 doi:10.20944/preprints202209.0215.v1

Article

Skin Cancer Classification and Comparison of Pretrained Models Performance using Transfer Learning

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

[1] Division of Enrollment Management, Texas A&M University-Commerce, 2600 W Neal St., Commerce, TX 75428, USA; ssingha@loemail.tamuc.edu
[2] Department of Marketing & Business Analytics, Texas A&M University-Commerce, 2600 W Neal St., Commerce, TX 75428, USA; proy@leomail.tamuc.edu
* Correspondence: ssingha@loemail.tamuc.edu; Tel.: +1-903-886-5076

Abstract: Skin cancer is an uncommon but serious malignancy. Dermoscopic images examination and biopsy are required for cancer detection. Deep learning (DL) is extremely effective in learning characteristics and predicting malignancies. However, DL requires a large number of images to train. Image augmentation and transferring learning were employed to overcome the lack of images issue. In this study we divided images into two categories: benign and malignant. To train and test our models, we used the public ISIC 2020 database. Melanoma is classified as malignant in the ISIC 2020 dataset. Along with categorization, the dataset was studied to demonstrate variation. The performance of three top pretrained models was then benchmarked in terms of training and validation accuracy. Three optimizers were employed to optimize the loss: RMSProp, SGD, and ADAM. Using ResNet, VGG16, and MobileNetV2, we obtained training accuracy of 98.73%, 99.12%, and 99.76%, respectively. Using these three pretrained models, we attained a validation accuracy of 98.39%.

Keywords: pretrained model; transfer learning; skin cancer; deep learning; ISIC 2020

1. Introduction

One of the most frequent types of cancer is skin cancer. Melanoma is responsible for 75% of skin cancer fatalities, according to the American Cancer Society. As a result, dermatologists examine each patient's moles for melanoma, the most common kind. UV radiation causes melonocyte cells in the human body to be damaged, leading in melanoma, a kind of skin cancer. Skin cancer is the most frequent malignancy, accounting for one-third of all malignancies globally, according to the World Health Organization (WHO). Basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and malignant melanoma are the most common kinds (MM). Non-melanocytic cancer is caused by BCC and SCC, and the majority of skin cancers are non-melanocytic, although melanoma is the deadliest because it spreads quickly if not diagnosed and treated early. Thus, several organs in the human body, such as malignant genes, hair color, and an increased incidence of benign melanocytic nevi and dysplastic nevi, are implicated in Melanoma diagnosis. Skin cancer is really caused by abnormal melanocyte cell growth, which affects surrounding tissues by multiplying and spreading through lymph nodes [1]. Skin cancer is presently a major public health concern, with over 123,000 new cases diagnosed each year worldwide. Surprisingly, melanoma accounts for 75% of skin cancer mortality. Furthermore, according to the American Cancer Society, 100,000 new instances of melanoma will be detected by the end of 2020, with 7,000 people dying from the condition. Melanoma is responsible for around 9000 deaths in the United States each year [2]. However, if this cancer is not detected early, the cost of treatment is significant, costing over $134,000 in its fourth stage [3]. Dermatologists first detect melanoma by examining photos of the cancer and moles, among other treatment options. They also check for "ugly ducklings" or outlier lesions around the moles that might be melanoma. If they're lucky, the outcomes will be precise and accurate. As a result, artificial intelligence (AI) has the potential to assist physicians in accurately identifying melanoma. AI-based detection technologies have the potential to increase dermatological accuracy. Existing AI approaches have not sufficiently considered this clinical environment. Dermatologists may increase their diagnostic accuracy if detection algorithms use into account "contextual" images inside the same patient to determine which shots depict a melanoma. If the AI detects skin cancer, it will have a significant influence on dermatological clinic operations. Convolutional Neural Network (CNN) has previously performed admirably 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 even tens of thousands of images, CNNs are remarkably accurate at identifying medical images. Skin lesion photographs, in our situation, are sufficient for skin cancer categorization into two binary classes: benign and malignant. In this study our contributions are as follows:

- To get the best accuracy, we used three pretrained models VGG, ResNet, and MobileNetV2;
- We utilized the ISIC 2020 database [10], a 48.9 GB dataset of 33,126 jpeg pictures in 1024x1024 pixels, 533 of which are malignant and the rest benign;
- We used SGD, ADAM, and RMSProp three optimization algorithms to minimize the loss;
- To solve the overfitting concerns, we applied a dropout layer and data augmentation;
- We evaluated the model’s performance using accuracy;
- We analyzed the ISIC dataset in order to see location of skin cancer, type of disease that causes cancer etc.

