

Original Paper

Skin Lesion Classification With Deep Convolutional Neural Network: Process Development and Validation

Arnab Ray, BTech; Aman Gupta, BTech; Amutha AI, MEng

SRM Institute of Science and Technology, Chennai, India

Corresponding Author:

Arnab Ray, BTech

SRM Institute of Science and Technology

SRM Nagar, Kattankulathur

Chennai, 603203

India

Phone: 91 8939336693

Email: ad733943@gmail.com

Abstract

Background: Skin cancer is the most common cancer and is often ignored by people at an early stage. There are 5.4 million new cases of skin cancer worldwide every year. Deaths due to skin cancer could be prevented by early detection of the mole.

Objective: We propose a skin lesion classification system that has the ability to detect such moles at an early stage and is able to easily differentiate between a cancerous and noncancerous mole. Using this system, we would be able to save time and resources for both patients and practitioners.

Methods: We created a deep convolutional neural network using an Inceptionv3 and DenseNet-201 pretrained model.

Results: We found that using the concepts of fine-tuning and the ensemble learning model yielded superior results. Furthermore, fine-tuning the whole model helped models converge faster compared to fine-tuning only the top layers, giving better accuracy overall.

Conclusions: Based on our research, we conclude that deep learning algorithms are highly suitable for classifying skin cancer images.

(*JMIR Dermatol* 2020;3(1):e18438) doi: [10.2196/18438](https://doi.org/10.2196/18438)

KEYWORDS

deep convolutional neural network; VGG16, Inceptionv3; Inception ResNet V2; DenseNet; skin cancer; cancer; neural network; machine learning; melanoma

Introduction

Skin Cancer

One in every three cancers diagnosed is skin cancer. Although melanomas represent fewer than 5% of all skin cancers, they account for approximately 75% of all skin cancer-related deaths and are responsible for over 10,000 deaths annually. Early detection of the mole would decrease the number of skin cancer deaths.

Skin cancer is significantly lower in India due to the presence of eumelanin in India's dark-skinned population, which provides some protection against the development of skin cancer. Still, skin cancer constituted 3.18% of all patients with cancer in India. Of this, 54.76% were basal cell carcinomas, while 36.91% were squamous cell carcinoma and malignant melanoma was

only 8.33%. The majority of patients were from rural areas (88%) and many were involved in agriculture (92%) [1].

Neural Networks in the Context of Skin Cancer

We searched for research papers that used neural networks in the context of skin cancer from Google Scholar, PubMed, Research Gate, and the ISIC (International Skin Imaging Collaboration) archive. We included the results in the literature survey. Deep learning has solved many complex modern problems. The increasing amount of data on the internet helps in this process. There is a huge improvement in image classification using convolutional neural networks (CNN). The first few layers of deep CNN (DCNN) can learn the general features of an image, which can be used for different models. Using fine-tuning, DCNN models trained on one data set can be reused for image classification of other data sets. By fine-tuning Inceptionv3, Esteva et al [2] proposed that, "CNN

achieves performance on par with all tested experts, demonstrating an artificial intelligence capable of classifying skin cancer with a level of competence comparable to dermatologists". Esteva and colleagues used their own obtained dermatologist-labelled data set consisting of 129,450 clinical images, including 3374 dermoscopy images. This data set includes 2032 skin diseases, belonging to 9 skin disease partitions. By fine-tuning Inceptionv3 on this data set, Esteva and colleagues achieved up to 66% accuracy classification on these 9 classes.

Another previously published study that used DCNN used AlexNet [3]. The data set consisted of 200 pictures. However, by image augmentation (ie, rotating all the pictures), 4400 images were made. This study used the transfer learning model, in which the AlexNet model was trained on ImageNet data, and the last layer was replaced with the softmax layer that is classified into melanoma, seborrheic keratosis, and nevus. For the change of weights, they used the stochastic gradient descent (SGD) algorithmic program. They were able to achieve an accuracy of 98%.

