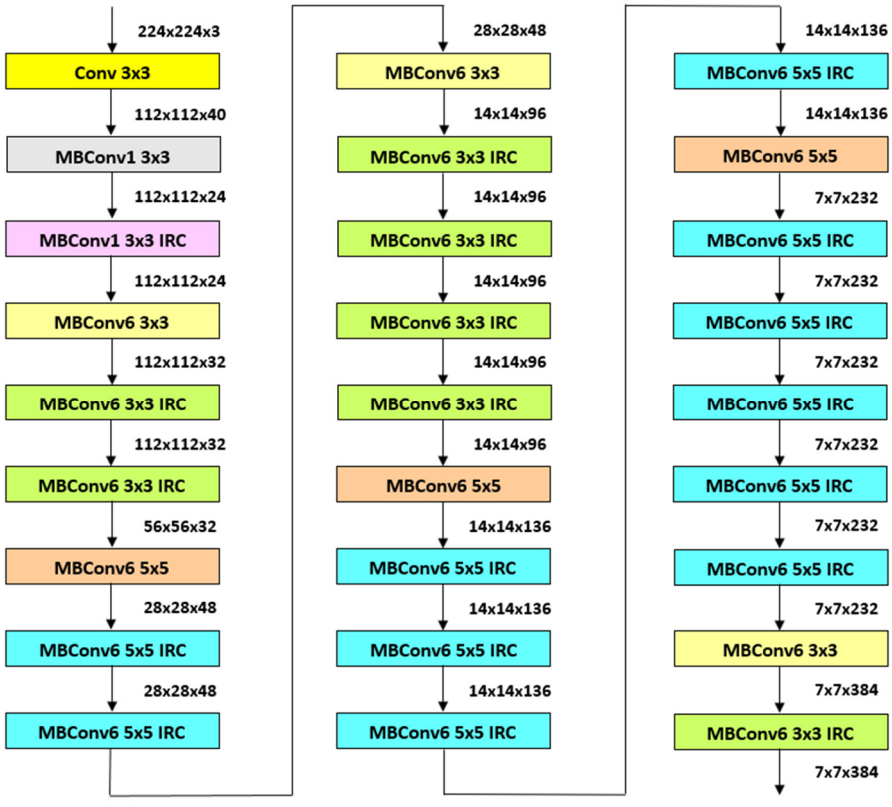


Fig. 6 EfficientNetB3 architecture

EfficientNetB3 consists of a convolution filter (Conv) block with a kernel size of 3×3, an MBConv1 block with a kernel size of 3×3, and MBConv6 blocks with kernel sizes of 3×3 and 5×5. Some of the MBConv6 blocks apply inverted residual connection (IRC). Filter kernel sizes of 3×3 and 5×5 are used in the EfficientNetB3 model to extract feature maps from the input images. EfficientNetB3 comprises 25 MBConv blocks differing in many characteristics such as feature maps expansion ratio, resolution, output layers kernel size, etc. MBConv1 with a kernel size of 3×3 and MConv6 with kernel sizes of 3×3 and 5×5 employ depthwise convolution along with batch normalization and activation layer. Additionally, layers of dropout and skip connection are integrated with MBConv6 3×3, and MBConv6 5×5 but omitted in MBConv1. In comparison to the baseline EfficientNetB0 model, EficientNetB3 comprises a larger network that helps to pull out detailed features which can infer better on new missions. The EfficientNetB3 model has the advantage of a broader network that abstracts superlative features and patterns employed for melanoma classification. Fig. 6 shows EfficientNetB3 architecture.

4. Results

The proposed model amalgamated data preprocessing and augmentation techniques, along with fine-tuned GAP layer and softmax output layer for classification, to improve the efficiency of lesion classification results. The ISIC dataset was distributed in an 80:20 ratio of training and testing batches. The training set consisted of 2637 images, and the testing set consisted of 660 images. All experiments are performed on a Google Colab notebook that furnished usage to NVIDIA Tesla GPU of size 12Gb K80 SMI 460.32.03. All the models are compiled employing the Adam optimization algorithm with a learning rate of 0.001. The models have been trained for 35 epochs holding a batch size of 32. The images are trained batch-wise and approximately take 29 to 33 seconds to train per epoch. Recall, confusion matrix, precision, accuracy, and F_1-score are computed to probe model potency [24]. These metrics are calculated on true positive (TP), true negative (TN), false positive (FP), and false negative (FN) [25].

