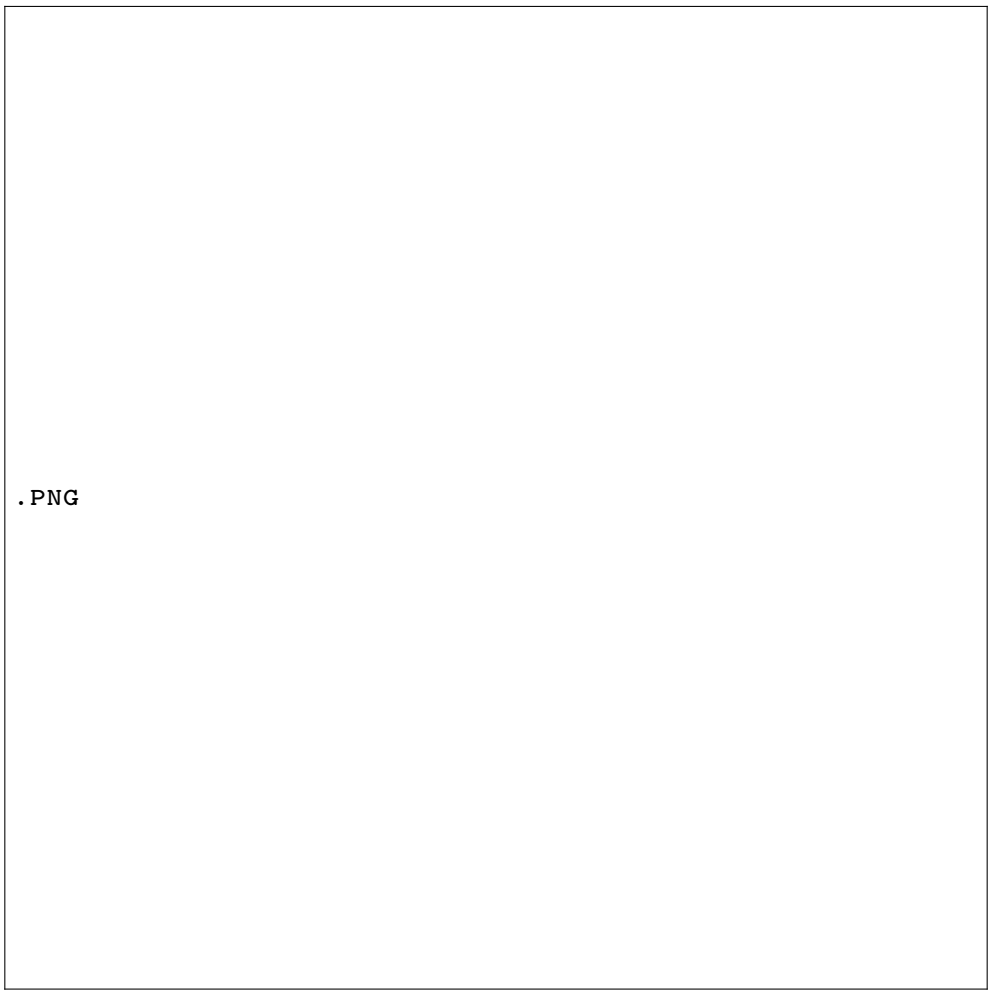
Image Analysis of Skin Lesion: Melanoma

Project 3 Report

Lukas Rasocha (lukr@itu.dk)
Thomas Fosdam Claudinger (thcl@itu.dk)
Martin Kirkegaard (marki@itu.dk)
Frederik Søgren Blunk Raisa (frai@itu.dk)
Yasmin Sarkhosh (yasa@itu.dk)



ORDINARY EXAM
FIRST YEAR PROJECT

Department of Computer Science, IT University
Copenhagen, Denmark
Date: April 23, 2021

1 Introduction

Skin cancer has been growing rapidly over the last years. Usually among population of fair-skinned seen in Europe, North America, and Australia (Nidhal et al., 2017).

