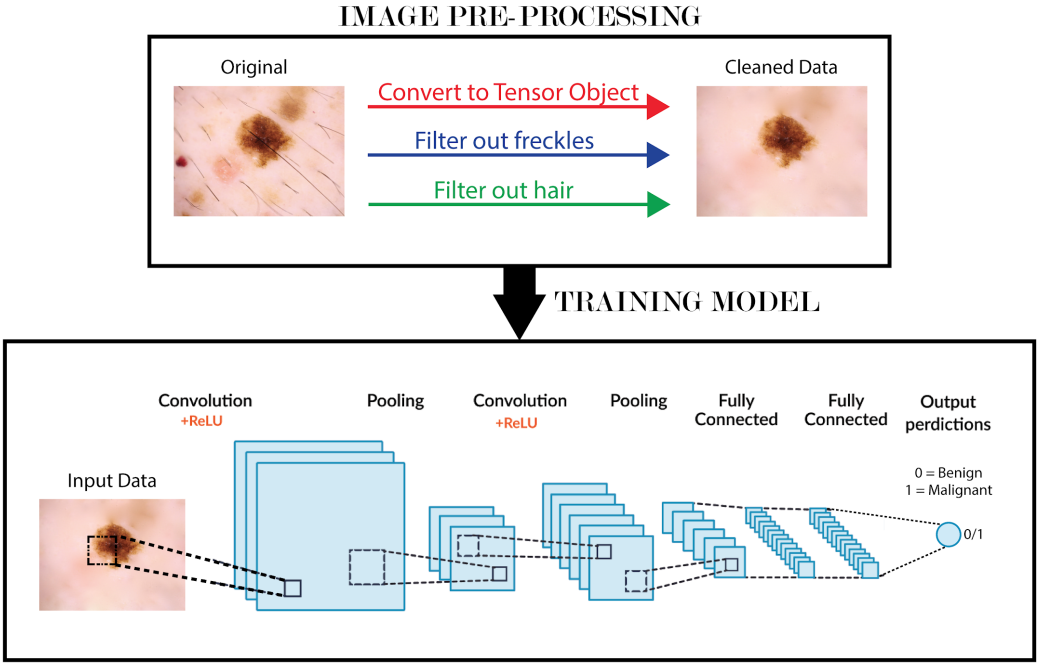
# Skin Cancer Classifier using Dermatoscopic Images of Pigmented Lesions

Meha Gupta
Samarth Sinha
Ines Bosch-Alfonso
Valentina Manferrari
Word Count: 1387
Penalty: 0

## Introduction

Skin cancer is the 19th most common form of cancer for humans [1], and the most prevalent in the United States where it is estimated that 1 in every 5 Americans will suffer from skin cancer at some point in their life [2]. As such, the prevention and early detection of skin cancer is crucial. Detecting skin cancer involves carefully analyzing one's moles, skin lesions and overall skin tone. Dermatologists are highly trained in this skill, but the process to diagnose a patient can be lengthy and costly due to the need for expert supervision from the doctor. Our project goal is to create a classifier that is able to classify an image of a skin lesion as being malignant or benign. Machine learning would be of appropriate use in this field as a classifier trained with images of skin cancer lesions would greatly aid in this early detection. This design would allow for those with non-medical knowledge to easily know the possibility of them having skin cancer by simply providing an image of the suspected lesion rather than having to go through the potentially costly process of being diagnosed by a dermatologist.

## Illustration




## Background & Related Work

Following the advancements made in computer vision and the introduction of publicly-available large dermoscopic datasets with different types of benign and malignant skin lesions, the use of dermoscopic images in skin cancer classification algorithms has become a popular field of research. In a landmark study, Esteva et al. [3] compared the performance of a CNN model (GoogleNet Inception-v3 Architecture), trained on a combined dataset of 129,450 clinical and dermatological images (representing 2032 different skin

lesions), with the performance of 21 board-certified dermatologists for skin cancer classification across three categories, and successfully demonstrated that the performance of the neural network was at par with dermatologist performance.

Another study by Brinker et al. [4] used enhanced deep learning methods to train a CNN model over 12,378 open-source dermoscopic images and used 100 images to compare the performance of the neural network with the performance of board-certified 157 dermatologists from 12 different universities in Germany. The CNN was found to outperform 136 of the 157 participating dermatologists.

