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

Skin Cancer Classification and Comparison of Pre-trained Models Performance using Transfer Learning

Subroto Singha[1)*], Priyangka Roy[2)]

[1)2)]*Texas A&M University-Commerce, USA*
2600 W Neal St, Commerce
[1)]ssingha@leomail.tamuc.edu, [2)]proy@leomail.tamuc.edu

Abstract

Background: Skin cancer can quickly become fatal. An examination and biopsy of dermoscopic pictures are required to determine if skin cancer is malignant or benign. However, these examinations can be costly.
Objective: In this research, we proposed deep learning (DL)-based approach to identify a melanoma, the most dangerous kind of skin cancer. DL is particularly excellent in learning traits and predicting cancer. However, DL requires a vast number of images.
Method: We used image augmentation and transferring learning to categorise images into benign and malignant. We used the public ISIC 2020 database to train and test our models. The ISIC 2020 dataset classifies melanoma as malignant. Along with the categorization, the dataset was examined for variation. The training and validation accuracy of three of the best pre-trained models were compared. To minimise the loss, three optimizers were used: RMSProp, SGD, and ADAM.
Results: We attained training accuracy of 98.73%, 99.12%, and 99.76% using ResNet, VGG16, and MobileNetV2, respectively. We achieved a validation accuracy of 98.39% using these three pre-trained models.
Conclusion: The validation accuracy of 98.39% outperforms the prior pre-trained model. The findings of this study can be applied in medical science to help physicians diagnose skin cancer early and save lives.

Keywords: Deep Learning, ISIC 2020, Pre-trained Model, Skin Cancer, Transfer Learning

Article history: Received 2 September 2022, first decision 26 September 2022, accepted 8 October 2022, available online 28 October 2022

I. INTRODUCTION

One of the most common types of cancer is skin cancer, with over 123,000 new cases diagnosed each year worldwide. According to the World Health Organization (WHO), skin cancer accounts for one-third of all malignancies globally. According to the American Cancer Society, melanoma is responsible for 75% of skin cancer fatalities. Around 100,000 new instances of melanoma were detected by the end of 2020, with 7,000 people dying from the condition. In the United States alone, melanoma is responsible for around 9000 deaths each year [2].

The most common kinds of skin cancer are basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and malignant melanoma (MM). The majority of skin cancers are non-melanocytic, which are caused by BCC and SCC. Melanoma is caused by UV radiation. Melanocyte cells grow abnormally and affect surrounding tissues by multiplying and spreading through lymph nodes [1]. This kind of cancer spreads quickly and can be deadly if not diagnosed and treated early.

The cost of treatment can reach USD 134,000 when skin cancer is already in the fourth stage [3]. Dermatologists first detect melanoma by examining photos of cancer and moles, among other treatment options. They also check for outlier lesions around the moles that might be melanoma. However, outcomes may not always be precise and accurate. Artificial intelligence (AI) has the potential to assist physicians in accurately identifying melanoma. AI-based detection technologies can increase dermatological accuracy. Existing AI approaches have not sufficiently considered this clinical environment. Dermatologists may increase their diagnostic accuracy if detection algorithms are used. If the AI detects skin cancer, it will significantly influence dermatological clinic operations. Convolutional neural network (CNN) has previously been used in image classification tasks such as breast cancer image classification [4], hyperspectral image classification [5], pap smear image classification [6], skin cancer and lesion classification [7 - 9], and others. Even when taught with hundreds or tens of thousands of images, CNNs accurately identify medical images.

* Corresponding author

ISSN 2443-2555 (online) 2598-6333 (print)
doi: http://dx.doi.org/10.20473/jisebi.8.2.218-225

In this study, skin lesion photographs are sufficient for skin cancer categorization into two binary classes: benign and malignant.

