MobileDerm

APS360 - Project Proposal

Matthew Lee, Rohan Ramdoss, John Xie

Word Count: 1383

1.0 Introduction

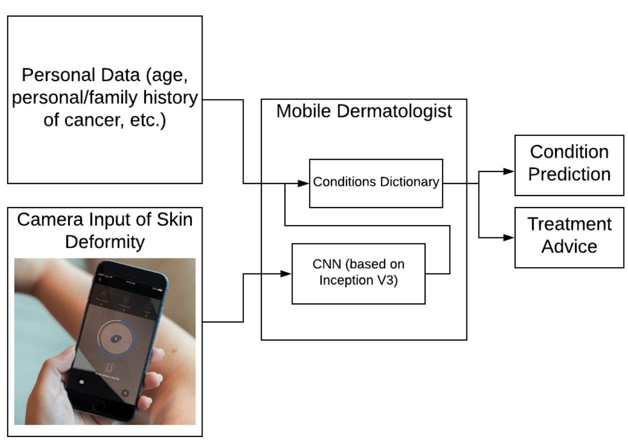
The idea we have decided to implement is called Mobile Dermatologist. We have decided to use CNN to more accurately diagnose different types of skin cancers, conditions, and deformities. Some of the signs and symptoms of skin cancer are: skin bumps, scaly red patches, wart like growth, pink skin lesions [1]. However these are similar to symptoms of different skin conditions and diseases which pose a challenge in correct diagnosis. By incorporating all types of skin conditions, we can reduce misdiagnosis and expand the scope of the project.

Our goal is to train a neural network model in order to be able to detect different forms of skin cancer and conditions from dermoscopic images. Skin cancer is primarily diagnosed visually, with an initial clinical screening followed potentially by dermoscopic analysis, a biopsy and histopathological examination. Automated Classification of skin lesions using images is a challenging task owing to the fine grained variability in the appearance of skin lesions. In the US, more than 9500 people in the US are diagnosed with skin cancer every day. More than two people die of the disease every hour [2].

There are a few existing projects that aim to help with diagnosis of skin cancer--however, we wish to extend the scope to anything skin related. Ideally we would like to have this as a mobile app where the user can scan over a skin deformity and have an initial diagnosis and treatment tips. We aim to provide a solution for people without immediate access to a dermatologist, due to distance or socioeconomic conditions.

Neural networks can be used to process complicated and a huge amount of information in a short time. Neural networks have the capability to surpass human reading capacity, meaning that it can see more precisely than the naked human eye. We believe that a properly trained Convolutional Neural Network can increase the efficiency and speed of detection and, in our case, put this capability to the hands of anyone with a smart device [3].

2.0 Illustration/Figure



As not all of our datasets include personal data, we will only use images to help predict the skin condition. However, we may choose to use user personal data to offer more relevant treatment advice. The "Conditions Dictionary" is just a mapping of a condition to possible recommended remedies. The condition prediction and treatment advice will appear on the app screen, if there is time available to implement this to mobile devices.

3.0 Background & Related Work:

3.1 Stnaford's CNN to diagnose all types of skin cancer:

https://cs.stanford.edu/people/esteva/nature/

3.2 SkinVision:

https://www.skinvision.com/

What we learned so far is Stanford developed CNN to majorly identify the most common skin cancers and the deadliest skin cancers. They used 129,450 clinical images include 2,032 different diseases, and trained the deep CNN (Inception v3) model and classified the skin lesion inputs.

SkinVision, on the other hand, is an commercial iOS/Android application to diagnose the skin diseases. The local application is mainly for inferencing by uploading user's skin photos to its online server and get skin health rating after 30 seconds.

4.0 Data Processing

1. Skin melanoma image dataset - https://www.isic-archive.com/#!/topWithHeader/onlyHeaderTop/gallery
2. HAM10000 image dataset of common pigmented skin lesions - https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T
3. Clinical Skin Disease Images - https://medicine.uiowa.edu/dermatology/education/clinical-skin-disease-images