In another study, the authors planned a mechanized strategy for malignant melanoma determination connected to an arrangement of dermoscopy photos [4]. Highlights removed relied upon using a multilayer perceptron (MLP) classifier and coevent network to distinguish between melanocytic nevi and melanoma. The authors proposed two different procedures for MLP: programmed MLP and conventional MLP. Both techniques were useful for the separation of melanocytic carcinoma with a high accuracy. Following this, the arrangement procedure was executed with an MLP classifier that involved two strategies: automatic MLP and traditional MLP. The MLP classifier displayed distinctive grouping accuracy. The programmed MLP planned 93.4% and 76% training and testing accuracy, respectively.

A different study used a model that uses support vector machine (SVM) learning algorithms [5]. Their model did not use annotated information. The feature transfer that they used allowed the system to draw similarities between observations of dermoscopic pictures and that of the natural world. It mimics the method specialists use to explain patterns in skin lesions. Two-fold cross-validation was performed 20 times for analysis (40 experiments in total), and two discrimination tasks were

examined: malignant melanoma versus atypical lesions, and malignant melanoma versus all nonmelanoma lesions. This approach achieved an accuracy of 93.1% for the primary task and 73.9% accuracy for the second task.

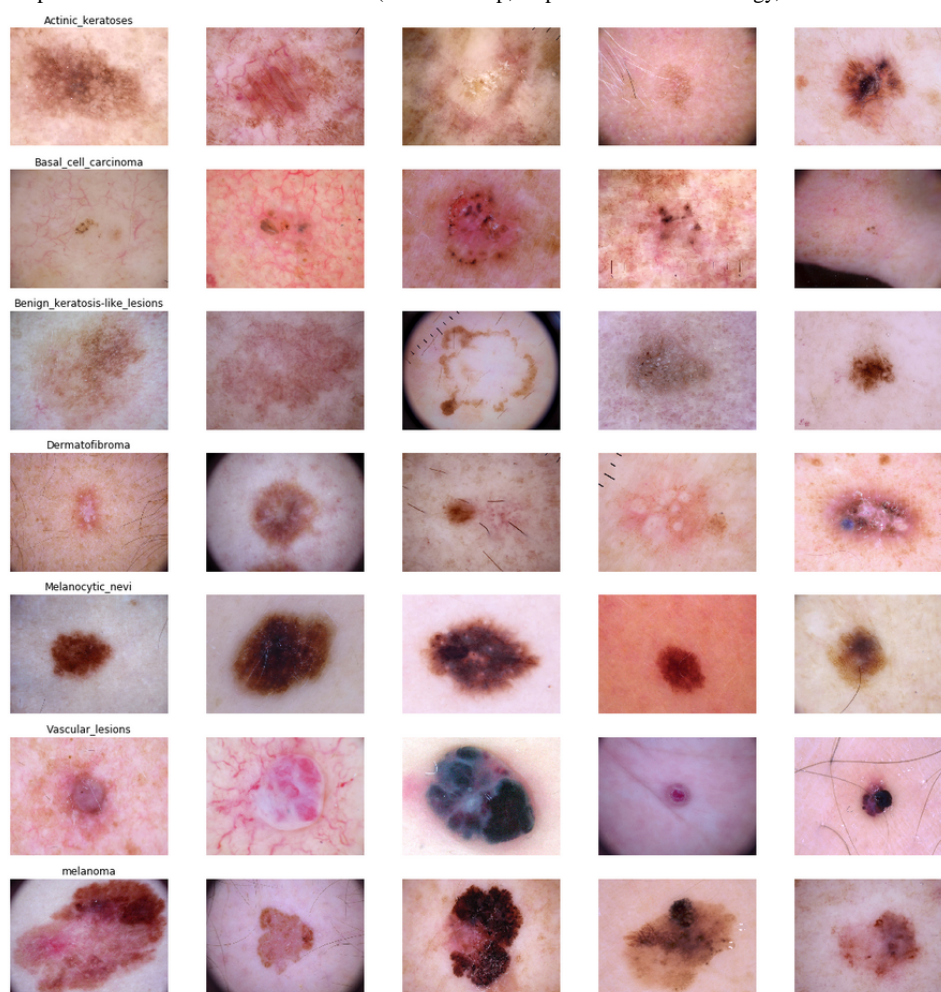
In another study, authors designed and modelled a system that can collect and combine past pigmented skin lesion (PSL) image results, their analysis, and corresponding observations and conclusions by medical experts, using a prototyping methodology [6]. One area of the system used computational intelligence techniques to research, process, and classify the images and their probable morphology. Trained medical personnel in remote locations can use mobile knowledge acquisition devices to take pictures of PSL and input the pictures into the planned system, which would classify the imaged PSL as malignant or benign.