Our contributions to this work are six-fold. Firstly, we employed three pre-trained models to get the highest accuracy: VGG, ResNet, and MobileNetV2. Secondly, we used the ISIC 2020 database [10], a 48.9 GB dataset of 33,126 1024x1024 jpeg images, 533 of which are cancerous and the rest benign. Thirdly, we applied three optimisation algorithms to reduce the loss: SGD, ADAM, and RMSProp. Fourthly, we used a dropout layer and data augmentation to alleviate overfitting problems. Then, we used accuracy to assess the model's performance. Finally, we examined the ISIC dataset to determine the location of skin cancer, the type of sickness that causes cancer, and so on.

II. LITERATURE REVIEW

Skin cancer was first detected by human eyes in the early twentieth century, [11] indicated by abnormal size, bleeding, and ulceration. This procedure, however, depends mostly on physicians' eye examinations rather than any sophisticated or advanced technology. Another detection approach, dermoscopy of epiluminescence microscopy, has been demonstrated to be 75% to 85% accurate [12]. A biopsy is conducted to see if a practitioner cannot determine whether a mole is a melanoma or not [11].

Several CNN-based classifiers have recently been explored, aiding dermatologists in identifying melanoma. In fact, [3] identified skin lesion images as benign or malignant with 95.23% accuracy. Their CNN was built with four convolutional layers, the ReLu activation function, and a softmax classifier. They used the ADAM and SGD algorithms to decrease neural network loss and add noise to increase accuracy. Using ISIC 2018 skin lesion pictures, the proposed classifier was trained and assessed. [13] proposed a new model that classifies skin lesions as benign or malignant using a novel regularizer technique. Their binary classifier could distinguish between benign and malignant cancers in images. They observed the AUCs for nevus vs melanoma lesion, seborrheic keratosis versus basal cell carcinoma lesion, seborrheic keratosis versus melanoma lesion, and solar lentigo versus melanoma lesion. They found the AUC scores were 0.77, 0.93, 0.85, and 0.86, respectively, for these three comparisons. Their model has an average accuracy of 97.49%. Their technique could assist doctors in categorizing various skin lesions.

Meanwhile, [2] developed a CNN architecture for skin lesion grouping to attain excellent dermoscopy picture group precision. They employed an approach that merged the group layers of four different deep neural network topologies. In terms of accuracy, their results indicated that their technique outperformed the other CNNs. They developed a new CNN model based on a novel regularizer. Furthermore, [8] assessed the efficacy of their CNN on 21 board-certified dermatologists utilizing biopsy-proven clinical images and two critical binary groupings. Their deep learning CNN surpassed dermatologists in detecting skin cancer using dermoscopy and digital imaging. In addition, [14] suggested a well-performing automated computerized system, and their technique included learning. They utilized 2000 pictures from ISIC 2017 and attained an accuracy rate of 93.6%.

The University of Oxford's Visual Geometry Group (VGG) released VGGNet [15], in which they obtained top performance in the ImageNet dataset, which has 1000 classes. The key contribution of VGGNet was its smaller filter (3x3) compared to other comparable architectures, such as AlexNet (11x11). Furthermore, VGGNet demonstrated that increasing the depth of the architecture is more important than expanding the width to enhance performance. VGG16 and VGG19 are two versions of VGGNets, with VGG19 being more expensive to train due to its more complex design and parameters. Fig. 1 shows the architecture of VGG16, which consists of 16 convolutional layers with 3x3 filter sizes. In addition, the architecture has five max-pooling layers with a size of 2x2.

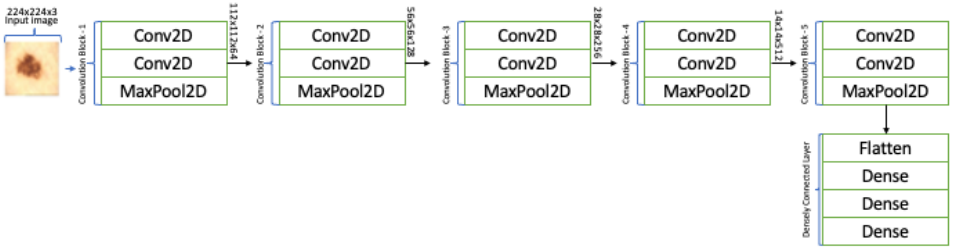

Fig. 1 Architecture of VGG-16

Another pre-trained architecture, ResNet [16], introduced deep networks by increasing network depth. Deeper networks introduced an issue known as vanishing gradient, which decreases neural network efficiency. To tackle the vanishing gradient problem, ResNet was proposed, with the key idea being bypassing one or more layers dubbed

'identity shortcut connections'. Fig. 2 shows the residual module in ResNet50 that helps prevent the vanishing gradient issue generated by deep networks.