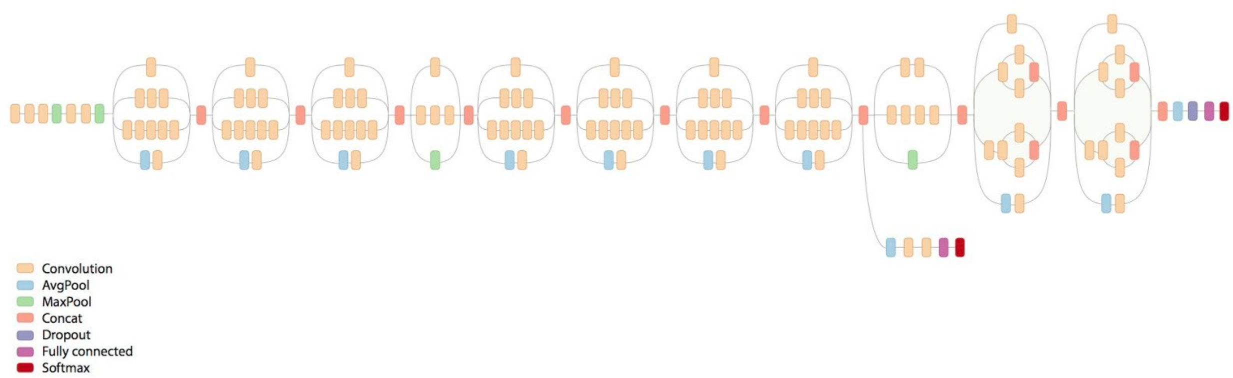
Since we are collecting data from many different sources, we will have to standardize the images in terms of image size, resolution, background, pigmentation, condition terminology, etc.. Moreover, different skin deformities happen in different parts of the body (finger, lip, nose) which can lead to varying contents in the foreground. Ideally we would like all images to have the skin deformity in the center with only the skin surrounding it as the background. For images that cannot be effectively edited to be of our ideal configuration and for images that belong to a condition we do not have much data on, we will have to omit those data points.

Moreover some datasets include conditions that are out of our scope (black hairy tongue, hair shaft, etc.). We will need to filter these conditions out using a predetermined set of skin conditions we wish to observe.

One major issue we run into with our datasets is that most of the images are of lighter skin subjects--due to the increased ease of distinguishing a darker skin deformity on a lighter skin background. We could try to augment the dataset by editing the images to emulate the deformity on darker skin, however, there still exist limitations of not using images of actual darker skinned subjects.

5.0 Architecture

As this project is based on taking information from pictures and classifying them, we will use a form of CNN. We will first try the deep CNN (Inception v3 or v4) model first according to the Stanford research paper [2].



Based on the benchmark of Inception v4, we can get 3180 ms/661 ms mean inference time on Pixel 2/iPhone 8 on TensorFlow framework [4]. Considering we should get similar performance on PyTorch, it is acceptable for our mobile implementation of this project.

6.0 Baseline Model

Unfortunately, we are not the dermatologists specializing in skin cancer, however, we do have accuracy comparison between dermatologists and trained neural network from Stanford research paper. We will use both dermatologists and CNN accuracies from that research paper to determine the performance of our CNN model on skin cancer prediction. As for an actual model we could implement, considering that our actual model will have hundreds of classifications, we are not sure of the feasibility of implementing a solely heuristic model without machine learning; it would be hard to make a complicated decision (over a hundred different possibilities) based on the fundamental and primitive attributes we could extract from an image without machine learning.

Extended Data Table 2 | General validation results

a.

Classifier	Three-way accuracy
Dermatologist 1	65.6%
Dermatologist 2	66.0%
CNN	69.4 ± 0.8%
CNN - PA	**72.1 ± 0.9%**

b.

Classifier	Nine-way accuracy
Dermatologist 1	53.3%
Dermatologist 2	55.0%
CNN	48.9 ± 1.9%
CNN - PA	**55.4 ± 1.7%**

c. **Disease classes: three-way classification**

0. Benign single lesions
1. Malignant single lesions
2. Non-neoplastic lesions

d. **Disease classes: nine-way classification**

0. Cutaneous lymphoma and lymphoid infiltrates
1. Benign dermal tumors, cysts, sinuses
2. Malignant dermal tumor
3. Benign epidermal tumors, hamartomas, milia, and growths
4. Malignant and premalignant epidermal tumors
5. Genodermatoses and supernumerary growths
6. Inflammatory conditions
7. Benign melanocytic lesions
8. Malignant Melanoma

