ISSN: 2577 - 8005

Research Article

Medical & Clinical Research

Detection of Melanoma Skin Cancer with Deep Neural Networks

Grism Patel, Pooja Ranawade and Yuling Yan*

Department of Bioengineering, Santa Clara University, Santa Clara, United States

***Corresponding author**
Dr. Yuling Yan, Department of Bioengineering, Santa Clara University, Santa Clara, United States. E-mail: yyan1@scu.edu

Submitted: 13 Mar 2019; **Accepted**: 21 Mar 2019; **Published**: 05 Apr 2019

Citation: *Grism Patel, Pooja Ranawade and Yuling Yan. (2019). Detection of Melanoma Skin Cancer with Deep Neural Networks. Med.Clin.Res, 4(4), 1-4.*

Abstract

Detection of skin cancer involves several steps of examinations first being visual diagnosis that is followed by dermoscopic analysis, a biopsy, and histopathological examination. The classification of skin lesions in the first step is critical and challenging as classes vary by minute appearance in skin lesions. Deep convolutional neural networks (CNNs) have great potential in multicategory image-based classification by considering coarse-to-fine image features. This study aims to demonstrate how to classify skin lesions, in particular, melanoma, using CNN trained on data sets with disease labels. We developed and trained our own CNN model using a subset of the images from International Skin Imaging Collaboration (ISIC) Dermoscopic Archive. To test the performance of the proposed model, we used a different subset of images from the same archive as the test set. Our model is trained to classify images into two categories: malignant melanoma and nevus and is shown to achieve excellent classification results with high test accuracy (91.16%) and high performance as measured by various metrics. Our study demonstrated the potential of using deep neural networks to assist early detection of melanoma and thereby improve the patient survival rate from this aggressive skin cancer.

Keywords: Melanoma, Convolutional Neural Networks, Detection of Skin Cancer

Introduction

With automated image diagnosis in health care industry it is predicted that the image analysis could create $3 Billion annual savings by giving specialists more time to focus on reviews that require greater attention and judgement [1]. Skin cancer remains the most prevalent of all cancers in United States, more common than all other cancer diagnoses combined.

The signs and symptoms of skin cancer maybe individual specific, which makes its detection more challenging. Although melanoma accounts for only two percent of all skin cancer cases, it is more likely than other types of skin cancers (basal and squamous cell cancers) to spread to the lymph nodes or metastasize to other parts of the body. Melanoma has doubled in incidence in recent decades and is increasing more rapidly than any other cancer [2]. An estimated 60,000 cases and over 8000 deaths annually associated with melanoma were reported in the U.S. with an average individual lifetime risk of melanoma approaching 1 in 75 [3].

Melanoma is the most aggressive and deadly of all skin cancer types, however, early detection of melanoma is critical, as the estimated 5-year survival rate for melanoma patients drops from over 99% if detected in its earliest stage to about 14% if detected in its latest stage [4]. The earlier detection of melanoma can also significantly decrease the treatment costs. Our research was inspired by "***Dermatologist - level classification of skin cancer with deep neural networks***" which makes use of CNN in form of Inception v3 to detect a variety of skin cancers [4].

However, we aim to scale down this approach to focus on the detection of melanoma characteristics. Most often, melanoma has no symptoms, and it may be difficult to differentiate atypical nevus from melanoma. We believe that an AI-assisted mobile app can be valuable in assisting at-home self-examinations whenever any new nevi or noticeable changes in shape or color are observed in existing nevi.

