Binary Classification of Skin Cancer using Pretrained Deep Neural Networks

Hadeer Hussein[1], Ahmed Magdy[1], Rehab Farouk[2], Khaled Abd Elsalam Ali [1]

[1] *Electrical Engineering Department, Faculty of Engineering, Suez Canal University, Ismailia, Egypt.*

[2] *Electrical Engineering Department, Faculty of Engineering, Port Said University, Port Said, Egypt.*

Abstract One of the most frequent kinds of cancer in the world is skin cancer. Clinical examination of skin lesions is essential to detect disease characteristics, but it is limited by long timeframes and a broad variety of interpretations. Computer vision is being used to detect diseases, help in diagnosis, and identify patient risks. This is particularly true for skin cancer, which may be lethal if not detected early on. Several computer-aided diagnosis and detection systems have already been developed to do this. Deep learning techniques have been developed to address these issues and assist dermatologists, as early and precise detection of skin cancer is critical to improve patient survival rates. In this paper, some pretrained deep neural networks are utilized for binary classification of skin cancer disease. They are used to classify between benign and malignant cancers in dermoscopic images. AlexNet, ResNet-18, SqueezeNet, and ShuffleNet are the used networks as transfer learning classifiers. In this study, we employed a Kaggle dataset titled "Skin Cancer: Malignant vs. Benign". The networks' maximum accuracy approaches 89%.

Keywords Binary classification, Deep learning, Pretrained deep neural networks, skin cancer.

I. INTRODUCTION

According to various data, cancer is the leading cause of death among people. Skin cancer is the most common type; it usually develops in skin that has been exposed to sunlight on a regular basis, although cancer can develop elsewhere on the body. Skin cancer is easily visible because it develops in the epidermis, the topmost layer of skin [1].

Melanoma is the worst type of skin cancer known in humans, causing pigmented moles to form on the skin [2]. Melanoma is produced by anomalies in the melanin-producing cells, which are responsible for the skin's pigment. A history of sunburn, a weakened immune system, pale skin, hereditary factors, and inappropriate exposure to UV light are all risk factors for melanoma [3].

Melanoma starts in the outer skin layer and spreads to the interior layers, where it finally connects with the blood and lymph arteries. When skin cancer is detected in its early stages, it has a better chance of being cured than when it is diagnosed later. Early detection of skin cancer, on the other hand, is expensive [4].

Because skin lesions are so similar, determining whether a lesion is malignant or benign can be difficult. A regular mole is frequently the same color as the skin, such as brown, black, or tan, with a distinct border that differentiates it from neighboring skin. Numerous systems, such as genetic algorithms, Artificial Neural Networks (ANNs), and Convolutional Neural Networks (CNNs), have been developed to assess skin discomfort and classify it as melanoma or benign [4]. All these methods have been shown to be less expensive, more efficient, and less painful than standard medical treatments. However, in many computer vision applications, CNNs and deep learning are the favored methodologies [5].

Deep neural networks that have been pre-trained are networks that have already learned to extract powerful and valuable properties from real-world images and use them as a starting point for learning a new task. They have been used to improve performance and reduce computing costs in a range of fields. These networks have been widely used in image classification field [6].

In this study, we identify and classify skin cancer using four pretrained networks: AlexNet, ResNet-18, SqueezeNet, and ShuffleNet as Transfer Learning (TL) classifiers. We used a Kaggle dataset in the classification experiments. The remaining sections of the paper are organized as follows. Section 2 contains studies on skin cancer categorization and detection that are relevant. Section 3 details the materials and procedures employed. Section 4 discusses the experimental outcomes. Finally, Section 5 demonstrates the research's conclusion.