(1) True positive (TP)- the correct class is positive and the predicted class is positive.

(2) False positive (FP)- the correct class is negative and the predicted class is positive.

(3) True negative (TN)- the correct class is negative and the predicted class is negative.

(4) False negative (FN)- the correct class is positive and the predicted class is negative.

Accuracy: It is an evaluation metric that finds the model's performance across all classes. It is a fraction of the sum of correct class predictions to the sum of total predictions.

$$Accuracy = \frac{TP + TN}{FN + TN + FP + TP} \tag{3}$$

Precision: It computes the ratio of total positive identification over the sum of total positive identification that is either categorized correctly or incorrectly.

$$Precision = \frac{TP}{FP + TP} \tag{4}$$

Recall: It computes the ratio of total positive identification over the sum of total positive input samples categorized precisely.

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

F_1-score: It is computed from precision and recall metrics that calculate the model's accuracy by giving higher importance to false negatives and false positives.

$$F_1 - score = \frac{2TP}{FN + FP + 2TP} \tag{6}$$

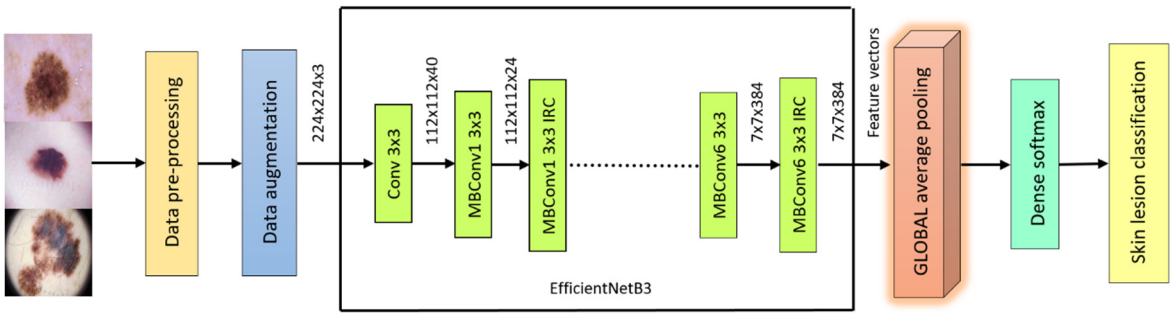


Fig. 7 Proposed methodology architecture

Validation of the effectiveness of the proposed framework is accomplished in two approaches. Primarily, four models are compiled to analyze the outcomes of various layers of EfficientNetB3 network architecture in the first approach. The training of these models is carried out on the preprocessed dataset, and the fully connected (FC) last layer of classification is fine-tuned to adjust to the binary (benign and malignant) categorization of the ISIC 2017 dataset. Model 1 depicts the baseline EfficientNetB3 model with no augmentation and no GAP layer. Model 2 refers to EfficientNetB3 with augmentation techniques applied to the network with no GAP layer. Model 3 depicts the GAP layer added to EfficientNetB3 network with no augmentation techniques applied. Model 4 refers to the fine-tuned proposed architecture presented in this study. Table 2 presents the methodology of each model.