Most researchers made use of the available transfer learning using pretrained architectures including Inception v3 for skin cancer detection and resnet50, resnet100, vgg16, Alexnet, xceptioin or Mobilenet to solve other image-based classification problems. In particular, Andre Esteva et al. trained Inception v3 for the multi-class classification of skin cancer and achieved dermatologist level classification results [4]. The reported accuracy using this model for the three class division of image set is 72%, while the accuracy for nine-class division is 55.3%. Most recently, Haenssle et al. trained Google's Inception v4 CNN architecture and used it to

detect in situ and invasive melanomas [5]. They have reported a specificity superior to dermatologist-level diagnosis (82.5% for CNN versus 75.7% for dermatologists) and an improved mean sensitivity to 88.9%. Lequan Yu et al. proposed a comprehensive two-stage approach based on very deep CNNs and a very deep (50 layers) fully-connected convolutional residual network (FCRN-50) for accurate skin lesion segmentation [6]. They presented similar classification results in comparison with VGG-16 and Google Net as measured by accuracy (0.855), average precision (0.624), specificity (0.931) and sensitivity (0.547). Nasr-Esfahani et al. proposed and trained an CNN model that achieved superior performance in comparison with several state-of-the-art methods presented in as measured by performance metrics including accuracy (0.81), sensitivity (0.82) and specificity (0.75) [7-10].

In this study, we have performed pre-processing of the images to focus on the lesion region of the image. Since our goal is to classify image characteristics into two categories, we decided to develop a new CNN architecture optimized for our purpose and at a significantly less computational cost. We also aim to train our own CNN model to further improve classification result with an ultimate goal of implementing a mobile app to assist early detection of melanoma for use at home or in the clinical setting.

Methods

Figure 1: Flowchart of the approach. Dividing of the dataset is done with the help of metadata provided by the ISIC achieve. Image processing is done to reduce the image size and remove extra skin pixels. After that CNN is applied to classify the cancer images.

We propose a two-category classification model, which represents two individual disease classes:
1). Melanoma, and 2). Nevus (birthmark or mole). The flowchart of the procedure is illustrated in figure 1.

The dataset used in our study are obtained from open-access repository, ISIC dermoscopic archive. These images come with metadata annotated as malignant (melanoma), carcinoma, and nevus (skin mole) respectively. The subset dataset contains 4050 images that were then divided into melanoma: 2025 and nevus: 2025, with the help of metadata. Subsequently, we proceeded with the image pre-processing.

Image Pre-processing

The pre-processing of these images provides two advantages: first, it improves computational efficiency (the pixel size for each original image is about 6000 by 4000), and second, it can automatically detect and zoom into the region of interest. The pre-processing involves removing periphery skin pixels and cropping the image, which is performed using Open CV. We multiply the original image (figure 2, middle) with its mask image (figure 2, left), which is followed by making a boundary around the skin mark and cropping the rest of the image to generate new input images each with reduced dimensions of 299*299*3 pixels (figure 2, right). With the preprocessing, we have achieved a shorter average running time of 4575 seconds (for each epoch) for all 14 layers of the network with a mini-batch size of 32.

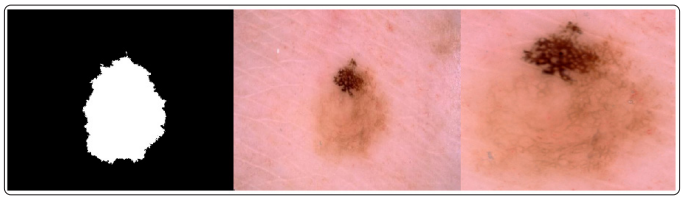

Figure 2 Left: Mask Image of same size as original Image; Centre: original image of size 6682*4442; Right: pre-processed image (2486*2771) after removal of the non-lesion part of the skin.

Proposed CNN Architecture

Rationale for a custom-built architecture

Many researchers tend to take advantage of the available transfer learning by using pretrained architectures such as Inceptionv3, Resnet50, Resnet100, Vgg16, Alexnet, Xceptioin or Mobilenet to solve image classification problems. Unfortunately, we have found these architectures inappropriate for our application after vigorously testing them on our datasets. First, the scarcity of data in our case does not play in favor for them. And second, these models are built to best perform for multi-class rather than binary classification (two classes) problems. Hence, we chose to build our own architecture to address these issues and deliver optimal results.