II. RELATED WORKS

Magdy et al. [7] proposed two ways for recognizing and classifying benign and malignant tumors in dermoscopic images. The first strategy leverages K-Nearest Neighbor (KNN) as a classifier, with PDLNs functioning as feature extractors. The second strategy optimizes its hyperparameters by merging AlexNet and grey wolf optimizer. The scientists also employed artificial neural networks, support vector machines, and CNNs to study two strategies for categorizing skin cancer images: Machine Learning (ML) and Deep Learning (DL). The studies were conducted on 4000 images from the ISIC archive collection, and the proposed methods beat other examined approaches, with some models achieving an accuracy of more than 99%. Mazoure et al. [8] proposed DUNEScan, a deep neural network-based online tool for analyzing uncertainty in Skin Cancer Diagnosis (SCD), was suggested. This method used many CNN models to predict skin cancer, including ResNet50T, EfficientNet, Inceptionv3, and MobileNetv2, and diagnostic uncertainty was determined using the following measures: average, and variance of learning models.

To diagnose SC, Tabrizchi et al. [9] employed an improved CNN model based on the visual geometry group’s VGG-16 architecture. The VGG model’s design was altered in this study so that it is more compatible with SCD diseases and can be detected with more accuracy than the original model. This upgrade includes changes to filter dimensions and ANN activation functions. Reis et al. [10] used CNNs to detect SCD and lesion sites. The input images are pre-processed in this model before being segmented with the UNet network. Depending on the segmentation findings, the lesion region is clipped, and this segment is used as the input for a CNN model called InSiNet to classify the input picture. Kousis et al. [11] researched Deep Learning (DL) algorithms and a smartphone app for effective skin cancer detection. They presented the XGBoost, a composite of the best eight DL models, and a composite of fifteen DL models.

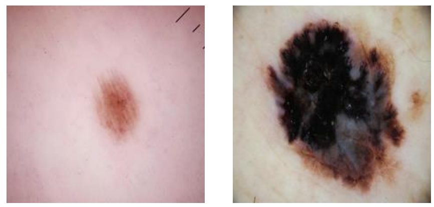
Shorfuzzaman [12] treated SCD with a DL ensemble model. This model, which incorporates several incomplete CNN classifications running concurrently in the form of an ensemble system, used the TL strategy. Finally, the results of these models were merged, and the final output was established using an integration model. Nawaz et al. [13] presented DL techniques for melanoma diagnosis. To extract visual attributes, CNN was used in this technique. These properties are then fed into two ANN models, one of which is a CNN and was used to identify the target sites. Furthermore, the second NN is a recurrent CNN, which locates the lesion. Finally, the obtained location was used to segment the lesion using the Fuzzy K-Means (FKM) approach. Thurnhofer-Hemsi and Domnguez [14] demonstrated a CNN architecture for detecting skin cancer. They stated that the findings of the DenseNet201 network were suitable for this application.

Manne et al. [15] presented a CNN-based categorization method for skin cancer. They demonstrated a totally automated computer system for classifying skin lesions. Three models were pre-trained to function as feature generators in this study: ResNet-18, AlexNet, and VGG16. After that, the retrieved characteristics are utilized to train support vector machines. Skin lesions were segmented and classified pixel by pixel using the suggested CNN. CNNs can segment, identify, and categorize skin lesions, according to Song et al. [16]. They used a loss function based on the Jaccard distance and the focal loss to regulate the unbalanced datasets.

III. MATERIALS AND METHODS

3.1 Dataset

In this study, a dataset called "Skin Cancer: Malignant vs. Benign" from Kaggle is used [17]. Kaggle is a great place to get datasets for data scientists and machine learners. This dataset offers a balanced collection of images of benign and malignant skin moles. This dataset contains 3297 pictures. The dataset is separated into two sections: train (2637 images) for training the models and test (660 images) for testing the accuracy of the trained models. Malignant (total of 1497 images) and benign (total of 1800 images) cancer images may be found in both sections. The image dimensions are (224×224). All data rights are related to the ISIC-Archive rights [18]. Figure 1 shows samples of this dataset.



(a) (b)

Figure 1. Samples of Kaggle dataset. (a) benign; and (b) malignant.

3.2 Pretrained networks

In this article, we used pretrained networks as classifiers to differentiate between benign and malignant skin cancer images. A pretrained model is a previously trained stored network on a big dataset, generally on a large-scale image-classification problem. We used four different networks, which are listed below: AlexNet, ResNet-18, SqueezeNet, and ShuffleNet.