Another group used a similar concept using DCNN. They trained their model on a data set of 129,450 images. They used the Inceptionv3 architecture model and classified images among 757 different melanoma classes. The accuracy achieved was 72%; this value was relatively low due to the high number of classes in this data set [2].

Another study used lesion segmentation as the first step of processing [7]. They identified morphological features specific to certain lesions. Preprocessing steps included changing the color channel, smoothing the image, removing hairs, etc. They modelled the algorithm as a binary classification model (ie, benign or malignant). Lesion-related morphological features (including diameter, color, and magnification) were used as the input to a number of classifiers. The best accuracy (79%) was found with the k-nearest neighbors (KNN) algorithm.

In this project, we used the HAM10000 data set obtained by ViDIR Group, Department of Dermatology, Medical University of Vienna. [Figure 1](#) shows example images from the data set that was used for this study.

In this study, we fine-tuned DCNNs and compared the performance of 4 DCNNs: VGG16, Inception-ResNet V2, Inceptionv3, and DenseNet-201. Each DCNN was fine-tuned from the top layers. Fine-tuning of all layers was performed with Inceptionv3 and DenseNet-201. Finally, we created an ensemble of Inceptionv3 and DenseNet-201 with all layers fine-tuned.

Figure 1. Example lesion photos from the HAM10000 data set (ViDIR Group, Department of Dermatology, Medical University of Vienna).

Methods

Exploratory Data Analysis

This step was performed to better understand the data and prepare the data for neural networks. In this project, we used the HAM10000 data set obtained by ViDIR Group, Department of Dermatology, Medical University of Vienna. The diagnostic accuracy for melanoma was significantly higher with dermoscopy compared to unaided eye diagnosis (respectively, log OR 4.0 [95% CI 3.0-5.1] versus log OR 2.7 [95% CI 1.9-3.4], an improvement of 49%, $P < .001$) [8]. The diagnostic

accuracy solely depended on the experience and knowledge of the examiner.

We observed that this data set is biased toward melanocytic nevi, as seen in Table 1. Hence, in the worst-case scenario, our neural network model will have an accuracy higher than 60%.

All the original images (450×600 pixels) were resized to 64×4-pixel RGB images for the baseline model and 192×256 pixels for fine-tuning models. The data set was split into 7210 training examples, 1803 validation examples, and 1002 test examples.

Table 1. Counts for each type of lesion in the data set

| Type of lesion | Number of images |
|----------------------|------------------|
| Melanocytic nevi | 6705 |
| Melanoma | 1113 |
| Benign keratosis | 1099 |
| Basal cell carcinoma | 514 |
| Actinic keratosis | 325 |
| Vascular lesions | 142 |
| Dermatofibroma | 115 |

