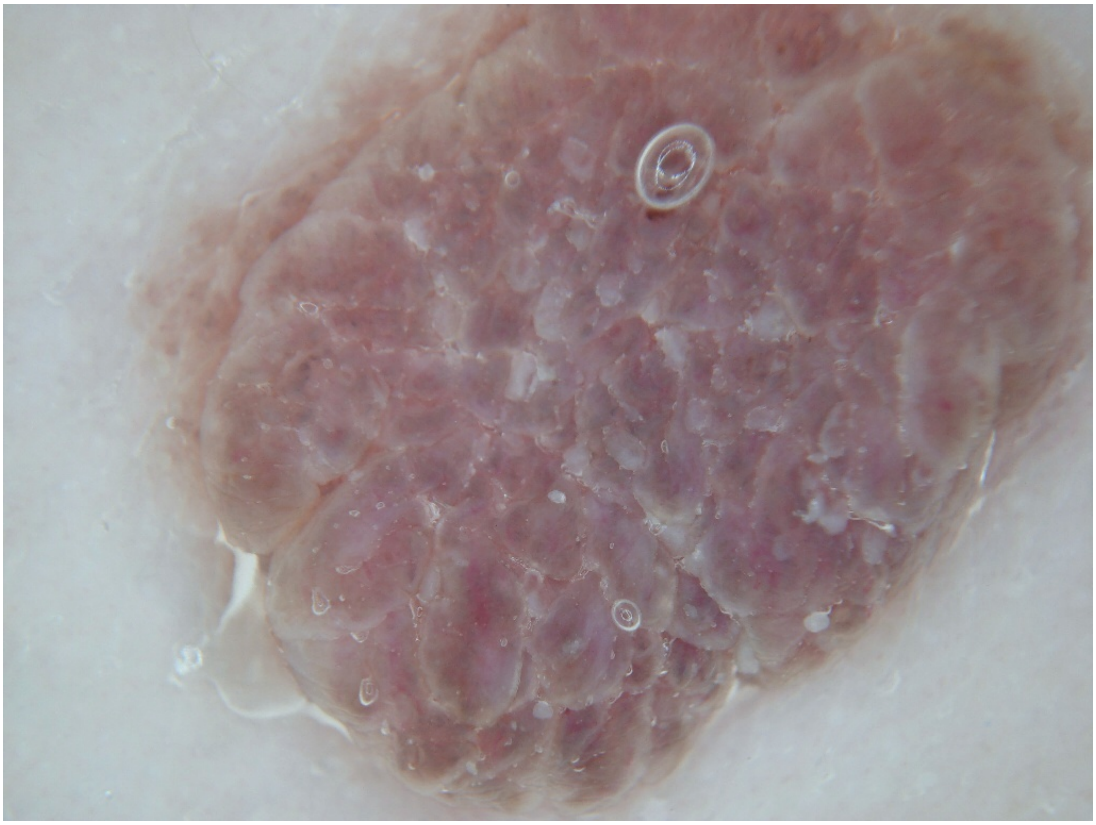
Analysis of Melanoma

Project 2 - Group G



Healthy skin lesion

Christina Aftzidis, *caft@itu.dk*
Germán Buttiero, *gebu@itu.dk*
Hubert Wójcik, *huwo@itu.dk*
Laurids Robert Holme Pedersen, *lrpe@itu.dk*
Paula Victoria Menshikoff, *pmen@itu.dk*

IT University of Copenhagen

April 8^{th} 2022

1 Introduction

We have been hired by the local dermatologist to help them predict possible melanoma cases. There are simply too many images for them to handle alone so they want our help.
The dermatologist has helped us out and told us that normally the skin lesions are classified by the 'ABCDE' scale ("What to look for: ABCDEs of melanoma", n.d.).

Asymmetry: A healthy skin lesion mimics the shape of a circle, but in general the more uniform the shape, the better.

Border: A healthy skin lesion retains a consistent border all around. The contrast between skin and skin lesion is therefore easy to see. A healthy border is also smooth meaning it does not consist of squiggly lines.

Colour: Uniform colouring means healthy skin lesion, even a red lesion can be healthy but you should still be more wary of that compared to a (non-red) skin lesion which has a consistent colouring throughout.