Malignant melanoma is the most deadly form of skin cancers, but also highly curable if detected in its early stages (Faziloglu1 et al., 2003). Therefore, developing accessible and proper diagnosis tools can prevent this type of cancer of becoming terminal. Over the years machine learning has been implemented to help support and detect diagnostic decisions among clinicians. However, if the data set does not consists of accurate measurements of feature characteristics whilst being demographically representative it could affect the lives of many and possibly result in deaths. Therefore, ensuring data set balance is crucial for preventing systematic disadvantage of certain groups of people in our society (Kinyanjui et al, 2019). To help us understand how we can approach this kind of data for diagnostic purposes, we will investigate:

"How is the relationship between feature characteristics and skin lesions, and can these features be used to distinguish between types of skin lesions?"

2 Data

This project consists of a collection of data sets including: 1) 150 dermoscopic JPEG format images from ISIC 2017 Challenge[1]. All the dermoscopic images are provided with a unique identifier consisting of 7-digits. 2) 150 binary segmentation masks corresponding to their dermoscopic images from the ISIC 2017 data set. These images are in PNG format and encoded as single-channel (grayscale) 8-bit PNGs, where each pixel is either 0, representing the background of the images, or areas outside the lesion, and 255 representing the foreground of the images, or areas inside the lesion (ISIC 2017). 3) Clinical features such as 'perimeter' and 'area' for each image. 4) Disease classification of each image into labels: Melanoma, Seborrheic Keratosis or non-diseased skin.
We decided to analyse these skin lesions by The ABC Rule commonly used for detecting skin abnormalities. In this project we will solely focus on A and B features (Nidhal et al, 2017).

The ABC Rule		
A: Asymmetry	B: Border	C: Colour
The two halves of a skin lesion does not match	The edges of the skin lesion are irregular, ragged and blurred	The skin lesion shows variation of brown/black colours

Table 1: Definition of The ABC Rule

Before we dove into the data, we checked it for missing values, which there are none of. Our results were achieved by a series of steps throughout the project. These steps included:

[1]https://challenge.isic-archive.com/landing/2017

2.1 Step 1: Pre-processing

The dermoscopic images can contain unwanted particles such as hair, air bubbles or markers for measuring size of lesions. We can minimise these by cropping images and filtering out noise with morphological operators. This step is particularly important for customising segmentation masks.

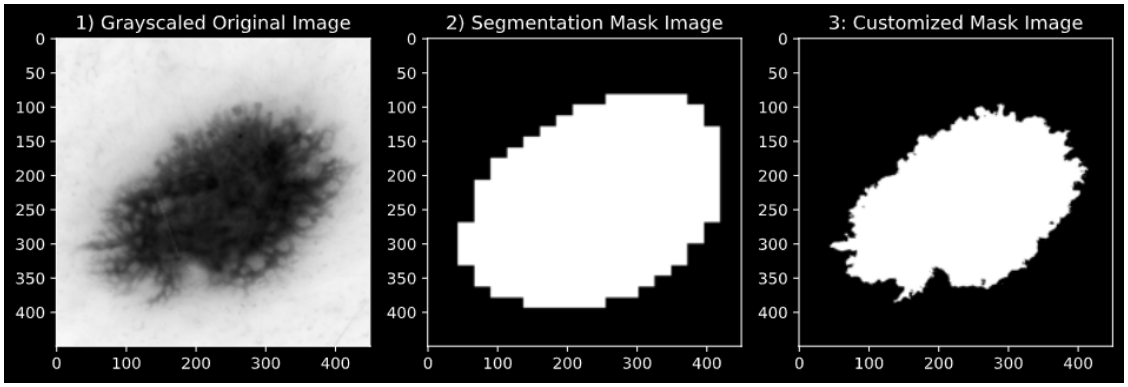


Figure 1: Visualisation of ISIC 2017 segmentation images and customised binary images

2.2 Step 2: Segmentation

The segmentation images are important for analysing the lesions properly. The images are divided into two regions; one for lesion and the other for skin. Thus, resulting in binary images.