Baseline Model

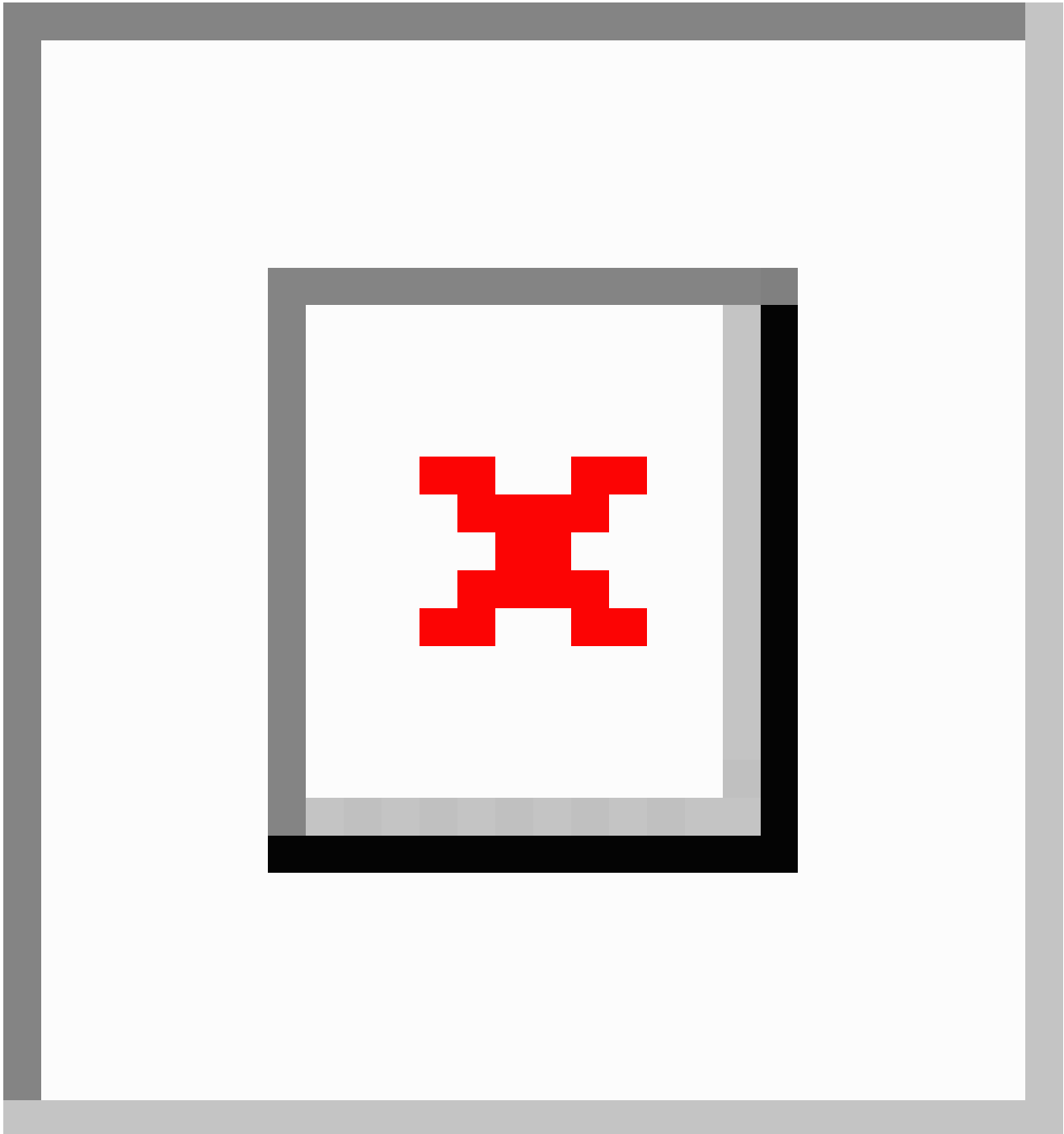
We built a baseline CNN to estimate the difficulty of the problem. Our architecture consisted of 6 layers: (1) a convolutional layer with 16 kernels each of size 3 and padding such that the size of the image is maintained, (2) a max-pooling layer with 2×2 window, (3) a convolutional layer with 32 kernels each of size 3 and padding to maintain size, (4) a max-pooling layer with 2×2 window, (5) a convolutional layer with 64 kernels each of size 3 and padding to maintain size, and (6) a max-pooling layer with 2×2 window.

To train the model, data augmentation was required. The learning rate was initialized at 0.01 and Adam Optimizer was used. The baseline model was trained for a total of 35 epochs.

VGG16 Model

VGG16 is a convolutional neural net architecture (Figure 2 [9]) that won the ImageNet competition in 2014 and is generally regarded as one of the best current vision models architecture. Even though it is an old model, we chose VGG16 because of its simplicity.