Here we show ninefold cross-validation classification accuracy with 127,463 images organized in two different strategies. In each fold, a different ninth of the dataset is used for validation, and the rest is used for training. Reported values are the mean and standard deviation of the validation accuracy across all $n = 9$ folds. These images are labelled by dermatologists, not necessarily through biopsy; meaning that this metric is not as rigorous as one with biopsy-proven images. Thus we only compare to two dermatologists as a means to validate that the algorithm is learning relevant information. **a**, Three-way classification accuracy comparison between algorithms and dermatologists. The dermatologists are tested on 180 random images from the validation set—60 per class. The three classes used are first-level nodes of our taxonomy. A CNN trained directly on these three classes also achieves inferior performance to one trained with our partitioning algorithm (PA). **b**, Nine-way classification accuracy comparison between algorithms and dermatologists. The dermatologists are tested on 180 random images from the validation set—20 per class. The nine classes used are the second-level nodes of our taxonomy. A CNN trained directly on these nine classes achieves inferior performance to one trained with our partitioning algorithm. **c**, Disease classes used for the three-way classification represent highly general disease classes. **d**, Disease classes used for nine-way classification represent groups of diseases that have similar aetiologies.

7.0 Ethical Considerations

This project is designed to function similarly to a dermatologist, skin cancer practitioners, and doctors. Although our goal is not to replicate the professional expertise of the aforementioned professions, an over-reliance on such a tool could lead to less consultation with dermatologists and doctors. Not only is this potentially financially threatening to professionals, but also, it can lead to mistreatment of patients as our project is not aimed to be a trained practitioner.

On a different note, the majority of our images are of lighter skin--due to the increased ease of distinguishing a darker skin deformity on a lighter skin background. Although we could try to preprocess the data to emulate the skin deformity on a darker skin, the model would still more accurately classify skin deformities on lighter skinned subjects.

8.0 Project Plan

The table below shows the list of tasks accompanied, its internal deadlines, and the accompanying checkpoints from the course.

Task	Internal Deadline	Relevant Checkpoint
Clean up and Preprocess Data	July 7th	Progress Meeting with TA Mentor (July 11-15)
Implement Baseline Model	July 13th	Progress Report (July 24th)
Experiment with Different Parameters	July 19th	Progress Report(July 24th)
Observe findings	July 26th	-
Summarizing Report	August 10th	Project Report August 15th
Presentation Slides	TBD	Presentation - TBD

1-2 hour meetings will be conducted at 5:30 on Tuesdays, every week. The method of communication is facebook messenger. Code suggestions in comments can be used to ensure that code does not get overwritten. Communicating from the TA mentor will be done through email. One team member will correspond with the TA directly on behalf of the team. Collaboration will be through Github.

9.0 Risk Register

1. A team member decides to drop the course
    - Reduce scope and reduce the amount of data images needed to clean.
2. Model takes longer to train than expected
    - Reduce scope (classes) and focus on just a few skin conditions
3. Cannot correctly detect skin conditions on darker skinned subjects (dataset augmentation was not helpful enough)
    - Model is only as good as the dataset--not much can be done, will have to address in our project presentation and final implementation
4. Too many different classes--final model has low accuracy due to conditions looking alike
    - Reduce scope (classes) to just skin conditions related to cancer

10.0 Link to Github/Colab

https://github.com/RRamdo233/APS360-Group-Project

11.0 References

[1] G. W. Cole, "Skin Cancer Symptoms, Signs, Types, and Treatment - MedicineNet." Available: MedicineNet, https://www.medicinenet.com/skin_cancer_overview/article.htm [Accessed: June 30, 2019].

[2] Esteva, A., Kuprel, B., Novoa, R., Ko, J., Swetter, S., Blau, H. and Thrun, S. "Dermatologist-level classification of skin cancer with deep neural networks." *Nature*, vol. 542, no.7639, pp.115-118. Available: https://cs.stanford.edu/people/esteva/nature/. [Accessed June 30, 2019]

[3] Y. B. Perez "AI is already changing how cancer is diagnosed - TNW.", May 1, 2019. [Online], Available: https://thenextweb.com/artificial-intelligence/2019/05/01/ai-is-already-changing-how-cancer-is-diagnosed/. [Accessed June 30 2019].

[4] "Performance Benchmarks," Tensorflow, [Online]. Available: https://www.tensorflow.org/lite/performance/benchmarks. [Accessed June 30 2019]