Table 2 Approach of models

Approach	Preprocessing	Augment	GAP	FC
Model 1	✓	×	×	✓
Model 2	✓	✓	×	✓
Model 3	✓	×	✓	✓
Model 4 (Proposed model)	✓	✓	✓	✓

Table 3 provides the result analyses of the above models on the ISIC dataset. From Table 3, it can be found that model 1 achieved an accuracy of 56.97%, a precision of 58%, which indicates that the pre-trained baseline EfficientNetB3 model suffered from an overfitting problem. Fig. 8 shows the accuracy curve of model 1. Model 2, with augmentation techniques applied to the network, gave poor results with an accuracy of 56.67% as compared to model 1. Fig. 9 shows the accuracy curve of model 2. Model 3 gave an accuracy of 85.75%, which indicated that the GAP layer boosted the performance of the EfficientNetB3 network. Fig. 10 shows the accuracy curve of model 3. Model 3 suffered from an overfitting problem, although the model was performing well on training data. Whereas for testing data, the model performed comparatively less, as observed in Fig. 10, which represents the accuracy curve of Model 3. Model 4 achieved the best results with an accuracy of 88.13% and also acquired a balanced precision value of 88%, which demonstrates that the proposed framework overcomes and handles the problem of model overfitting. Fig. 11 represents the accuracy curve of Model 4 (proposed framework).

Table 3 Result summary of models

Approach	Recall	Accuracy	F_1-score	Precision
Model 1	58.00	56.97	58.00	58.00
Model 2	57.00	56.67	53.00	56.00
Model 3	86.00	85.75	85.00	86.00
Model 4 (Proposed model)	88.00	88.13	88.00	88.00