Training of the CNN model

Phase 1

Using 2032 training images and 200 validation images, we showed the initial model (Figure 3) was over fitting. To address this issue, we introduced a new layer of Dropout with a probability of 0.5. This modification reduced overfitting but not by a great extent. Considering this result is most likely due to the small number of training images, in the next phase, we tried to change the division of data.

Phase 2

After changing the ratio of training images to validation images to 90:5 and reserving the remaining 5% for testing, we had 3838 training images and 212 validation images. In this phase we also introduced batch normalization and removed dropout. Batch normalization reduces the amount by which the hidden units will change values i.e. covariance shift, and allows each layer to learn independently of other layers. We then added one more fully connected with batch normalization layer. Now, we are using 14 Layers with three convolution layers, three max pool layers, one flatten, three fully Connected layers with batch normalization layer and one output layer, as shown in figure 4.

Description of the architecture in figure 4:

1. For feature extraction, we used three layers of convolution and max pooling with 32 kernels each of size 3*3 and input image of size 299*299*3. The activation function was ReLu (rectified linear unit).
2. Then a flatten Layer was added to convert two-dimensional image matrix to one dimensional vector.
3. This is followed by the classifier neural network of three fully connected layers and batch normalization for each layer to avoid overfitting.
4. The final output layer with sigmoid activation function is used to classify each image into one of the two classes.