2. Materials and Methods

2.1 Background

Skin cancer was first discovered using human eyes in the early twentieth century. [11] identified a few examples of such approaches, including 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 was conducted to see if a practitioner was unable to determine whether a mole is melanoma or not [11]. Several CNN-based classifiers have recently entered the picture, aiding dermatologists in better identifying melanoma. In fact, [3] identified skin lesion images as benign or malignant with 95.23 percent 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. They employed SGD and ADAM to 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 that 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 were 0.77, 0.93, 0.85, and 0.86, respectively. Their model has an average accuracy of 97.49 percent. Their technique benefited doctors in categorizing various skin lesions. [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 beat the other CNNs. In this study, 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 grouping use cases. 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 percent.

2.1 Methodology

This section discusses data preparation, data augmentation, the architectures of pretrained models, and evaluation metrics. Three pretrained models' performance was compared using ISIC 2020 images.

2.1.1. Data Preprocessing

Labeling the images is the first stage in data preparation. Each image in the ISIC 2020 dataset was labeled with a benign or malignant target. We divided the dataset into two parts: train and validation. The pretrained 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.

2.1.2. Data augmentation

Data augmentation is a strategy for correcting data imbalance. The most prevalent approaches used in data augmentation are oversampling and undersampling. To correct the class imbalance, oversampling uses exact copy duplicates or modified copies of the original data from the minority class. 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.

2.1.3. Pretrained model's architectures

Convolutional Neural Networks (CNN) are excellent in image classification and detection. CNNs generate feature maps by stacking convolutional layers and pooling layers. Feature maps are high-level features derived from CNNs at each level of convolution. As an output layer for picture categorization, a final fully linked layer is employed. However, the layer stack in CNNs can vary greatly, and there is no previous 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. Thus, transfer learning is a solution in which we employ pretrained models that have previously been trained on some dataset and we can simply use the pretrained weights and put our layers on top of the pretrained architecture as per our issue needs. We used three common pretrained models in our methodology: ResNet, VGG, and MobileNet. Imagenet was used to train all of these models. Four layers were added: a flatten layer, a dense layer, a dropout layer, and an output dense 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 crossentropy. Pretrained models such as VGG, ResNet, and MobileNetV2 are frequently utilized in transfer learning for feature extraction or fine tuning. Each model is trained on ImageNet, abbreviated ILSVRC for ImageNet Large Scale Visual Recognition Challenge. The goal of this image classification challenge is 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 object classifications we encounter in our daily lives, such as dog and cat breeds, household products, automobile types, and so on. As a result, pretrained models are highly excellent at extracting features with high accuracy while requiring less training time.

2.1.4. VGGNet Architecture

The University of Oxford's Visual Geometry Group (VGG) releases VGGNet, in which they obtained top performance in the ImageNet dataset, which has 1000 classes. The key contribution of VGGNet was its smaller filter (3x3) in comparison 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 in order 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 more parameters. We utilized VGG16 to train the ISIC 2020 dataset in our work.

2.1.5. ResNet

Other designs 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 to bypass one or more layers dubbed 'identity shortcut connections.'

2.1.6. MobileNetV2

Google introduced the MobileNetV2 concept. Because of its lightweight and minimal complexity, this architecture is appropriate for mobile devices. Version 1 of MobileNet featured depthwise separable convolution, whereas version 2 introduced a superior module called inverted residual.

2.1.7. Evaluation Metrics

We used accuracy as an assessment criterion to compare the performance of the pretrained models on the ISIC 2020 dataset. We did not evaluate additional metrics like F1 score, Precision, Recall, and so on since our major purpose was to assess the performance of pretrained 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.