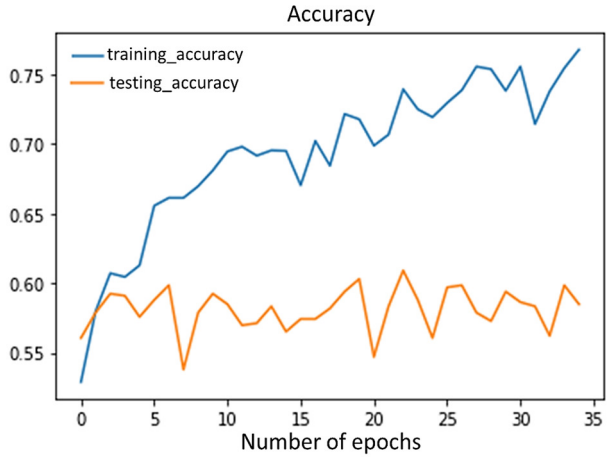


Fig. 8 Accuracy curve of Model 1

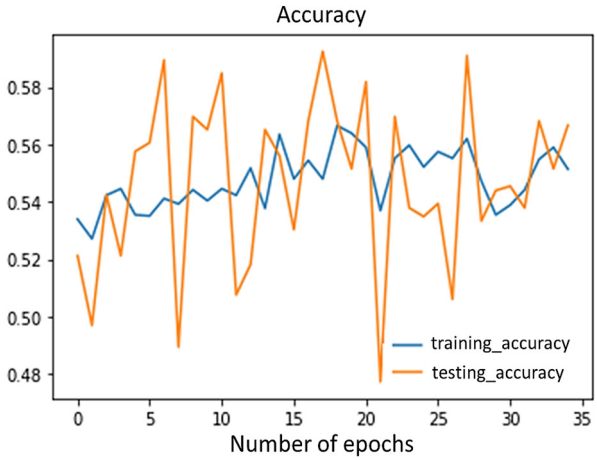


Fig. 9 Accuracy curve of Model 2

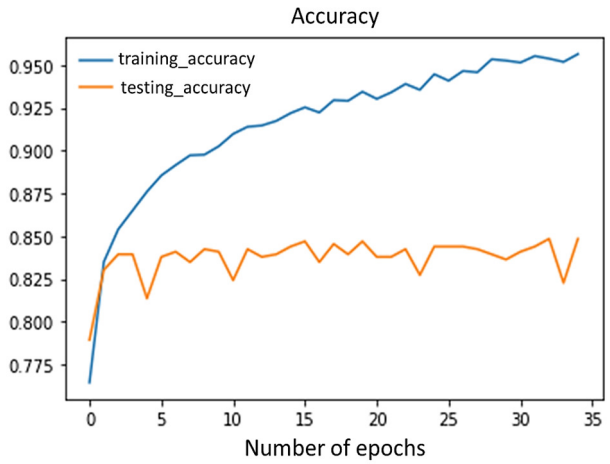


Fig. 10 Accuracy curve of Model 3

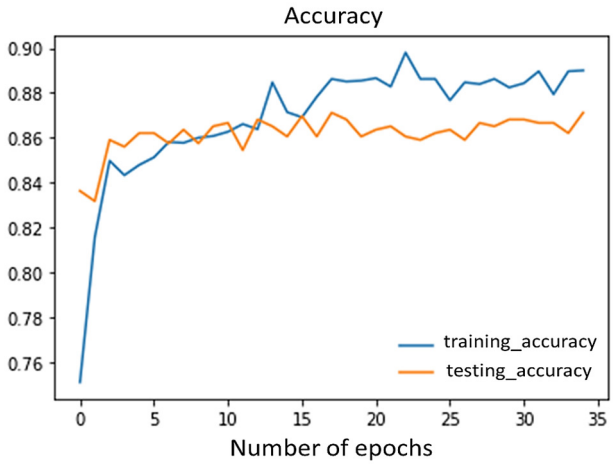


Fig. 11 Accuracy curve of Model 4

In the second approach, for evaluation of the proposed model, its result is compared to other advanced pre-trained models, namely InceptionResNetV2 [26], ResNet50 [27], EfficientNet B0-B2 [22], and InceptionV3 [28] over the ISIC dataset for the melanoma classification task. To investigate the potentiality of the proposed model, it is compared with advanced networks like InceptionResNetV2, ResNet50, InceptionV3, and EfficientNetB0-B2 models. These advanced neural models are pre-trained on the ImageNet dataset that outputs feature vectors for 1000 categories. Therefore, it is important to adjust these models on the target ISIC dataset comprising only two types of benign and malignant classes. To fit these models over the ISIC dataset, the last classification layer of these models is altered using the softmax layer with two classes.

Table 4 Comparative analyses of various evaluation metrics

Approach	Dataset	Recall	Accuracy	F_1-score	Precision
Proposed model	ISIC 2017	88.00	88.13	88.00	88.00
InceptionV3	ISIC 2017	83.00	82.73	83.00	83.00
EfficientNetB0	ISIC 2017	60.00	60.15	59.00	65.00
ResNet50	ISIC 2017	85.00	84.85	85.00	85.00
EfficientNetB1	ISIC 2017	56.00	55.75	55.00	55.00
EfficientNetB2	ISIC 2017	66.00	58.48	66.00	66.00
InceptionResNetV2	ISIC 2017	83.00	83.33	83.00	83.00

Table 4 indicates the comparative metrics assessment of the proposed model with other advanced pre-trained networks. The proposed design excelled over other models and gave an accuracy of 88.13%, recall of 88%, precision of 88%, and F_1-score of 88%. ResNet50 model achieved an accuracy of 84.85% and performed well compared to InceptionV3,

InceptionResNetV2, and EfficientNetB0-B2 models. Fig. 12 shows the accuracy curve of ResNet50. EfficientNetB1 achieved the lowest accuracy of 55.75%, with an F_1-score of 55%. Fig. 13 shows the accuracy curve of InceptionResNetV2. InceptionV3, EfficientNetB0, EfficientNetB2 and InceptionResNetV2 achieved an accuracy of 82.73%, 60.15%, 58.48%, and 83.33% respectively. Confusion matrices of each model are also examined to check false negative and false positive class counts. Fig. 14 depicts the confusion matrix of fine-tuned proposed model in comparison with InceptionV3, ResNet50, InceptionResNetV2.