3.2.1 AlexNet

AlexNet is a CNN model that has a significant impact on deep learning applications in computer vision. It comfortably won the ImageNet LSVRC-2012 competition in 2012 (15.3% botch rates versus 26.2% blunder rates in the runner-up, which is VGG-16). The configuration of the organization was like Yann LeCun et alLeNet, but deeper, with more channels per layer and stacked convolutional layers. Convolutions, maximum pooling, dropout, information growth, ReLU initiations, and stochastic gradient descent with force all played significant roles. After each convolutional and totally related layer, it adds ReLU initiations. Furthermore, dropout is used to cope with overfitting rather than regularization [19].

3.2.2 ResNet-18

The ResNet-50 model won the ILSVRC-2015 competition with a 3.57% error rate and an input picture size of 224×224 pixels. ResNet, the well-known deep learning model, was created by Shaoqing Ren, Kaiming He, Jian Sun, and Xiangyu Zhang. ResNet-18 has 18 layers, but ResNet-50 has 50 layers, with two or three convolutional layers in each layer [20].

3.2.3 SqueezeNet

SqueezeNet features 68 layers with 1.2 million learnable parameters, including numerous 2D convolution, ReLU, max-pooling, and concatenation layers. To avoid overfitting, a dropout layer is also provided. Over a million pictures from the ImageNet database of various objects are used to retrain the pretrained SqueezeNet network. It has a large feature set and a greatly reduced architecture, allowing it to attain greater accuracies while using less computational power and training time. SqueezeNet has performed well with TL in several studies [21].

3.2.4 ShuffleNet

ShuffleNet, a CNN architecture tailored for mobile devices with 10-150 MFLOPs of computing capability, has been announced by Megvii Inc (also known as Face++). The ShuffleNet utilizes pointwise group convolution and channel shuffling to reduce computation costs while maintaining accuracy. On ImageNet classification, it has a smaller top-1 error than the MobileNet system and a 13x real-time speedup over AlexNet while maintaining comparable accuracy [22].

This work uses TL to test pretrained deep networks as classifiers. TL is a ML technique that repurposes a model created for one task for another. It is frequently used when there is a paucity of training data. Data augmentation, on the other hand, can help overcome the data challenge. Because malignant and benign lesions are so similar, distinguishing and classifying them takes a long time, which is why we require TL. Because TL is more effective at classifying associated lesions, it is the preferred technique. TL networks are trained on enormous datasets, and their model weights are fixed before changing the last few layers for a different dataset. To start learning a new task, a pretrained network might be employed. TL is faster and easier to use than manually training a network with randomly supplied weights [23].

IV. RESULTS AND DISCUSSIONS

4.1 System Implementation

The implemented frameworks were established and evaluated on the following software and hardware configurations:

- <u>Processor:</u> Intel (R) Core (TM) i7-9750H CPU @ 2.60GHz 2.59 GHz.
- <u>Operating system:</u> Windows 10 Pro.
- <u>Installed RAM:</u> 16.0 GB (15.9 GB usable).
- <u>System type:</u> 64-bit operating system, x64-based processor.
- <u>Compiler:</u> MATLAB R2020b.

4.2 Performance Metrics

We used five metrics to estimate the networks' performance:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{F1 score} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{2}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{4}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5}$$

- <u>True Positive (TP):</u> is the number of accurately classified malignant samples.
- <u>True Negative (TN):</u> is the number of samples accurately identified as benign.
- <u>False Positive (FP):</u> is the number of benign samples misdiagnosed as malignant.
- <u>False Negative (FN):</u> is the number of malignant samples that were misidentified as benign.

4.3 Experimental results

The results and parameters used in the tested networks are explained in this section.

We began by loading data, and then ran a pretrained network. The network's convolutional layers gathered visual characteristics, which the final learnable and classification layers used to classify the input image. The output classes of the network are determined by the classification layer. We replaced the classification layer with a new one that did not include class labels. We set the learning rates of the network's previous layers to zero in order to freeze their weights. The settings of the frozen layers were not changed by the network during training. Freezing the weights of many early layers can substantially speed up network training since the gradients of the frozen layers do not need to be calculated.