Figure 2. VGG16 architecture.



On the ImageNet data set, VGG16 achieved an accuracy of 90.1% for top-5 and 71.3% for top-1.

Data augmentation was performed to increase the data set image count. Fine-tuning was performed on the model by removing the top, fully-connected layers that were then replaced with

following: (1) a max-pooling layer, (2) a fully connected layer with 512 units, (3) a dropout layer with 0.5 rate, and (4) a softmax activation layer for 7 types of skin lesions.

The first step included freezing all layers in VGG16 and performing feature extraction for newly added layers. After 3 epochs, we unfroze the final convolutional block of VGG16 and started fine-tuning a model for 20 epochs. The learning rate was set to 0.001 and Adam Optimizer was used. VGG16 was fine-tuned for a total of 30 epochs.

Inception Model

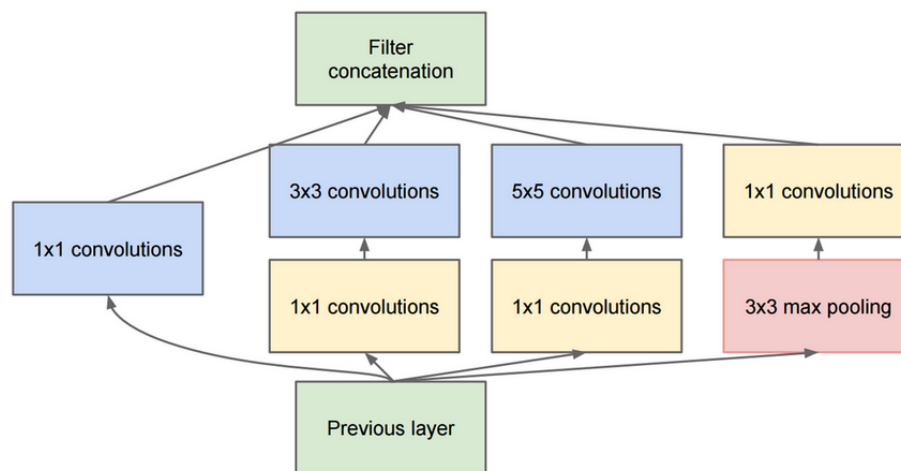
Inceptionv3 produced an accuracy of 93.7% for top-5 and 77.9% for top-1 on the ImageNet data set. The Inception module has

1×1 , 3×3 , and 5×5 convolutions, all in parallel (Figure 3 [10]). The intention was to let the network decide, through training, what information would be learned and used. It also allows for multi-scale processing; the model can recover low-level features via small convolutional layers and high-level features with large convolutional layers.

We fine-tuned all layers of Inceptionv3 and the top two inception blocks with batch normalization layers. Inceptionv3 was fine-tuned for 20 epochs.

Additionally, we tried Inception-ResNet, a variant of Inception. It uses a residual connection, which has become necessary for training very deep convolutional models. The same training strategy used for Inceptionv3 was used for Inception-ResNet.