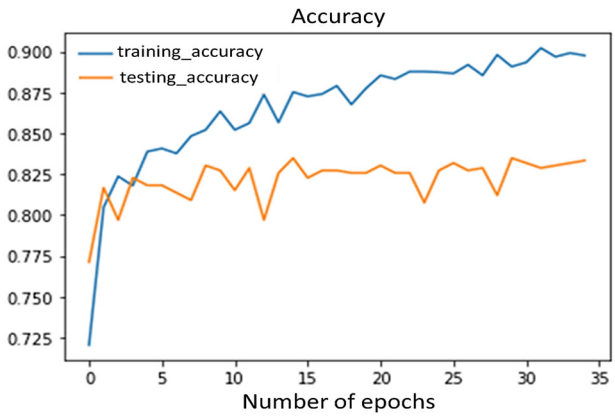


Fig. 12 ResNet50 accuracy curve

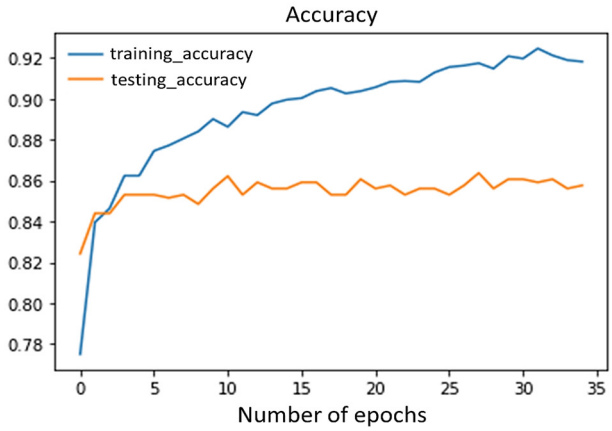
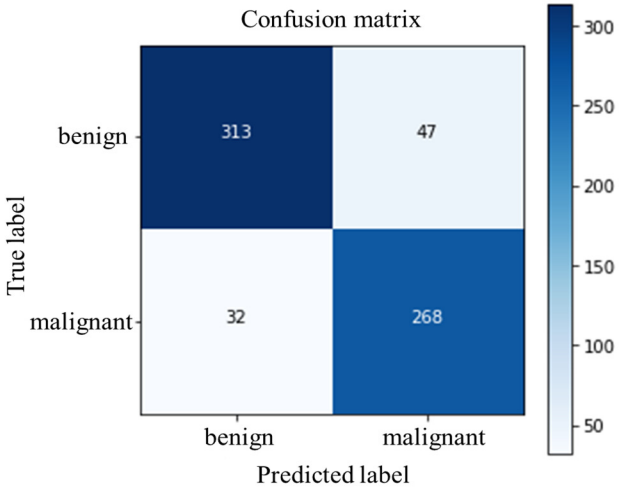
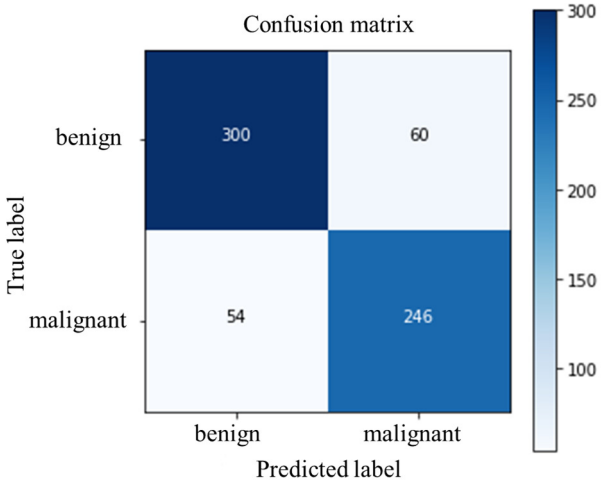


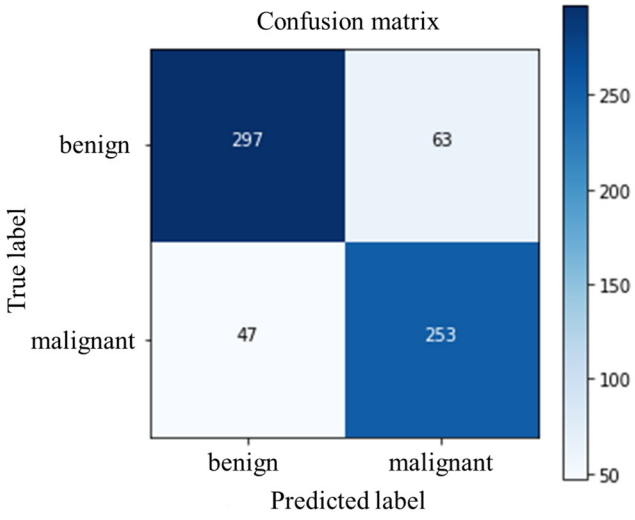
Fig. 13 InceptionResNetV2 accuracy curve