Diameter: The diameter of a skin lesion usually stays below 6mm across. Should yours grow beyond that, get checked.

Evolution: Typically a skin lesion forms and then stops growing and changing. If you notice a sudden change in a skin lesion it might be a sign that it is not healthy.

In order to help the dermatologist, we train a classifier to predict melanoma from images of skin lesions trying to answer the question

Is it possible to predict melanoma from an image of a skin lesion?

2 Data

We were provided with 150 images of skin lesions (.jpg). We also got their equivalents in black and white (.png) in order to facilitate the process of selecting only the lesion from an image (masking). Two csv files enabled us to get to know the area and parameter of each lesion, and whether an image displays a melanoma.

However, we eventually have decided to use a larger dataset with two thousand images (1744 after removing unknown values), where there was additional information with the sex of a person as well as their age. The data set was taken from the ISIC 2017 challenge (Rotemberg et al., 2021). It was needed for our research question.

For every iteration in the dataset, an instance of our custom-made Picture class was initiated, where we provided it with two attributes - the image and it's segmented version (black and white). The main image was loaded using Matplotlib's `plt.imread()` function, whereas for loading the other image, we used the

`PIL.Image.open()` from the Pillow library, as it was needed for our assymetry-investigating function to work.

Task 1 entailed us trying to predict whether a lesion is cancerous or not. The same 50 images where investigated by all members of our group. What we could assess were the first three parts of the ABCDE method - asymmetry, border and color.

The rules we followed were:

1. Asymmetry:
   (a) 0 if the image is symmetric
   (b) 1 if there are signs of asymmetry
   (c) 2 if the image is (almost) completely asymmetric
2. Border:
   (a) 0 if the border looks smooth
   (b) 1 if there are parts of the border that are uneven
   (c) 2 if the border looks rough
3. Color:
   (a) 0 if there is no lesion's color variability
   (b) 1 if there are signs of lesion color variability
   (c) 2 if the lesion's color variability is high

Finally, everyone assigned an overall value (0 or 1) for each image, which indicated whether we think the lesion is cancerous - 0 if it is not, 1 if it is melanoma.

The overall accuracy of this group's judgement is considered low. Only 7 images (out of 50) were managed to be guessed correctly by every member of the group. On average, a person guessed that 59,6% of the lesions were melanoma. It turned out that only 6% of the pictures were showing cancerous lesions. This example demonstrates how inaccurate human's assessment can be, which is a crucial argument for implementing more machine learning models in the health industry.

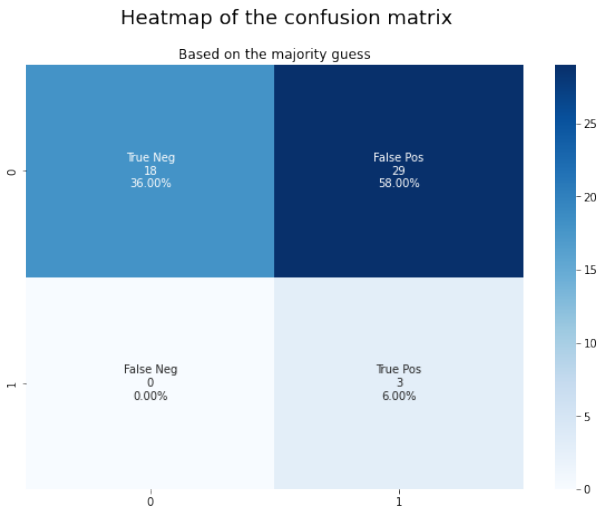


Fig. 1: Heatmap of our manual guesses

When discussing our staggeringly bad accuracy, there emerged reasons for such a phenomenon. Firstly, we all felt that it was better to check a lesion as cancerous if we were not sure. Moreover, the experiment was conducted at the beginning of our project, when we had less knowledge about melanoma features.

3 Feature Extraction

The features selected to check whether the lesion is a melanoma were asymmetry, compactness and color variability. These features were quantitatively measured after the images were preproccessed (segmented images needed to be cut in order to get an even width and length required for asymmetry processing).