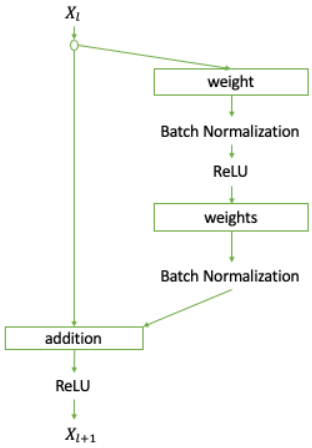

Fig. 2 Residual module in ResNet

Google introduced MobileNetV2 [17] for its lightweight and minimal complexity. This architecture is appropriate for mobile devices. Version 1 of MobileNet featured depth-wise separable convolution, whereas Version 2 introduced a superior module called inverted residual. Fig. 3 shows the convolutional blocks of MobileNetV2 in which the first layer is 1x1 convolution with ReLU6, the second layer is for depth-wise convolution, and the third layer brings non-linearity using 1x1 convolution.

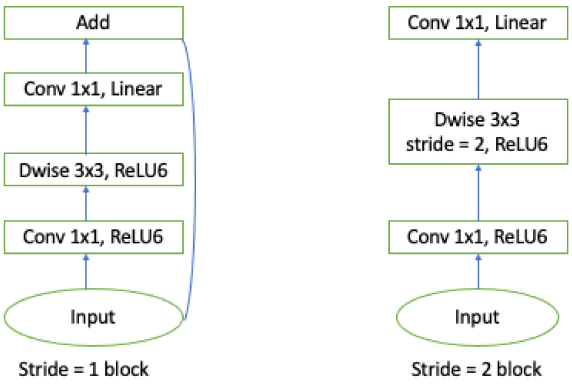

Fig. 3 MobileNetV2

Alongside the pre-trained models, data augmentation [18] is a commonly used strategy for correcting data imbalance. The most general approaches used in data augmentation are oversampling and under-sampling. To correct the class imbalance, oversampling uses exact copy duplicates or modified copies of the original data from the minority class. In addition to data augmentation and pre-trained models, transfer learning [19] is a formal process of using pre-trained models' learnt knowledge in a similar problem. All of the discussed pre-trained models were trained on ImageNet. Thus, the trained pre-trained models can be used in our problem to accelerate the training process and obtain faster prediction using fewer training images.

In literature, most methods used pure deep learning techniques to build models from scratch or manual techniques, such as physicians inspecting the mole to identify cancer. All of these techniques require ample time and are costly at times. In our next section, we present our proposed methodology to address the research gap and the techniques.

III. METHODS

This section discusses data preparation, data augmentation, the architectures of the pre-trained models, and evaluation metrics. Three pretrained models' performance was compared using ISIC 2020 images. Fig. 4 depicts the whole experimental setup. We employed data preparation first, followed by data augmentation in the second stage. Three models were chosen for the pre-trained training. Except for the output layer, we used all layers. A flattened layer, dense layer, dropout layer, and final output layer were added. After we concluded training, we used accuracy to evaluate the performance of our trained model.

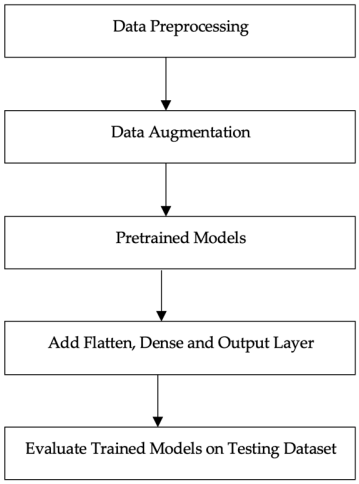

Fig. 4 Experimental setup

A. Data Pre-processing

Labelling the images is the first stage in data preparation. Each image in the ISIC 2020 dataset was labelled with a benign or malignant target. We divided the dataset into two parts: training and validation. The pre-trained models were trained using a train set. We employed an 80/20 data split, with 80% of images used for training and 20% used for validation. The final stage in data preparation was to rescale the images from 0 to 1. Because we employed RGB images with a pixel range of 0 to 255, rescaling images decreased training time and eliminated image pixel inconsistencies.