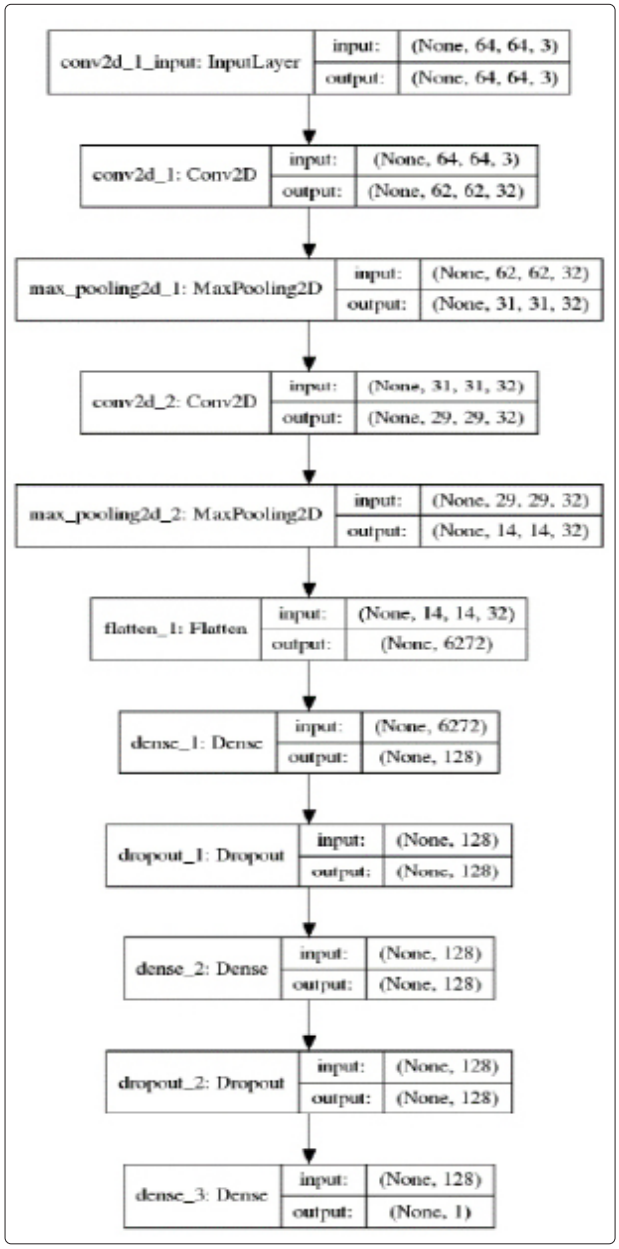

Figure 3: Phase 1 Convolutional Neural Network Architecture

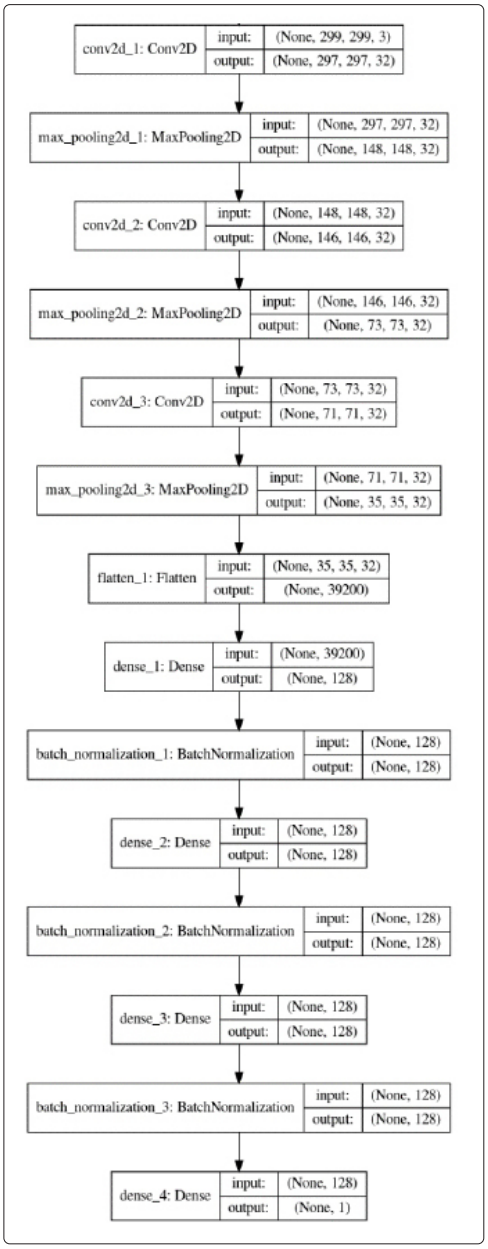

Figure 4: Phase 2 Convolutional Neural Network Architecture

Results

With the current model and architecture (figure 4), we achieved a training accuracy of 0.9633 with loss of 0.0965, a validation accuracy of 0.9116 and loss of 0.3499, and a test accuracy of 0.9198 (Table 1).

Table 1: Results for Training, Validation, Test

	Training Dataset (90%)	Validation Dataset (5%)	Test Dataset (5%)
Accuracy	0.9633	0.9116	0.9198

The confusion matrix and other performance metrics including precision, recall, and F1 score are also calculated and shown in Tables 2&3. Two representative image samples representing the true positive and true negative prediction results are shown in figures 5 &6.

Table 2: Confusion Matrix to Show the No. of Results in Each Category

Classification Result	Predicted Melanoma	Predicted Nevi
Positive (True Melanoma)	101 (TP)	5 (FN)
Negative (True Nevus)	12 (FP)	94 (TN)

Table 3: Accuracy, Precision, Recall and F1 Score Performance Metrics for the Proposed CNN Model

Accuracy	Precision	Recall	F1 Score
0.9198	0.8938	0.9528	0.9224

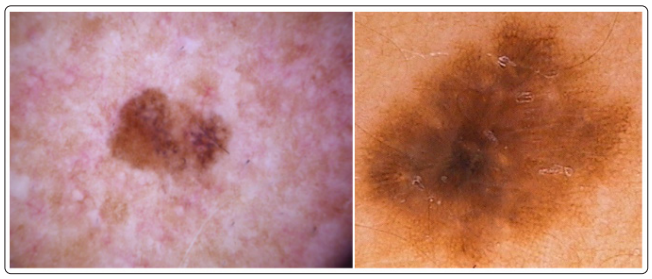

Figure 5: True Positive Prediction **Figure 6:** True Negative Prediction

The computational efficiency and the number of training parameters used in our model are summarized as follows:

1. Efficiency

Our model is computationally more efficient compared to other abovementioned approaches. The time cost for per epoch was 1.50 hours and the total training time was 15 hours for 10 epochs.