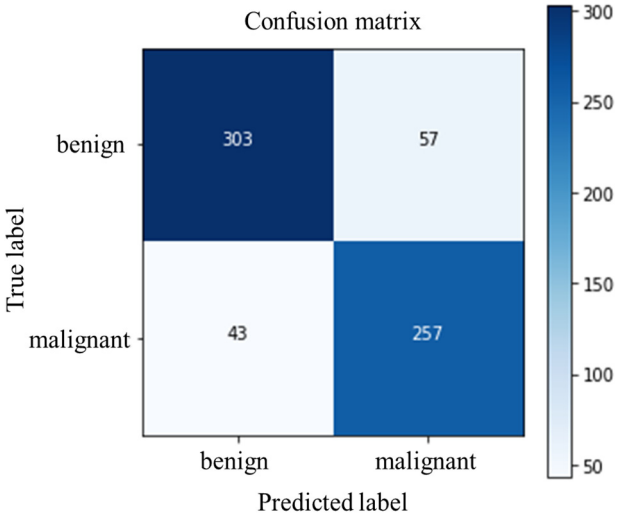
(a) Proposed model confusion matrix



(b) InceptionV3 model confusion matrix



(c) InceptionResNetV2 model confusion matrix



(d) ResNet50 model confusion matrix

Fig. 14 Confusion matrix comparison

5. Discussion

To further validate the efficacy of the proposed model, a comparative analysis is carried out with other state-of-art methods. Table 5 provides a comparative evaluation of the proposed model with other methods. The experimental study provided by Naronglerdrit et al. [5] over the HAM10000 dataset provided an accuracy of 97.12%, but less recall value of 85.18% using the ResNet-101 deep learning model indicates the model suffers from the problem of overfitting. Categorical accuracy of the MobileNet model with data augmentation proposed by Chaturvedi et al. [7] provided an overall accuracy of 83.15%, precision of 89%, recall of 83%, and F_1-score of 83% only. Ashim et al. [10], presented an experimental analysis of pre-trained networks such as VGG16, ResNet50, EfficientNet, DenseNet, and Xception for melanoma classification over the Kaggle dataset consisting of only 660 images of benign and malignant classes. Xception, EfficientNetB0, DenseNet, VGG16, and ResNet models furnished training accuracy of 78.41, 78.44%, 81.94%, 72.37%, and 88.61% respectively.

Le et al. [12], presented a ResNet50-based deep learning classifier over the HAM10000 dataset and achieved an average accuracy of 93%, precision of 81%, recall, and F_1-score of 80% only. The low value of other evaluation metrics indicates that the model proposed by [12] suffers from model overfitting problem. Acosta et al. [15], proposed a deep learning model based on ResNet152 and also presented a comparison of 20 other state-of-art models trained and tested over the ISIC 2017 dataset. According to the study carried out by [15], the proposed ResNet152 achieved the highest accuracy of 87.2%, the sensitivity of 82%, and the F_1-score of 84.8%. Rezaoana et al. [16] performed an exploratory analysis of the proposed CNN model with other networks like VGG-16 and VGG-19. The proposed approach by [16] achieved an F_1-score of 76.92%, precision of 76.16%, and recall of 78.15%. It can be noticed from Table 5 that the proposed EfficientNetB3 model accomplished better results compared to other methods concerning metrics such as recall, precision, and F_1-score. These metrics are consistent with the accuracy of the proposed model and thus handle the problem of model overfitting.