Furthermore, Tschandl et al. [5] tested a combined CNN model trained on 7895 dermoscopic and 5829 close-up images of lesions on 2072 unknown cases and compared results with 95 medical personnel of which 62 were board-certified dermatologists. The research concluded that the performance of neural networks matched human expert performance under experimental conditions.

Despite these early successes, some of the aforementioned authors have identified some key challenges with training deep neural networks to reliably classify benign and malignant skin lesions. One important challenge in the research area is the lack of a large-scale public benchmark which includes data from different diversities of the world population [6].

## Data Processing

Our data source is the ISIC 2020 Challenge dataset which includes images and labels for examples of both malignant and benign skin lesions [7]. There are 33,126 dermoscopic images of over 2000 different patients. The images are all .jpg and are all sized 1024x1024 pixels. The labels on the data are as follows :

- image_name - unique identifier for the image
- patient_id - unique patient identifier
- sex - the sex of the patient
- age_approx - approximate patient age
- anatom_site_general_challenge - anatomic location of image
- diagnosis - detailed diagnosis information
- benign_malignant - indicator of malignancy
- target - categorical target variable: 0 represents benign, 1 represents malignant

It should be noted that different cases include multiple images of the same lesion to see progression over time. Furthermore, for our purposes, we shall only be using the target label for training therefore we need to clean out the other unnecessary labels. Finally, we are considering searching for a supplementary dataset with a greater diversity in skin tones as most in this dataset are of lighter skin.

## Architecture

We will be applying a filter to remove possible freckles and hair so that the features of the skin lesion are clear. We will be creating a CNN based on our Lab 2 code. For now, we do not know the specific architecture or specific hyperparameter values, but we plan to use deep CNNs.

## Baseline Model

We select the Support Vector Machine (SVM) model from scikit-learn [8] as the baseline to compare our neural network against. SVM provides a simple supervised machine learning algorithm that builds an optimal hyperplane in multi-dimensional space to separate different classes of data [9]. We believe this model to be a good baseline for our design because as per a recent study, among different classifiers (like Random Forests and kNNs), SVMs gave the best performance on skin lesion classification [10].

## Ethical Considerations

It is very important that the training dataset accurately represents the source population in order to avoid the generation of a biased model. Since our database provides us with images at a dermatoscopic level, the ethnicity of the patient could potentially affect the performance of our model. In addition, our training data is limited to a certain number of skin cancer types and this can be a limitation to the model when it encounters rare forms which are underrepresented in data or different forms of cancer from what is available in the training set since this can lead to misclassifications by the model. To successfully deploy our trained model in practice, the efficacy of the model needs to undergo further ethical considerations such as liabilities in case of a prediction error.

## Project Plan

Our team will closely work together using Zoom Video Calls when important team decisions are needed and for the crucial parts of the project (i.e. data cleaning, model training and testing). Tasks will be split evenly and carried out individually but final revision will always be done together before each submission. We will use a Messenger chat group to communicate individual progress and any question or doubt we have to the rest of the team. At the end of each meeting we will also always choose a date for the next meeting in order to keep internal deadlines clear and commitment high; ideally we will meet once a week after classes. To make sure individual work runs smoothly and no one overwrites each other's text and/or code we will use Google Docs and Google Colab so that we can easily add comments to different sections and keep track of the who did what and last edits.

| Team Member | Task | Internal Deadline | Completion |
|---|---|---|---|
| Valentina | Draft for Ethical Considerations and Risk | 02/09/2021 | Done |

| | | | |
|---|---|---|---|
| | Register | | |
| Ines | Draft for Introduction and Data processing | 02/09/2021 | Done |
| Meha | Draft for Background and Related Work | 02/09/2021 | Done |
| Samarth | Joined late due to partners dropping out. Minor draft editing. | 02/11/2021 | Done |
| Team | Final Revision of Project Proposal | 02/12/2021 | Done |
| Ines | Data Preprocessing | 02/26/2021 | - |
| Meha | Develop baseline SVM model | 03/05/2021 | - |
| Samarth | Develop Deep CNN | 03/12/2021 | - |
| Valentina | Train Deep CNN | 03/19/2021 | - |
| Ines | Test Deep CNN & Compare with Baseline | 03/26/2021 | - |
| Team | Complete Final Report and Presentation Slides | 04/03/2021 | - |