2. Number of parameters

The number of parameters used in our model is lower than that used in any other existing models. For example, AlexNet has 60 million parameters, VGGNet has three times more than that of AlexNet, and Inception v1 has 7 million parameters. The proposed architecture uses only 5,071,809 (around 5 million) parameters which makes it more efficient.

To test the performance of the classifier, we calculated the following common used metrics and the results are summarized in Table 4.

- TPF = True Positive Fraction (Sensitivity) = TP/ (TP+FN) = 101/ (101+5) = 95.3%
- FNF = False Negative Fraction (1-Sensitivity) = 4.7%
- TNF = True Negative Fraction (Specificity) = TN/ (TN+FP) = 94/ (94+12) = 88.7%
- FPF = False Positive Fraction (1-Specificity) = 11.3%
- PPV = Positive Predicted Value=TP/(TP+FP) = 101/ (101+12) = 89.4%
- NPV = Negative Predicted Value=TN/(TN+FN) = 94/ (94+5) = 94.9%

Table 4: Results of Test Performances

TPF (Sensitivity)	FNF	TNF (Specificity)	FPF	PPV	NPV
95.3%	4.7%	88.7%	11.3%	89.4%	94.9%

Conclusion

The proposed Neural Network consists of 14 layers with three convolution layers and three max pool layers trained on 3838 images and validated on 212 images. We used 32 kernels each of a size of 3*3 matrix. Using the proposed model and parameters described above, we have achieved a training and validation accuracy of 96.33% and 91.16% respectively, and with a training loss of 0.096 and validation loss of 0.34 after 10 epochs. The results demonstrated the potential of using our trained CNN model to detect malignant melanoma with an accuracy that is comparable to a human expert. With more pathological datasets and further improvements our algorithm may be used to assist early detection and staging of skin cancer, especially melanoma, for mobile home use or use in the clinical setting.

References

1. Brian Kalis, Matt Collier, Richard Fu (2018) "10 Promising AI Applications in Health Care", https://hbr.org/2018/05/10-promising-ai-applications-in-health-care.
2. Dennis LK (1999) Analysis of the melanoma epidemic, both apparent and real: data from the 1973 through 1994 surveillance, epidemiology, and end results program registry. Arch Dermatol 135: 275-280.
3. Agnessa G Goodson, Douglas Grossman (2009) "Strategies for early melanoma detection: approaches to the patient with nevi". J Am Acad Dermatol 60: 719-738.
4. Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, et al. (2017) "Dermatologist - level classification of skin cancer with deep neural networks". Nature. 542: 115-118.
5. H A Haenssle, C Fink, R Schneiderbauer, F Toberer, T Buhl, et al. (2018) "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists." Annals of Oncology 29: 1836-1842.
6. Lequan Yu, Hao Chen, Qi Dou, Jing Qin, and Pheng-Ann Heng (2017) "Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks." IEEE TRANSACTIONS ON MEDICAL IMAGING. 36: 994-1004.
7. Nasr-Esfahani E, Samavi S, Karimi N, Soroushmehr SM, Jafari MH, et al. (2016) "Melanoma detection by analysis of clinical images using convolutional neural network." Conf. Proc. IEEE Eng. Med. Biol. Soc. 2016: 1373-1376.
8. I Giotis, N Molders, S Land, M Biehl, M F Jonkman, N Petkov (2015) "MED-NODE: A computerassisted melanoma diagnosis system using non-dermoscopic images," Expert Systems with Applications, Elsevier 42: 6578-6585.
9. C Munteanu, S Cooclea (2009) "Spotmole-melanoma control system," Available: http://www.spotmole.com/
10. E Zagrouba, W Barhoumi (2004) "A preliminary approach for the automated recognition of malignant melanoma". Image Analysis & Stereology. 23: 121-135.