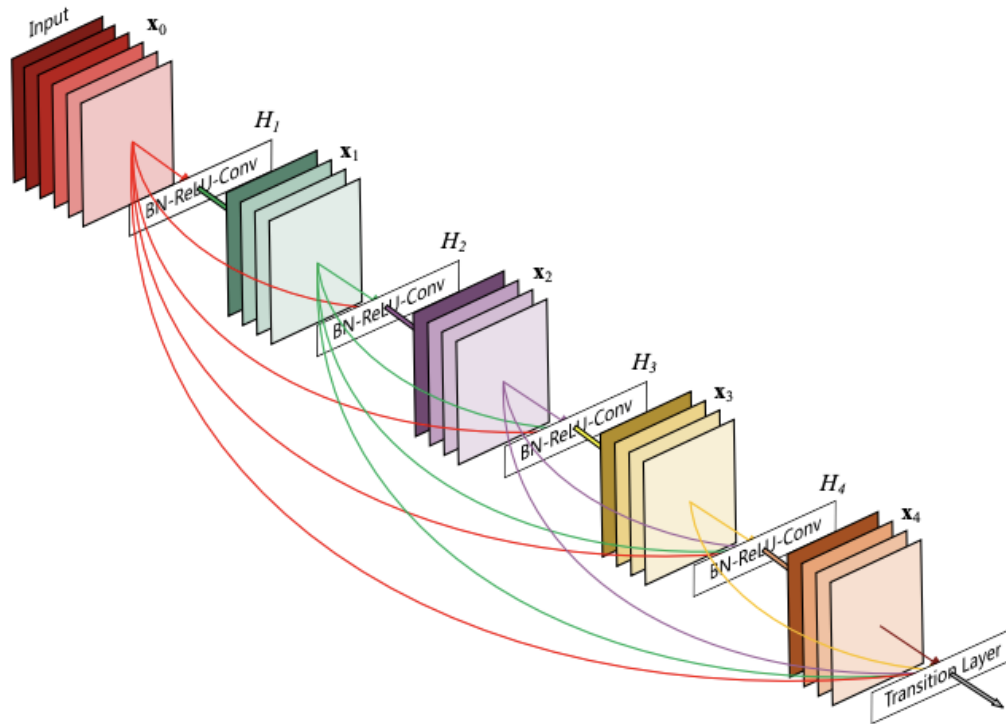
Figure 3. Inceptionv3 architecture. Published with permission.



DenseNet Model

This is a new architecture that performed exceptionally well in the ImageNet data set competition, giving an accuracy of 93.6% in top-5 and 77.3% on top-1. DenseNet has 4 dense blocks and uses approximately 20 million parameters (Figure 4 [11]).

In a dense block, one layer generates feature maps through a composite function, consisting of three consecutive operations: batch normalization, ReLU (rectified linear activation unit), and a 3×3 convolution. We used DenseNet-201, which uses 4 dense blocks, and we performed two types of fine-tuning on it: (1) fine-tuning on the last dense block (32 layers; Part A), and (2) fine-tuning on the whole network (Part B). Part A was trained for 27 epochs and Part B was trained for 20 epochs.

Figure 4. DenseNet architecture. Published with permission.

Results

Table 2 shows the classification results from each model when the top layers were fine-tuned (Part A). Table 3 displays the classification results for each model when all layers were fine-tuned. All experiments were performed on a laptop with GPU NVIDIA 1050Ti. To speed up processing times, Google Colab (P100 GPU) was used.

From training a custom model, it was clear that the problem cannot be solved by a simple CNN model with a few layers.

Therefore, we incorporated fine-tuning of the pretrained model. By hypertuning the pretrained model that had over 100 layers, we achieved better results. Fine-tuning all layers (Part B) gave us better results than fine-tuning only the top layers (Part A). Crucially, Part B was trained for fewer epochs, which helped the model converge faster. However, in both cases, DenseNet gave us better results than Inceptionv3. Using the concepts of ensemble learning, we created an ensemble of Inceptionv3 and DenseNet-201. This combination achieved a further improved accuracy of 88.8% on the validation set and 88.5% on the test set.

Table 2. Fine-tuning the top layers.

| Model | Validation (%) | Test (%) | Test loss | Depth (layers) |
|---------------------|----------------|----------|-----------|----------------|
| Custom model | 77.48 | 76.54 | 0.646671 | 11 |
| VGG16 | 79.82 | 79.64 | 0.708 | 23 |
| Inceptionv3 | 79.935 | 79.94 | 0.7482 | 315 |
| Inception-ResNet V2 | 80.82 | 82.53 | 0.6691 | 784 |
| DenseNet-201 | 85.8 | 83.9 | 0.691 | 711 |

Table 3. Fine-tuning all layers.

| Model | Validation (%) | Test (%) | Test loss |
|---|----------------|----------|-----------|
| Inceptionv3 | 86.92 | 86.826 | 0.6241 |
| DenseNet-201 | 86.696 | 87.725 | 0.5587 |
| Ensemble (Inceptionv3 and DenseNet-201) | 88.8 | 88.52 | 0.41156 |

Discussion

Our results indicate that deep learning algorithms are highly suitable for classifying skin cancer images. Additionally, by

using the concepts of fine-tuning and the ensemble learning model, improved results were achieved. Finally, we found that fine-tuning the whole model helped the model converge faster

compared with fine-tuning only the top layers, giving an overall better accuracy.