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

,

3. Results

3.1. Dataset

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 is 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. Figure 1 depicts the position of the skin from which the images were taken.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2022 doi:10.20944/preprints202209.0215.v1

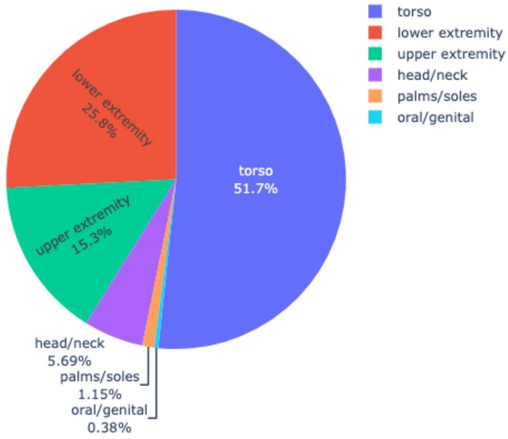

Figure 1. Location of skin.

Dermatologists classified benign and malignant skin tumors into nine categories. Figure 2 depicts the disease's eight classifications; the unknown class was not included.

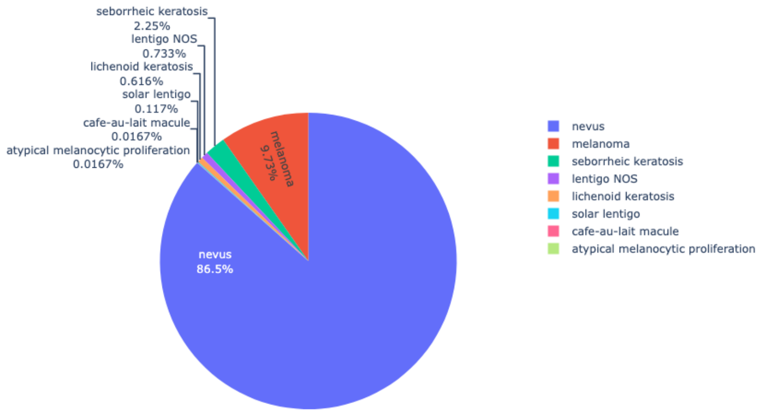

Figure 2. Various classes of disease.

86.5% of the individuals with moles had nevus disease, which was classed as benign. Melanoma patients were diagnosed with malignant cancers. Table 1 explains how nine types of disorders are classified as benign or malignant skin malignancies. Figure 3 shows sample images of the benign and malignant classes.

Table 1. Disease classes divided into benign and malignant.

Disease Classes	Benign	Malignant
atypical melanocytic proliferation	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

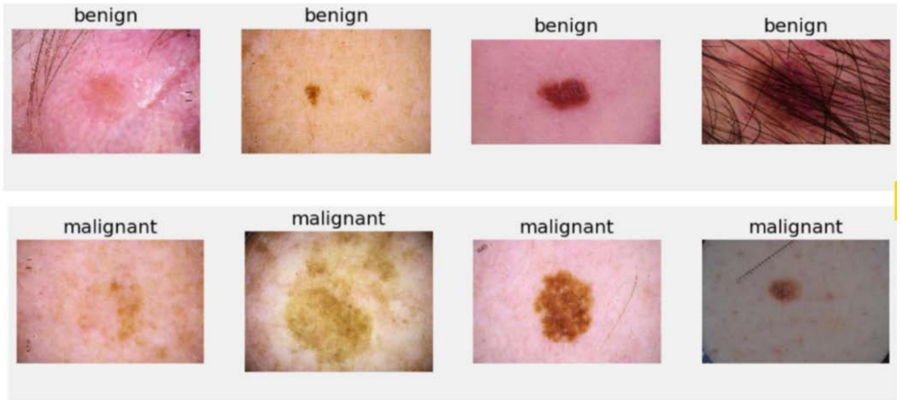

Figure 3. Images of Benign and Malignant Skin Cancers.

3.2 Experimental Setup

Figure 4 depicts the whole experimental setup. We employed data preparation first, followed by data augmentation in the second stage. Three models were chosen for pretrained training. Except for the output layer, we used all layers. A flatten 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.