Asymmetry: quantitative measure that evaluates how symmetric a lesion is. This feature is measured by folding the lesion mask image (black and white pixels) and calculating the ratio of the amount of pixels that, when overlapping the image, differ in value, and the size of the folded image. This process was repeated from different angles in order to get a better estimation of the asymmetry. A complete symmetric image (i.e. a circle) would have a value of 0.

Compactness: this feature is measured as the ratio of the area of an object to the area of a circle with the same perimeter. A circle is used as the object with the most compact shape, presenting a value of 1. Therefore, the closer the compactness of a lesion is to 1, the more compact the shape is.

Color variability: both color and mask image were used for this measurement. By using the mask on the color image we were able to extract just the lesion. We separated the image in three different color dimensions (red, green, blue) and then we applied the variance formula for every color and returned the mean of the values. The higher the value, the more color variability the lesion has.

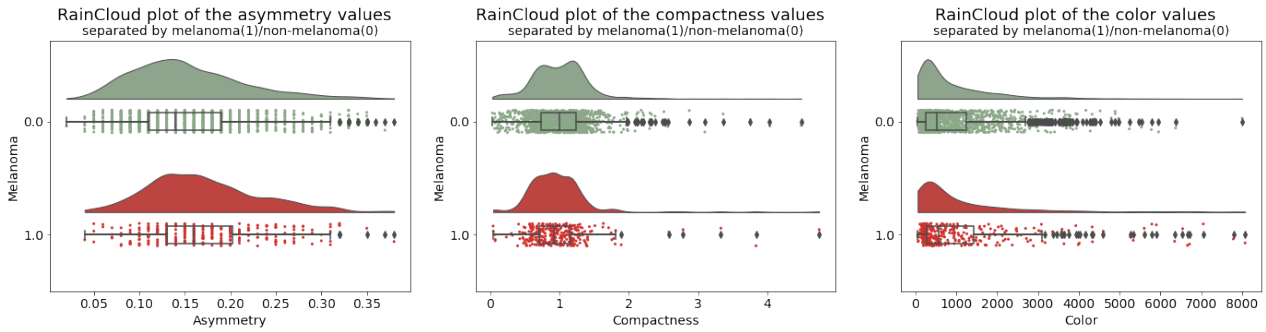


Fig. 2: Raincloud plots of the different features

4 Classification

During the classification step we split the dataset of 2000 pictures from the 2017 challenge Rotemberg et al., 2021 into a train, a validation, and a test set. The features we take into account are asymmetry, color variability, and compactness. In order to have a more accurate model, we balance our dataset. We do this by oversampling.

We use the training set to train three different classifiers: the nearest neighbours classifier, the decision tree and the Gaussian Naive Bayes classifier.

When running the algorithm, the Naive Bayes classifier has both low accuracy and low ROC AUC scores and is therefore discarded.

For the nearest neighbour classifier, we run the code multiple times for multiple different neighbours, with a new train set each time and kept track of which amount of k neighbours had the highest accuracy and Area under the Receiver Operating Characteristic curve (ROC AUC) score. The highest accuracy and ROC AUC score was achieved with k = 5. However, it is still lower than the score of the decision tree classifier which is why we decide to use the decision tree.

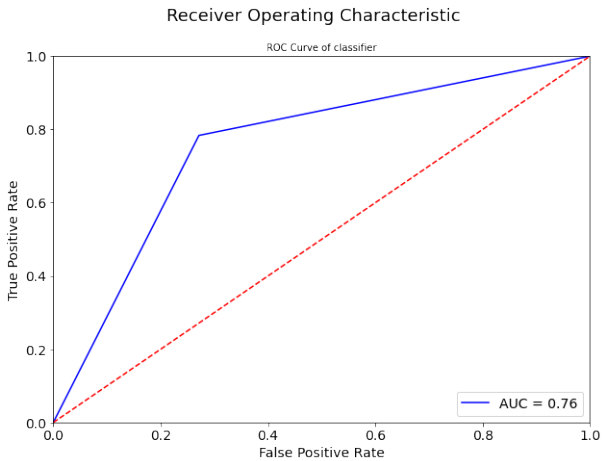


Fig. 3: Plot of the ROC curve with the DecisionTree classifier

For our open question we also used the 2000 picture dataset. The data comes with a set of metadata containing the sex and approximate age of the skin lesion host. For some images this is unknown and as part of our data cleaning these were removed from our image set.