2.2.1 Step 2.1: Asymmetrical Shape

The asymmetrical shape of each image were analysed by binary images. Firstly, we rotated each image by 10 degrees 5 times. For each rotation we: 1) cropped the image and calculated center coordinates, 2) divided images into two horizontal halves by center points, 3) flipped the one half, and overlapped these, 4) calculated the intersection and the symmetric differences for each side, and lastly 5) we extracted the minimum value for the left and the right horizontal symmetric differences performed for each rotation. As the output are in pixels, we converted each value by dividing the area of each half with the symmetric difference. These values are presented as 'Asymmetric decimals' and saved into a comma-separated file for further analysis.

2.2.2 Step 2.2: Border

The border of a lesion was found by subtracting a copy of the lesion's segmentation (from the original segmentation) with a slightly smaller size, leaving only an outline of the original lesion. Since pixels in the segmentation image have value of 1 for lesions and 0 everywhere else, the border was found by summing all the pixels over the newly created outline.

2.2.3 Step 2.3: Area

Area of a lesion was calculated for binary images, where the area of the lesion was the sum of all the pixels in the segmentation (since pixels around the lesion have value of 0).

2.2.4 Step 2.4: Compactness

Compactness is a dimensionless property that is defined as the ratio of the area of an object to the area of a circle with the same perimeter.

$$Compactness = \frac{Perimeter^2}{4\pi * Area}$$

A circle is an object with the most compact shape (objects with a complete circle shape will have compactness 1).

2.3 Step 3: Classification

There are models, that are more advanced like the Convolutional Neural Network (CNN) and others, more basic models like k-Nearest Neighbours (KNN). The choices are endless, and they all contribute with different things. The flow chart on Scikit-Learn[2] guided us to pick the best fitted model for this project. The recommended models were the KNN or the Support Vector Classification (SVC). For this project we chose the KNN model.

This classifier takes one parameter, namely the number of neighbours we are interested in[3]. To find a K, K starts at $k = 1$ to $k = \sqrt{N}$ where n is the number of data points. Then our model system uses majority rule to identify the class, i.e. the class with the highest number of votes (Hassanat et al, 2014).

3 Results and discussion

We used different approaches regarding the training of our KNN-model. We started with choosing an appropriate K. We tried to run a bunch of different predictions using a different K in the range $K = \{1..\sqrt{N}\}$ We then chose the K with the highest accuracy closest to $\sqrt{N}$. The accuracy can vary greatly when working with classifiers tested on small amounts of data. As our sample size is small, we decided to approach the 5-fold cross validation to validate our results.

3.1 Relationship between Features and Test Accuracy

We tried a several combinations of features firstly by using all categories of features, meaning no dimension reduction and with 3 labels, $\{Melanoma, Keratosis, None\}$. In this test we got an accuracy of 45% when doing cross validation. Then we gradually reduced the feature space. However, not by using Principal Component Analysis PCA, but more a "Trial and Error" method. We discovered which features have little to no impact. Features such as "Average color", "deviation_of_colour" and "Area". The feature mix, which yields the highest accuracy score, was "compactness" and "asymmetry".

[2] https://scikit-learn.org/stable/tutorial/machine_learning_map/

[3] (small value leads to unstable decisions, large values to expensive computations)