B. Data Augmentation

To execute data augmentation at random, we utilized rotation range=20, width shift range=0.2, height shift range=0.2, and horizontal flag=True. Data augmentation was utilized exclusively in the training dataset to keep the model impartial, whereas the testing dataset was kept unmodified save for rescaling the pictures between 0 and 1.

C. Pretrained Model's Architectures

CNN is excellent in image classification and detection [20]. It generates feature maps by stacking convolutional layers and pooling layers. Feature maps are high-level features derived from CNNs at each level of convolution. A final fully linked layer is employed as an output layer for picture categorization. However, the layer stack in CNNs can vary greatly, and there is no existing knowledge of what is the best architecture for a given dataset. Furthermore, a lack of data may result in incorrect classification, and training a CNN might take a long time to converge [20]. Thus, transfer learning is a solution in which we employ pre-trained models trained on some datasets. The layers can be put on top of the pre-trained architecture.

Our methodology used three common pre-trained models: ResNet, VGG, and MobileNet. Imagenet [21] was used to train all of these models. Four layers were added: a flattened layer, a dense layer, a dropout layer, and a dense output layer. With a learning rate of 0.0001, models were optimized using RMSProp, SGD, and ADAM. We utilized an early stopping strategy to avoid overtraining the models, in which we stopped training the models when the validation accuracy reached its maximum and was no longer improving.

We trained each epoch for 100 steps, and our loss function was binary cross-entropy. Pretrained models such as VGG, ResNet, and MobileNetV2 are frequently utilised in transfer learning for feature extraction or fine-tuning. Each model is trained on ImageNet, abbreviated ILSVRC for ImageNet Large Scale Visual Recognition Challenge. This

image classification challenge aims to train a model that can properly classify an input image into 1,000 unique item categories. Models are trained using 1.2 million training photographs, 50,000 validation photos, and 100,000 testing photos. These 1,000 image categories illustrate daily object classifications, such as dog and cat breeds, household products, automobile types, etc. As a result, pre-trained models are highly effective at extracting features with high accuracy and less training time.

D. Evaluation Metrics

We used accuracy (1) as an assessment criterion to compare the performance of the pre-trained models on the ISIC 2020 dataset. We did not evaluate additional metrics [22] like F1 score, Precision, Recall, and so on since our major purpose was to assess the performance of pre-trained models to determine if the method eliminates the need to design neural network architecture from scratch and achieves high accuracy in a short training period.

$$Accuracy\ (\%) = \frac{(TP + TN)}{(TP + TN + FP + FN)} * 100 \tag{1}$$

IV. RESULTS

We utilized a freely available dataset from the International Skin Imaging Collaboration [10]. There are 33,126 dermoscopic training pictures and 10,982 test images in the dataset. The training dataset was skewed, which is expected given the nature of medical data. There were 32542 benign images and 584 malignant images among all training data. That suggests that just 1.7% of the patients had aggressive cancer. The torso was the location of skin cancer in 51.7% of cases. Fig. 5 depicts the position of the skin from which the images were taken.