We investigated four pretrained networks: AlexNet, ResNet-18, SqueezeNet, and ShuffleNet. As TL classifiers, these networks were tested. We utilized an augmented image data store to automatically resize the training images because each network required a different size of input image. On the training images, we specified additional augmentation techniques such as randomly reflection along the horizontal axis, randomly translation up to 30 pixels, horizontally scaling and vertically scaling up to 10%, and rotation up to 30%. The following are the network training hyperparameters: Maximum epoch count of 100, mini-batch size of 32, and initial learning rate of 1×10^{-4}.

Figures 2 and 3 depict the ResNet-18 confusion matrix and AlexNet training progress, respectively. Table 1 shows the performance measurements of the pretrained networks in skin cancer classification. In detecting skin cancer classifying into benign and malignant, Table 1 displays that ResNet-18 attains the highest accuracy of 89.3939%, 87.0968% precision, 88.5246% f1 score, 90% sensitivity, and 88.8889% specificity. AlexNet achieves 88.3333% accuracy, 87.0432% precision, 87.188% f1 score, 87.3333% sensitivity, and 89.1667% specificity. SqueezeNet achieves 85.9091% accuracy, 87.6364% precision, 83.8261% f1 score, 80.3333% sensitivity, and 90.5556% specificity. ShuffleNet achieves 87.4242% accuracy, 85.1133% precision, 86.3711% f1 score, 87.6667% sensitivity, and 87.2222% specificity.