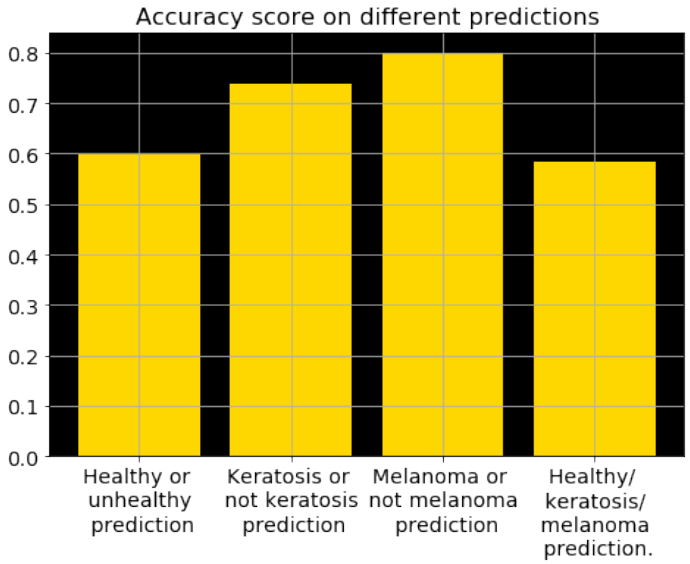


Figure 2: KNN-model predicting different categories. Y-axis is accuracy of a given model

3.2 Prediction of skin lesions and Test Accuracy

3.2.1 Test case 1: non-diseased skin and diseased skin

Next, we tried to predict the labels for non-diseased skin an, diseased skin. As the data set provided had fairly a 50 / 50 data points for these two categories, ideally for training the data set. From Figure 2 the model for this category had accuracy about 60 %. One explanation for the low accuracy could be poorly chosen features for training the model, or assuming we could merge melanoma and keratosis under one category since both skin conditions could exhibit different behaviours under the selected features.

3.2.2 Test case 2: Keratosis

Then we examined the prediction of the concrete skin conditions. The model was now trained to predict keratosis from our earlier selected features. Now the model performed slightly better, predicting about 70 % correctly. This result however needs to be examined more in detail. Since now the categories are highly unbalanced we need to look into sensitivity and specificity. Specificity in this case is 95 % which makes sense since the majority of the data is non-keratosis (so the model is more likely to predict everything as non-keratosis). Whereas sensitivity is only 33 % which doesn't necessarily mean that the model nor the features are wrong, it could be just the imbalanced amount of data.

3.2.3 Test case 3: and Melanoma

Trying to predict melanoma showed similar problems as stated in the previous case. Furthermore, classifying melanoma showcased the most extreme situation where the sensitivity was 0 % and specificity 100 %, this means that all the predictions on our test set were non-melanoma, identifying exactly 0 lesions with melanoma.

This problem could be solved either by oversampling or undersampling and training the model on our oversampled (or undersampled) data set.

3.2.4 Test case 4: Melanoma, Keratosis and Non-diseased

The last test was done on 3 different categories: Melanoma, Keratosis and Non-diseased. This led to an accuracy about 57 %, possibly due to poorly selected features, data imbalance or classifying into 3 labels considering the low amount of data we had.

4 Limitations

A sample size of 150 dermoscopic images consisting a weak representation of diverse skin colours is not considered to be a representative sample of a whole population. This has contributed to be one major limitation of this project. Nonetheless, photographically technicalities such as over or underexposure can inflict with the natural lighting of the ISIC 2017 images, possibly skewing the results of our testing (Maio, 2021).