Table 5 Comparative analyses of the proposed methodology with other methods

Reference	Approach	Dataset	Accuracy	Recall	F_1-score	Precision
Proposed model	EfficientNetB3	ISIC 2017	88.13	88.00	88.00	88.00
Naronglerdrit et al. [5]	ResNet-101	HAM10000	97.12	85.18	-	-
Chaturvedi et al. [7]	MobileNet	HAM10000	83.15	83.00	83.00	89.00
Ashim et al. [10]	EfficientNetB0	Kaggle	78.41	-	-	-
	ResNet		88.61	-	-	-
Le et al. [12]	ResNet50	HAM10000	93.00	81.00	80.00	80.00
Acosta et al. [15]	ResNet152	ISIC 2017	87.20	82.00	84.80	-
Rezaoana et al. [16]	Proposed CNN	Kaggle ISIC	79.45	78.15	76.92	76.16
	VGG-16		69.57	68.89	67.77	65.67
	VGG-19		71.19	69.45	68.95	68.54

6. Conclusion

In this study, an automated computer-aided diagnostic system using fine-tuned EfficientNetB3 deep neural model for melanoma classification is proposed that efficiently classifies lesion images into benign and malignant classes. Skeptical skin lesions are routinely marked with surgical ink markers, and the presence of hair artifacts often influences the classification analysis. An automated preprocessing model is employed, and it can effectively remove surgical ink markers and hair artifacts. A broad variety of data augmentation schemes are utilized to prevent overfitting and improve the overall performance of the proposed model. The weights of EfficientNetB3 models are fine-tuned by appending an additional layer of GAP, and softmax layer to adjust to the ISIC 2017 dataset. Extensive experimental analyses are carried out with other popular pre-trained CNN networks like InceptionResNetV2, InceptionV3, EfficientNetB0-B2, and ResNet50. The analytic results indicate that the proposed model achieves robust and higher classification results. Empirical findings demonstrated the effectiveness of the proposed model in the malignant melanoma classification task.

The future study might be to employ the proposed model on diverse repositories of ISIC datasets, such as the HAM10000 dataset, ISIC 2019, etc., consisting of more than 10000 images of pigmented skin lesions. It can also include testing the proposed model on multi-labeled lesion classification dataset and building fine-grained neural networks for melanoma classification with improved accuracy. The study can be extended to develop an efficient lightweight smartphone application integrating the proposed methodology for skin lesion classification with a deep learning-based segmentation model.

Conflicts of Interest

The authors declare no conflict of interest.

Statement of Ethical Approval

For this type of study, statement of human rights is not required.

Statement of Informed Consent

For this type of study, informed consent is not required.

References

[1] "Cancer Facts and Figures 2021," https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2021/cancer-facts-and-figures-2021.html, December 01, 2021.

[2] S. Sonthalia, S. Yumeen, and F. Kaliyadan, "Dermoscopy Overview and Extradiagnostic Applications," https://www.ncbi.nlm.nih.gov/books/NBK537131/, August 13, 2021.

[3] K. Munir, H. Elahi, A. Ayub, F. Frezza, and A. Rizzi, "Cancer Diagnosis Using Deep Learning: A Bibliographic Review," Cancers (Basel), vol. 11, no. 9, article no. 1235, August 2019.

[4] "ISIC Challenge Datasets," https://challenge.isic-archive.com/data/, December 01, 2021.

[5] P. Naronglerdrit, I. Mporas, M. Paraskevas, and V. Kapoulas, "Melanoma Detection from Dermatoscopic Images Using Deep Convolutional Neural Networks," International Conference on Biomedical Innovations and Applications (BIA), pp. 13-16, November 2020.

[6] N. Siddique, S. Paheding, Md. Z. Alom, and V. Devabhaktuni, "Recurrent Residual U-Net with EfficientNet Encoder for Medical Image Segmentation," Pattern Recognition and Tracking XXXII, pp. 1-10, April 2021.

[7] S. S. Chaturvedi, K. Gupta, and P. S. Prasad, "Skin Lesion Analyser: An Efficient Seven-Way Multi-Class Skin Cancer Classification Using MobileNet," International Conference on Advanced Machine Learning Technologies and Applications, pp. 165-176, February 2020.