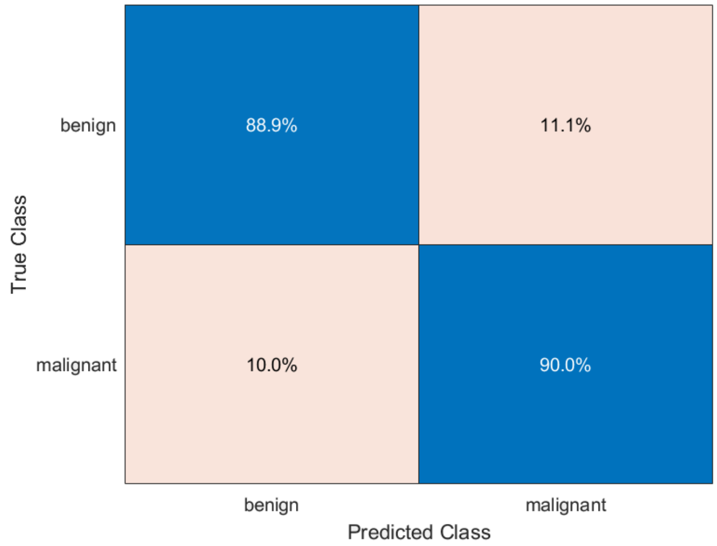


Figure 2. Confusion matrix of ResNet-18

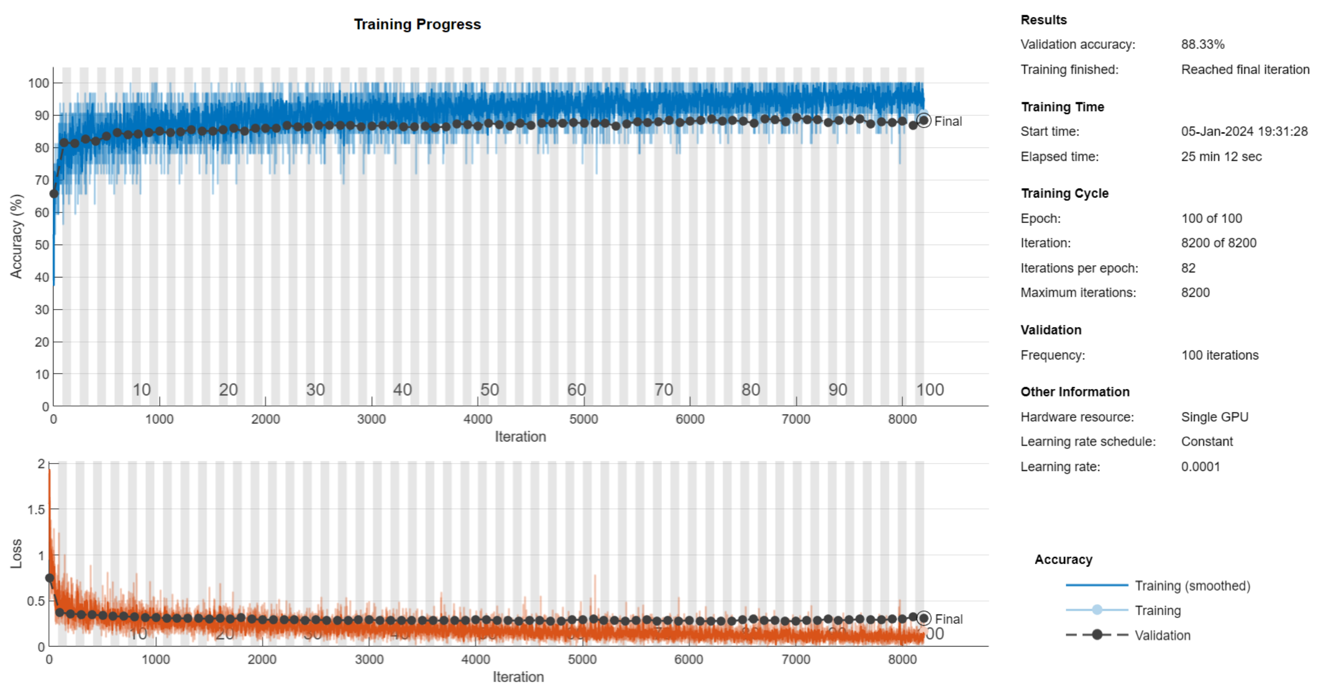


Figure 3. Training Progress of AlexNet

Table 1. Performance of pretrained networks in binary classification of skin cancer

Pretrained networks	Precision (%)	F1 score (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
AlexNet	87.0432	87.188	87.3333	89.1667	88.3333
ResNet-18	87.0968	88.5246	90	88.8889	89.3939
SqueezeNet	87.6364	83.8261	80.3333	90.5556	85.9091
ShuffleNet	85.1133	86.3711	87.6667	87.2222	87.4242

V. CONCLUSION

Medical illness classification research seeks to benefit patients and physicians. Using computer-based solutions allows doctors to make better decisions. A precise technique fosters trust in a disease diagnosis. It might be difficult to discern between benign and malignant lesions when diagnosing skin cancer. Skin cancer is one of the worst cancers in the world. Early discovery and diagnosis of skin lesions is crucial for choosing the best course of treatment for the patient and, in the case of malignant lesions, improving the patient's chances of survival. In this paper, we tested four pretrained networks, which are AlexNet, ResNet-18, SqueezeNet, and ShuffleNet as TL classifiers. These networks were utilized in dermoscopic images to detect and classify benign and malignant cancers. They were tested on samples from the Kaggle dataset named "Skin Cancer: Malignant vs. Benign". These networks have a maximum detection accuracy of 89% in detecting skin cancer, according to the outcomes.

References

[1] H. E. Kanavy and M. R. Gerstenblith, "Ultraviolet Radiation and Melanoma," *Seminars in Cutaneous Medicine and Surgery,* vol. 30, p. 222–228, December 2011.

[2] B. G. Goldstein and A. O. Goldstein, "Diagnosis and management of malignant melanoma," *American family physician,* vol. 63, p. 1359–1369, 2001.

[3] Y. Goncharova, E. A. S. Attia, K. Souid and I. V. Vasilenko, "Dermoscopic Features of Facial Pigmented Skin Lesions," *ISRN Dermatology,* vol. 2013, p. 1–7, February 2013.

[4] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature,* vol. 542, p. 115–118, January 2017.