## Risk Register

| | Risk | Likelihood | Actions |
|---|---|---|---|
| 1 | The database we have selected does not provide the type (or sufficient amount) of information we need in order to correctly do the training. | Medium | We will explore both the options of changing the dataset entirely or integrating the current one with extra data from other new databases. We will base the research on all available platforms such as Kaggle, Github, FiveThirtyeight and Quandl. |
| 2 | The architecture model we choose turns out not to be the most appropriate one for our scope. | Medium-Low | We will induce an urgent meeting to discuss together which other architectures/baseline models are worth trying. Each team member will then be assigned to research one of the architectures/models previously discussed with short internal deadlines. After accurate research is done we will meet again to share our |
| 3 | The baseline model we choose turns out not to be the most appropriate one for our scope. | | |

| | | | |
|---|---|---|---|
| | | | results and decide which one would be the new best one for our project. |
| 4 | A team member drops the course | Low | We will contact the course coordinator for further instructions since we would become a group of 3 students. We will still be prepared to bring the project to completion, perhaps narrowing down the scope a little in order to reduce the workload per student. |

## Link to Colab Notebook

https://colab.research.google.com/drive/1e071Qm2ocoXUipYl5VVBy5Uz09LN2pN4?usp=sharing

# References


[1] "Skin cancer statistics", World Cancer Research Fund International. 2018 [Online]. Available: https://www.wcrf.org/dietandcancer/cancer-trends/skin-cancer-statistics

[2] "Skin Cancer Facts & Statistics", The Skin Cancer Foundation, The Skin Cancer Foundation. 2017 [Online] Available: https://www.skincancer.org/skin-cancer-information/skin-cancer-facts

[3] "Dermatologist-level classification of skin cancer with deep neural networks", Esteva, A., Kuprel, B., Novoa, R. et al., Nature 542, 115–118. 2017 [Online]. Available: https://doi.org/10.1038/nature21056

[4] "Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task", Brinker, Titus J.Ludwig-Peitsch, Wiebke et al., European Journal of Cancer, Volume 113, 47 - 54. 2019 [Online]. Available: https://www.ejcancer.com/article/S0959-8049(19)30221-7/fulltext

[5] "Expert-Level Diagnosis of Nonpigmented Skin Cancer by Combined Convolutional Neural Networks", Tschandl P, Rosendahl C, Akay BN, et al., JAMA Dermatology. 2019 [Online]. Available: https://europepmc.org/article/pmc/pmc6439580

[6] "Automated Dermatological Diagnosis: Hype or Reality?", Navarrete-Dechent, Cristian et al., Journal of Investigative Dermatology, Volume 138, Issue 10, 2277 - 2279. 2018 [Online]. Available: https://www.jidonline.org/article/S0022-202X(18)31991-2/fulltext

[7] "A patient-centric dataset of images and metadata for identifying melanomas using clinical context.", Rotemberg, V., Kurtansky, N., Betz-Stablein, B., Caffery, L., Chousakos, E., Codella, N., Combalia, M., Dusza, S., Guitera, P., Gutman, D., Halpern, A., Helba, B., Kittler, H., Kose, K., Langer, S., Lioprys, K., Malvehy, J., Musthaq, S., Nanda, J., Reiter, O., Shih, G., Stratigos, A., Tschandl, P., Weber, J. & Soyer, P., Sci Data 8, 34. 2021 [Online] Available: https://doi.org/10.1038/s41597-021-00815-z

[8] "Support Vector Machines — scikit-learn 0.24.1 documentation", Scikit-learn.org, 2020. [Online]. Available: https://scikit-learn.org/stable/modules/svm.html

[9] "A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification.", Statnikov, A., Wang, L. & Aliferis, C.F. BMC Bioinformatics 9, 319. 2008 [Online]. Available: https://doi.org/10.1186/1471-2105-9-319

[10] "Detection of Skin Cancer Using SVM, Random Forest and kNN Classifiers", Murugan, A., Nair, S.H. & Kumar, K.P.S. J Med Syst 43, 269. 2019 [Online]. Available: https://doi.org/10.1007/s10916-019-1400-8