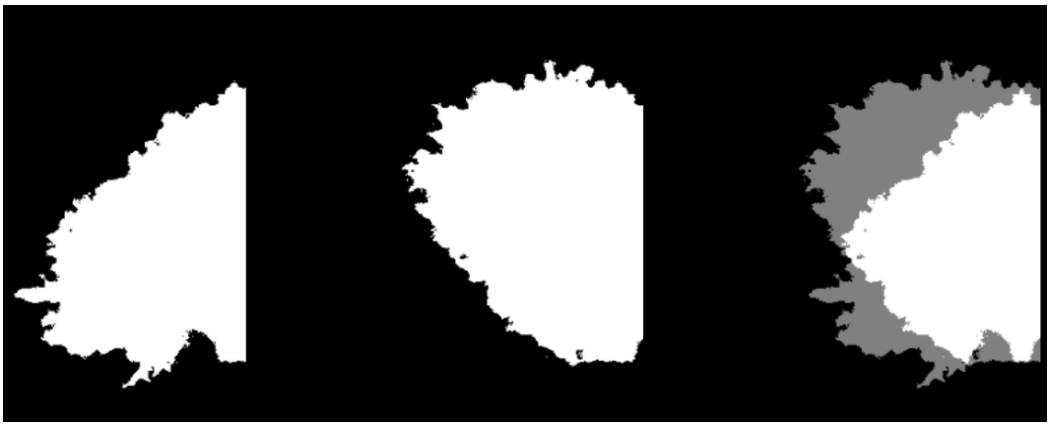


Figure 3: The grey area shows the non-overlapping part of the two horizontal halves

To check whether the model meets requirements for being representative for a whole population, the provided data sets were missing information such as gender, age and demographics. Not only to establish if the data covers the vast majority of a population, but also useful for determining whether the model is better for predicting skin lesions when compared to gender, age and/or demographics (Lashbrook, 2016).

5 Concluding remarks and future work

In this project we approached image processing techniques for analysing patterns between characteristics derived by The ABC rule and clinical features to skin lesions. Considering mentioned limitations, the small sample size and the skewed amount of data, the model is not able to deliver a prediction accurate enough to determine whether a skin legion is diseased skin (Keratosis, Melanoma) or non-diseased skin (either Keratosis nor Melanoma).

5.1 Future work

The results are based foremost from analysing dermoscopic images. Ensuring their quality is crucial for creating accurate and solid algorithms for diagnostic purposes. However, we experienced these images to differ from clinic to clinic depending on their instruments, and technical

knowledge and skills. Creating a general framework for clinicians could increase dermoscopy quality, but also make the data collection process among clinics and countries more identical. Thus, preventing potential data imbalances.

Furthermore, exploring and testing methods to implement standard categorisations for different kinds of skin tone in machine learning, will not only factor all groups of people, but also strengthen the accuracy for future research among this field. If combining a clinical framework for clinics and standard skin tone patterns in pixel data analysis, detecting skin abnormalities might probably give more accurate classifications (Lashbrook, 2016).

6 Disclosure statement

All the code in the final notebook was organised by ..... However all group members contributed to it through writing code in older versions of the main notebook and their own personal notebooks.

7 References and Bibliography

Faziloglu, Y., Stanley, J., Moss, R., Stoecker, W., McLean, R. (2003). *Colour histogram analysis for melanoma discrimination in clinical images*. Skin Res Technol. 2003 May ; 9(2): 147–156. [online].
Available: https://pubmed.ncbi.nlm.nih.gov/12709133/

Hassanat B. A., Abbadi A. M., Altarawneh A. G., Alhasanat, A. A. (2014). *Solving the Problem of the K Parameter in the KNN* (IJCSIS) International Journal of Computer Science and Information Security, Vol. 12, No. 8, August 2014. [online].
Available: https://arxiv.org/ftp/arxiv/papers/1409/1409.0919.pdf

Kinyanjui, K., Odonga, T., Cintas, C., Codella, N., Rameswar, P., Sattigeri, P., Varshne, R. (2019). *Estimating Skin Tone and Effects on Classification Performance in Dermatology Datasets*. NeurIPS 2019 Workshop on Fair ML for Health, Vancouver, Canada. [online].
Available: https://rpand002.github.io/data/NeurIPSW2019.pdf

Lashbrook, A. (2016).*AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind*. The Atlantic, Health. [online].
Available: https://www.theatlantic.com/health/archive/2018/08/machine-learning-dermatology-skin-color/567619/

Maio, A. (2021). *What is Overexposure in Photography How to Fix It*. Studiobinder. [online].
Available: https://www.studiobinder.com/blog/what-is-overexposure-in-photography/

Nidhal K. EL A., Faisal, F. (2017). *Detection and Analysis of Skin Cancer from Skin Lesions*. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 19 (2017) pp. 9046-9052 [online].
Available: https://www.ripublication.com/ijaer17/ijaerv12n19139.pdf