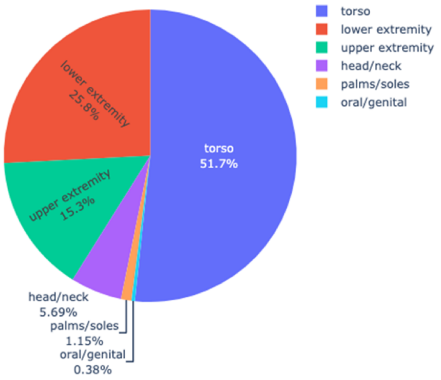

Fig. 5. Location of skin

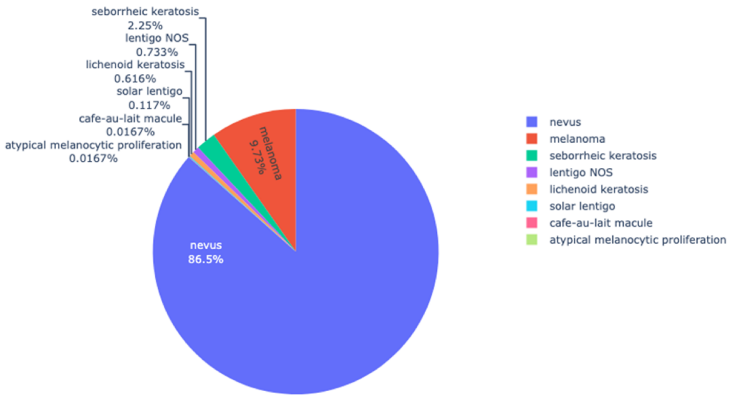

Fig. 6. Various classes of disease

Dermatologists classified benign and malignant skin tumours into nine categories. Fig. 6 depicts the disease's eight classifications—the unknown class was not included. The data show that 86.5% of the individuals with moles had nevus disease, which was classed as benign. Melanoma patients were diagnosed with malignant cancers. Table 1 shows nine types of disorders classified as benign or malignant skin malignancies. Fig. 7 shows the images of the benign and malignant classes.

TABLE 1
DISEASE CLASSES DIVIDED INTO BENIGN AND MALIGNANT

Disease Classes	Benign	Malignant
atypical melanocytic prolif-eration	1	0
cafe-au-lait macule	1	0
lentigo NOS	44	0
ichenoid keratosis	37	0
Melanoma	0	584
Nevus	5193	0
seborrheic keratosis	135	0
solar lentigo	7	0
unknown	27124	0

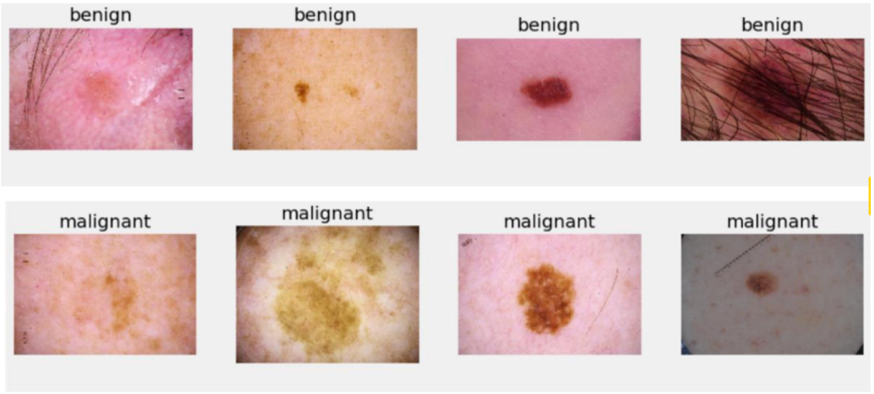

Fig. 7 Images of Benign and Malignant Skin Cancers

V. DISCUSSION

Using a massive dataset of 33,126 images, we achieved a validation accuracy of 98.39%. Using the pre-trained model ResNet and VGGNet, [23] attained an accuracy of 0.9208 and 0.8870. [24] used MobileNetV2 and obtained an accuracy of 0.8500. Table 2 compares training accuracies from the literature versus our proposed model's training accuracy. The greatest accuracy across these pre-trained models was 0.8500 in [24]. One pre-trained model was utilized in these architectures by integrating stages such as data augmentation, data standardization, etc. [24] employed a MobileNetV2 using ISIC-Archive and other similar datasets. However, none of the designs outperformed our attained accuracy. Furthermore, we tackled the challenges of overfitting and data preparation while minimising the need to develop models from scratch.