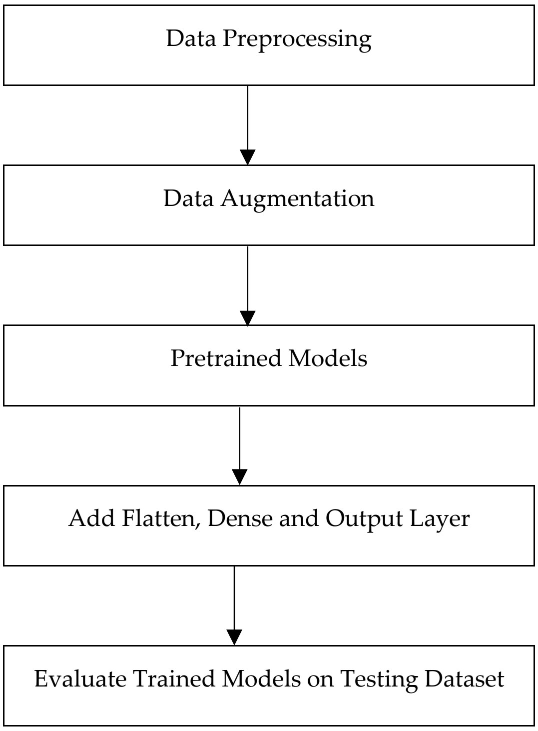

Figure 4. Experimental Setup.

Table 2. Summary of the result of the experiment

Pretrained Model	**Training Accuracy**	**Validation Accuracy**

 doi:10.20944/preprints202209.0215.v1

VGG16	0.9912	
ResNet	0.9873	0.9839
MobileNetV2	0.9976	

4. Discussion

Using a massive dataset of 33,126 images, we achieved a validation accuracy of 98.39%. Using the pretrained model AlexNet, they attained an accuracy of 0.7853 and 0.9086 in [15] and [16]. Similarly, in [15], ResNet, VGGNet, and Xception were used to obtain accuracy of 0.9208, 0.8870, and 0.9030, respectively. Table 3 displays testing accuracy from the literature. The greatest accuracy across all of these pretrained models was 0.9420 in [17]. One pretrained model was utilized in all of these architectures by integrating stages such as data augmentation, data standardization, and so on. [16] employed a hybrid of AlexNet, deep convolutional neural network (DCNN), and support vector machine (SVM). However, none of the designs outperformed our attained accuracy. Furthermore, we tackled the challenges of overfitting and data preparation while minimizing the need to develop models from scratch.

Table 3. Pretrained models accuracies in literature.

Pretrained Model Family	Accuracy
AlexNet	0.7853[15], 0.9086[16]
ResNet	0.9208[15], 0.9093[16]
VGGNet	0.8870[15], 0.8602[16]
Xception	0.9030[15]
DenseNet	0.8833[16]
MobileNet	0.8826[16]
DCNN	0.9143[16]
AlexNet + SVM + DCNN	0.9420[17]

5. Conclusions

Melanoma is one of the most serious skin malignancies, shortening people's lives dramatically. However, early identification avoids any problematic issue. AI can help detect this sort of cancer at an early stage. In this work, we employed pretrained models to compare performance while taking an evaluation metric into account (accuracy). We used RMSProp, SGD, and ADAM to optimize the models. Pretrained models were employed to get the maximum accuracy while spending the least amount of time creating models from scratch. Furthermore, we addressed the issue of overfitting and offered alternative data processing techniques with dataset insights. We achieved a validation accuracy of 98.39%, outperforming the prior pretrained 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: Conceptualization, S.S.; methodology, S.S. and P.R.; software, S.S.; validation, S.S. and P.R.; formal analysis, S.S.; investigation, S.S.; resources, S.S. and P.R.; data curation, S.S.; writing—original draft preparation, S.S. and P.R.; writing—review and editing, P.R. and S.S.; visualization, S.S.; supervision, S.S.; project administration, P.R.; funding acquisition, S.S. and P.R. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

 doi:10.20944/preprints202209.0215.v1

Data Availability Statement: The datasets used or analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest: We declare that there is no conflict of interest.