[8] R. Zhang, "Melanoma Detection Using Convolutional Neural Network," IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), pp. 75-78, January 2021.

[9] Y. Zhang and C. Wang, "SIIM-ISIC Melanoma Classification with DenseNet," IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), pp. 14-17, March 2021.

[10] L. K. Ashim, N. Suresh, and C. V. Prasannakumar, "A Comparative Analysis of Various Transfer Learning Approaches Skin Cancer Detection," 5th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 1379-1385, June 2021.

[11] Y. Chen, Y. Zhu, and Y. Chang, "CycleGAN Based Data Augmentation for Melanoma Images Classification," Proceedings of the 3rd International Conference on Artificial Intelligence and Pattern Recognition, pp. 115-119, June 2020.

[12] D. N. T. Le, H. X. Le, L. T. Ngo, and H. T. Ngo, "Transfer Learning with Class-Weighted and Focal Loss Function for Automatic Skin Cancer Classification," https://arxiv.org/abs/2009.05977, September 13, 2020.

[13] M. Manzo and S. Pellino, "Bucket of Deep Transfer Learning Features and Classification Models for Melanoma Detection," Journal of Imaging, vol. 6, no. 12, pp. 1-15, November 2020.

[14] M. A. Kadampur and S. A. Riyaee, "Skin Cancer Detection: Applying a Deep Learning Based Model Driven Architecture in the Cloud for Classifying Dermal Cell Images," Informatics in Medicine Unlocked, vol. 18, no. 100282, pp. 1-6, 2020.

[15] M. F. J. Acosta, L. Y. C. Tovar, M. B. G. Zapirain, and W. S. Percybrooks, "Melanoma Diagnosis Using Deep Learning Techniques on Dermatoscopic Images," BMC Medical Imaging, vol. 21, no. 1, pp. 1-11, January 2021.

[16] N. Rezaoana, M. S. Hossain, and K. Andersson, "Detection and Classification of Skin Cancer by Using a Parallel CNN Model," IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), pp. 380-386, December 2020.

[17] H. Talebi and P. Milanfar, "Learning to Resize Images for Computer Vision Tasks," IEEE/CVF International Conference on Computer Vision (ICCV), pp. 497-506, October 2021.

[18] M. K. Tekleyohannes, C. Weis, N. Wehn, M. Klein, and M. Siegrist, "A Reconfigurable Accelerator for Morphological Operations," IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), pp. 186-193, May 2018.

[19] S. Chatterjee, D. Dey, and S. Munshi, "Integration of Morphological Preprocessing and Fractal Based Feature Extraction with Recursive Feature Elimination for Skin Lesion Types Classification," Computer Methods Programs in Biomedicine, vol. 178, pp. 201-218, September 2019.

[20] J. K. Winkler, C. Fink, F. Toberer, A. Enk, T. Deinlein, R. Hofmann-Wellenhof, et al., "Association between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition," JAMA Dermatology, vol. 155, no. 10, pp. 1135-1141, October 2019.

[21] A. Mikołajczyk and M. Grochowski, "Data Augmentation for Improving Deep Learning in Image Classification Problem," IEEE International Interdisciplinary PhD Workshop (IIPhDW), pp. 117-122, May 2018.

[22] M. Tan and Q. Le, "Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks," International Conference on Machine Learning, pp. 6105-6114, June 2019.

[23] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4510-4520, June 2018.

[24] M. Lin, Q. Chen, and S. Yan, "Network in Network," https://arxiv.org/abs/1312.4400, December 16, 2013.

[25] S. R. Salian and S. D. Sawarkar, "Melanoma Skin Lesion Classification Using Improved EfficientnetB3," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 8, no. 1, pp. 45-56, March 2022.

[26] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-V4, Inception-ResNet and the Impact of Residual Connections on Learning," 31st AAAI Conference on Artificial Intelligence (AAAI'17), pp. 4278-4284, February 2017.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, June 2016.

[28] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818-2826, June 2016.