TABLE 2
PRETRAINED MODEL'S ACCURACIES IN LITERATURE VS OUR METHOD

Research	Training Accuracy			Validation Accuracy
	ResNet	VGGNet	MobileNet	
[23]	0. 9208	0. 8870	-	-
[24]	-	-	0. 8500	-
Our proposed method	0.9873	0.9912	0.9976	0.**9839**

Figures 5 and 6 show the distribution of the ISIC 2020 dataset. The torso is the largest anatomical site from where the skin images were collected, while oral/genital is the least common category. Similarly, we see that only 9.73% of the diagnosis were actually melanoma, i.e., malignant. Seeing the distribution from these two figures, we may correlate that torso is a common skin site of malignant images. Thus, if a mole is in the torso location of the skin, most likely, we need to investigate the mole closely for melanoma.

Finally, our proposed methodology has incorporated the data augmentation technique that addresses the issue of data imbalance. In our case, we had only 584 instances of malignant images. Thus, the dataset is highly imbalanced towards benign classes. Due to data augmentation, we artificially balanced the dataset during training. In addition to data augmentation, our proposed method took care of the overfitting issue by incorporating the dropout layer. The dropout layer randomly drops the neurons from the model, reducing the model complexity and decreasing the overfitting. Overfitting is the tendency of a model to memorise the training dataset and perform poorly in the validation/testing dataset. In our case, our training accuracies and validation accuracy were almost the same, which means the dropout layer worked well to tackle the overfitting issue. Hence, incorporating data augmentation, dropout layer with the data pre-processing and transfer learning techniques in our proposed model were the reasons to outperform the previous similar models.

VI. CONCLUSIONS

In this work, we employed pre-trained models to compare performance while considering an evaluation metric (accuracy). We used RMSProp, SGD, and ADAM to optimise the models. Pre-trained models were employed to obtain the maximum accuracy while spending the least time creating models from scratch. Furthermore, we addressed the overfitting issue and offered alternative data processing techniques with dataset insights. We achieved a validation accuracy of 98.39%, outperforming the prior pre-trained model's performance despite the need for a complex model. The findings of this study can be applied in medical science to help physicians diagnose skin cancer early and save lives.

Author Contributions: *Subroto Singha*: Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing, Supervision. *Priyangka Roy*: Software, Investigation, Data Curation, Writing - Original Draft.

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

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

REFERENCES

[1] S. Hosseinzadeh Kassani and P. Hosseinzadeh 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.

[2] F. Zou, L. Shen, Z. Jie, W. Zhang, and W. Liu, "A Sufficient Condition for Convergences of Adam and RMSProp," 2019, pp. 11127–11135. Accessed: Aug. 23, 2022. [Online]. Available: https://openaccess.thecvf.com/content_CVPR_2019/html/Zou_A_Sufficient_Condition_for_Convergences_of_Adam_and_RMSProp_CVPR_2019_paper.html

[3] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Dec. 2014, doi: 10.48550/arXiv.1412.6980.

[4] M. S. Ali, M. S. Miah, J. Haque, M. M. Rahman, and M. K. Islam, "An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models," *Machine Learning with Applications*, vol. 5, p. 100036, Sep. 2021, doi: 10.1016/j.mlwa.2021.100036.

[5] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using Convolutional Neural Networks," in *2016 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2016, pp. 2560–2567. doi: 10.1109/IJCNN.2016.7727519.

[6] N. M. Astudillo, R. Bolman, and N. M. Sirakov, "Classification with Stochastic Learning Methods and Convolutional Neural Networks," *SN COMPUT. SCI.*, vol. 1, no. 3, p. 119, Apr. 2020, doi: 10.1007/s42979-020-00126-x.

[7] T.-C. Pham, C.-M. Luong, M. Visani, and V.-D. Hoang, "Deep CNN and Data Augmentation for Skin Lesion Classification," in *Intelligent Information and Database Systems*, Cham, 2018, pp. 573–582. doi: 10.1007/978-3-319-75420-8_54.

[8] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, Art. no. 7639, Feb. 2017, doi: 10.1038/nature21056.