[5] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing,* vol. 187, p. 27–48, April 2016.

[6] X. Han, Z. Zhang, N. Ding, Y. Gu, X. Liu, Y. Huo, J. Qiu, Y. Yao, A. Zhang, L. Zhang, W. Han, M. Huang, Q. Jin, Y. Lan, Y. Liu, Z. Liu, Z. Lu, X. Qiu, R. Song, J. Tang, J.-R. Wen, J. Yuan, W. X. Zhao and J. Zhu, "Pre-trained models: Past, present and future," *AI Open,* vol. 2, p. 225–250, 2021.

[7] A. Magdy, H. Hussein, R. F. Abdel-Kader and K. A. E. Salam, "Performance Enhancement of Skin Cancer Classification Using Computer Vision," *IEEE Access,* vol. 11, p. 72120–72133, 2023.

[8] B. Mazoure, A. Mazoure, J. Bédard and V. Makarenkov, "DUNEScan: a web server for uncertainty estimation in skin cancer detection with deep neural networks," *Scientific Reports,* vol. 12, January 2022.

[9] H. Tabrizchi, S. Parvizpour and J. Razmara, "An Improved VGG Model for Skin Cancer Detection," *Neural Processing Letters,* vol. 55, p. 3715–3732, July 2022.

[10] 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, p. 643–662, January 2022.

[11] I. Kousis, I. Perikos, I. Hatzilygeroudis and M. Virvou, "Deep Learning Methods for Accurate Skin Cancer Recognition and Mobile Application," *Electronics,* vol. 11, p. 1294, April 2022.

[12] M. Shorfuzzaman, "An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection," *Multimedia Systems,* vol. 28, p. 1309–1323, April 2021.

[13] M. Nawaz, Z. Mehmood, T. Nazir, R. A. Naqvi, A. Rehman, M. Iqbal and T. Saba, "Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering," *Microscopy Research and Technique,* vol. 85, p. 339–351, August 2021.

[14] K. Thurnhofer-Hemsi and E. Domínguez, "A Convolutional Neural Network Framework for Accurate Skin Cancer Detection," *Neural Processing Letters,* vol. 53, p. 3073–3093, October 2020.

[15] R. Manne, S. Kantheti and S. Kantheti, "Classification of Skin cancer using deep learning, ConvolutionalNeural Networks-Opportunities and vulnerabilities-A systematic Review," *International Journal for Modern Trends in Science and Technology, ISSN,* vol. 6, p. 2455–3778, 2020.

[16] L. Song, J. Lin, Z. J. Wang and H. Wang, "An End-to-End Multi-Task Deep Learning Framework for Skin Lesion Analysis," *IEEE Journal of Biomedical and Health Informatics,* vol. 24, p. 2912–2921, October 2020.

[17] C. Fanconi, "Skin cancer: malignant vs. Benign," [Online]. Available: https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign/data (accessed on 2 January 2024).

[18] ISIC-Archive, "Content and Layout of the Archive," [Online]. Available: https://www.isic-archive.com/#!/topWithHeader/tightContentTop/about/isicArchiveContent.

[19] S. Suganyadevi and V. Seethalakshmi, "DarkCVNet: Optimized Pneumonia and Covid 19 detection using CXR Images," July 2022.

[20] K. Vyshnavi, A. Kintali, K. Sravan, V. Kothuru and V. Gowtham, *Performance Analysis of Pneumonia Detection Using ResNets Technique,* 2021.

[21] A. Saad, A. Usman, S. Arif and M. Liwicki, "Bearing Fault Detection Scheme Using Machine Learning for Condition Monitoring Applications," in *ICMAME 2023 Conference Proceedings*, 2023.

[22] X. Zhang, X. Zhou, M. Lin and J. Sun, "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[23] S. Jain, U. Singhania, B. Tripathy, E. A. Nasr, M. K. Aboudaif and A. K. Kamrani, "Deep Learning-Based Transfer Learning for Classification of Skin Cancer," *Sensors,* vol. 21, p. 8142, December 2021.