Conflicts of Interest

None declared.

References

1. Lal ST, Banipal RP, Bhatti DJ, Yadav HP. Changing Trends of Skin Cancer: A Tertiary Care Hospital Study in Malwa Region of Punjab. *J Clin Diagn Res* 2016 Jun;10(6):PC12-PC15 [FREE Full text] [doi: [10.7860/JCDR/2016/18487.8051](https://doi.org/10.7860/JCDR/2016/18487.8051)] [Medline: [27504344](https://pubmed.ncbi.nlm.nih.gov/27504344/)]
2. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017 Jan 25;542(7639):115-118. [doi: [10.1038/nature21056](https://doi.org/10.1038/nature21056)]
3. Hosny KM, Kassem MA, Foaud MM. Skin Cancer Classification using Deep Learning and Transfer Learning. 2018 Presented at: 9th Cairo International Biomedical Engineering Conference (CIBEC); December 20-22; Cairo, Egypt p. 90-93. [doi: [10.1109/CIBEC.2018.8641762](https://doi.org/10.1109/CIBEC.2018.8641762)]
4. S.Mabrouk M, A.Sheha M, Sharawy A. Automatic Detection of Melanoma Skin Cancer using Texture Analysis. *IJCA* 2012 Mar 31;42(20):22-26. [doi: [10.5120/5817-8129](https://doi.org/10.5120/5817-8129)]
5. Codella N, Cai J, Abedini M, Garnavi R, Halpern A, Smith JR. Deep Learning, Sparse Coding, and SVM for Melanoma Recognition in Dermoscopy Images. In: *Machine Learning In Medical Imaging*. Cham: Springer; 2015:118-126.
6. Okuboyejo DA, Olugbara OO, Odunaike SA. Automating Skin Disease Diagnosis Using Image Classification. 2013 Presented at: Proceedings of the World Congress on Engineering and Computer Science; October 23-25; San Francisco.
7. Codella N, Rotemberg V, Tschandl P, Celebi ME, Dusza S, Gutman D, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the International Skin Imaging Collaboration (ISIC). 2018 Presented at: IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018); April 4; Washington, DC, USA.
8. Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. *Lancet Oncol* 2002 Mar;3(3):159-165. [doi: [10.1016/s1470-2045\(02\)00679-4](https://doi.org/10.1016/s1470-2045(02)00679-4)] [Medline: [11902502](https://pubmed.ncbi.nlm.nih.gov/11902502/)]
9. VGG16 architecture. 2020 Jun 02. Classez et segmentez des données visuelles URL: <https://tinyurl.com/ya49hbpbk>
10. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2015 Jun [FREE Full text]
11. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. arXiv 2016 Aug 25 [FREE Full text]

Abbreviations

CNN: convolutional neural network
DCNN: deep convolutional neural network
ISIC: International Skin Imaging Collaboration
KNN: k-nearest neighbor
MLP: multilayer perceptron
PSL: pigmented skin lesion

Edited by G Eysenbach; submitted 26.02.20; peer-reviewed by A Chitranshi, K Ray; comments to author 03.03.20; revised version received 03.03.20; accepted 21.03.20; published 07.05.20

Please cite as:

Ray A, Gupta A, Al A

Skin Lesion Classification With Deep Convolutional Neural Network: Process Development and Validation

JMIR Dermatol 2020;3(1):e18438

URL: <http://derma.jmir.org/2020/1/e18438/>

doi: [10.2196/18438](https://doi.org/10.2196/18438)

PMID:

©Arnab Ray, Aman Gupta, Amutha Al. Originally published in JMIR Dermatology (<http://derma.jmir.org>), 07.05.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Dermatology Research, is properly cited. The complete bibliographic

information, a link to the original publication on <http://derma.jmir.org>, as well as this copyright and license information must be included.