[9] R. R. Winkelmann, A. S. Farberg, N. Tucker, R. White, and D. S. Rigel, "Enhancement of International Dermatologists' Pigmented Skin Lesion Biopsy Decisions Following Dermoscopy with Subsequent Integration of Multispectral Digital Skin Lesion Analysis," *J Clin Aesthet Dermatol*, vol. 9, no. 7, pp. 53–55, Jul. 2016.

[10] U.-O. Dorj, K.-K. Lee, J.-Y. Choi, and M. Lee, "The skin cancer classification using deep convolutional neural network," Multimed Tools Appl, vol. 77, no. 8, pp. 9909–9924, Apr. 2018, doi: 10.1007/s11042-018-5714-1.M. Mateen, J. Wen, Nasrullah, S. Song, and Z. Huang, "Fundus Image Classification Using VGG-19 Architecture with PCA and SVD," *Symmetry*, vol. 11, no. 1, Art. no. 1, Jan. 2019, doi: 10.3390/sym11010001.

[11] Q. Gao, S. Lim, and X. Jia, "Hyperspectral Image Classification Using Convolutional Neural Networks and Multiple Feature Learning," *Remote Sensing*, vol. 10, no. 2, Art. no. 2, Feb. 2018, doi: 10.3390/rs10020299.

[12] Z. A. Oraibi, H. Yousif, A. Hafiane, G. Seetharaman, and K. Palaniappan, "Learning Local and Deep Features for Efficient Cell Image Classification Using Random Forests," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Oct. 2018, pp. 2446–2450. doi: 10.1109/ICIP.2018.8451287.

[13] D. Theckedath and R. R. Sedamkar, "Detecting Affect States Using VGG16, ResNet50 and SE-ResNet50 Networks," SN COMPUT. SCI., vol. 1, no. 2, p. 79, Mar. 2020, doi: 10.1007/s42979-020-0114-9.

[14] K. M. Hosny, M. A. Kassem, and M. M. Fouad, "Classification of Skin Lesions into Seven Classes Using Transfer Learning with AlexNet," J Digit Imaging, vol. 33, no. 5, pp. 1325–1334, Oct. 2020, doi: 10.1007/s10278-020-00371-9.

[15] I. Z. Mukti and D. Biswas, "Transfer Learning Based Plant Diseases Detection Using ResNet50," in 2019 4th International Conference on Electrical Information and Communication Technology (EICT), Dec. 2019, pp. 1–6. doi: 10.1109/EICT48899.2019.9068805.

[16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2018, pp. 4510–4520. Accessed: Sep. 28, 2022. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2018/html/Sandler_MobileNetV2_Inverted_Residuals_CVPR_2018_paper.html

[17] D. A. van Dyk and X.-L. Meng, "The Art of Data Augmentation," Journal of Computational and Graphical Statistics, vol. 10, no. 1, pp. 1–50, Mar. 2001, doi: 10.1198/10618600152418584.

[18] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[19] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 173, pp. 24–49, Mar. 2021, doi: 10.1016/j.isprsjprs.2020.12.010.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun. ACM, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[21] H. Dalianis, "Evaluation Metrics and Evaluation," in Clinical Text Mining: Secondary Use of Electronic Patient Records, H. Dalianis, Ed. Cham: Springer International Publishing, 2018, pp. 45–53. doi: 10.1007/978-3-319-78503-5_6.

[22] S. Banerjee, S. K. Singh, A. Chakraborty, A. Das, and R. Bag, "Melanoma Diagnosis Using Deep Learning and Fuzzy Logic," *Diagnostics*, vol. 10, no. 8, Art. no. 8, Aug. 2020, doi: 10.3390/diagnostics10080577.

[23] J. Rashid et al., "Skin Cancer Disease Detection Using Transfer Learning Technique," Applied Sciences, vol. 12, no. 11, Art. no. 11, Jan. 2022, doi: 10.3390/app12115714.

[24] R. Indraswari, R. Rokhana, and W. Herulambang, "Melanoma image classification based on MobileNetV2 network," Procedia Computer Science, vol. 197, pp. 198–207, Jan. 2022, doi: 10.1016/j.procs.2021.12.132.