References

1. Kaur, R.; GholamHosseini, H.; Sinha, R.; Lindén, M. Melanoma Classification Using a Novel Deep Convolutional Neural Network with Dermoscopic Images. Sensors 2022, 22, 1134, doi:10.3390/s22031134.
2. Harangi, B. Skin Lesion Classification with Ensembles of Deep Convolutional Neural Networks. Journal of Biomedical Informatics 2018, 86, 25–32, doi:10.1016/j.jbi.2018.08.006.
3. Astudillo, N.M.; Bolman, R.; Sirakov, N.M. Classification with Stochastic Learning Methods and Convolutional Neural Networks. SN COMPUT. SCI. 2020, 1, 119, doi:10.1007/s42979-020-00126-x.
4. Spanhol, F.A.; Oliveira, L.S.; Petitjean, C.; Heutte, L. Breast Cancer Histopathological Image Classification Using Convolutional Neural Networks. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN); July 2016; pp. 2560–2567.
5. Gao, Q.; Lim, S.; Jia, X. Hyperspectral Image Classification Using Convolutional Neural Networks and Multiple Feature Learning. Remote Sensing 2018, 10, 299, doi:10.3390/rs10020299.
6. Bora, K.; Chowdhury, M.; Mahanta, L.B.; Kundu, M.K.; Das, A.K. Pap Smear Image Classification Using Convolutional Neural Network. In Proceedings of the Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing; Association for Computing Machinery: New York, NY, USA, December 18 2016; pp. 1–8.
7. Brinker, T.J.; Hekler, A.; Utikal, J.S.; Grabe, N.; Schadendorf, D.; Klode, J.; Berking, C.; Steeb, T.; Enk, A.H.; Kalle, C. von Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review. Journal of Medical Internet Research 2018, 20, e11936, doi:10.2196/11936.
8. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. Nature 2017, 542, 115–118, doi:10.1038/nature21056.
9. Pham, T.-C.; Luong, C.-M.; Visani, M.; Hoang, V.-D. Deep CNN and Data Augmentation for Skin Lesion Classification. In Proceedings of the Intelligent Information and Database Systems; Nguyen, N.T., Hoang, D.H., Hong, T.-P., Pham, H., Trawiński, B., Eds.; Springer International Publishing: Cham, 2018; pp. 573–582.
10. The ISIC 2020 Challenge Dataset Available online: https://challenge2020.isic-archive.com/ (accessed on 23 August 2022).
11. Banerjee, S.; Singh, S.K.; Chakraborty, A.; Das, A.; Bag, R. Melanoma Diagnosis Using Deep Learning and Fuzzy Logic. Diagnostics 2020, 10, 577, doi:10.3390/diagnostics10080577.
12. Winkelmann, R.R.; Farberg, A.S.; Tucker, N.; White, R.; Rigel, D.S. Enhancement of International Dermatologists' Pigmented Skin Lesion Biopsy Decisions Following Dermoscopy with Subsequent Integration of Multispectral Digital Skin Lesion Analysis. J Clin Aesthet Dermatol 2016, 9, 53–55.
13. Albahar, M.A. Skin Lesion Classification Using Convolutional Neural Network With Novel Regularizer. IEEE Access 2019, 7, 38306–38313, doi:10.1109/ACCESS.2019.2906241.
14. Mahbod, A.; Schaefer, G.; Wang, C.; Ecker, R.; Ellinge, I. Skin Lesion Classification Using Hybrid Deep Neural Networks. In Proceedings of the ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); May 2019; pp. 1229–1233.
15. Architectures on Melanoma Detection. Tissue and Cell 2019, 58, 76–83, doi:10.1016/j.tice.2019.04.009.
16. Ali, M.S.; Miah, M.S.; Haque, J.; Rahman, M.M.; Islam, M.K. An Enhanced Technique of Skin Cancer Classification Using Deep Convolutional Neural Network with Transfer Learning Models. Machine Learning with Applications 2021, 5, 100036, doi:10.1016/j.mlwa.2021.100036.
17. Dorj, U.-O.; Lee, K.-K.; Choi, J.-Y.; Lee, M. The Skin Cancer Classification Using Deep Convolutional Neural Network. Multimed Tools Appl 2018, 77, 9909–9924, doi:10.1007/s11042-018-5714-1.