After cleaning the dataset we split it up by sex and trained classifiers on both sets of images to see if there was a performance difference between the two. We also use the decision tree classifier for this task. The classifiers for female skin lesions performed a bit better than the ones for male skin lesions.

5 Discussion

As commented earlier, the subjectivity in interpreting the ABCDE rule among humans lead to different outcomes which were not conclusive. This is why, the proposed solution involved the systematisation of the main stages of the image analysis process; segmentation, feature extraction and classification. As opposed to using an individual ABCDE rule feature, a combination of features could lead to better accuracy in melanoma detection (She et al., 2007). As stated in several papers, the automatic recognition of melanoma has several challenges (Ali et al., 2020). First of all, the low contrast between skin lesion and normal skin region makes it difficult to segment accurate lesion areas. Secondly, both melanoma and non-melanoma lesions may have high degree of similarity. And finally, the variation of skin conditions; namely skin color, natural hairs, among patients produce different appearance of melanoma in terms of color and texture.

In order to address the aforementioned challenges we segment the skin lesion to form a region of interest, as proposed in (Ali et al., 2020). From this region of interest the measures are extracted and asymmetry, color variation and compactness are evaluated. The main advantage of extracting the features using the automated methods discussed in this paper is the ability to provide an objective second opinion to the patient or even the dermatologist which would otherwise be prone to subjectivity. On the other hand, many of the features extracting methods rely on the segmentation mask which could be degraded due to the presence of different factors (like hair, pens that dermatologist use to measure the skin lesion). Additionally, the methods also lack the ability to reflect real-world measures like the diameter measurement.

Overall, the classifier was not totally accurate because our specificity is 73% while our sensitivity is 78%. Meaning that, there are more false positives than false negatives which is important because melanoma are malignant and need to be treated.

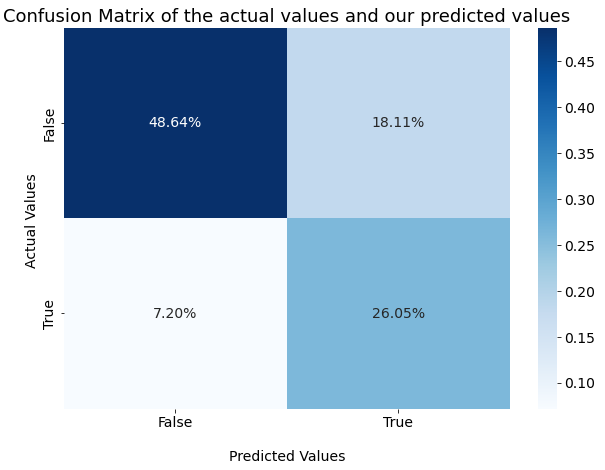


Fig. 4: Heatmap of our classifier

6 Limitations

6.1 Limits in our data-set

Our dataset was composed of 2000 images of light to intermediate colored skin patients. However, none of the patients had dark skin. Which means that our algorithm may not be as accurate with those patients.

6.2 Limits due to time

In terms of validity, the concern here is that we only had two weeks to carry out this project. The team members for this projects had no professional background on the domain. However, this research was based on previous work done by professionals in the domain. (She et al., 2007)(Ali et al., 2020) (Maglogiannis and Doukas, 2009)

6.3 Skin lesions

Some images do not completely cover the skin lesion and the compactness score is therefore skewed.

7 Conclusion

In conclusion, we realised that the classification on the individual features have a limited discrimination capability, as opposed to the classification on the features combined with the lowest and highest ROC AUC scores reported in the study are 0.69 and 0.77 with the implementation of our algorithm.
This research will serve as a starting point for researchers interested in automating the ABCDE rule. Nonetheless, future work must be done with more public data sets and a more diverse group of patients to enhance the outcomes in this domain. The main reason for this is because there is still a margin of performance improvement for both skin lesion segmentation and classification. Additionally, we still need to bring other aspects of patients life styles into the picture to get a broader understanding on how and why melanomas appear.

References

Ali, A.-R. H., Li, J., & Yang, G. (2020). Automating the ABCD Rule for Melanoma Detection: A Survey [Conference Name: IEEE Access]. *IEEE Access*, *8*, 83333–83346. https://doi.org/10.1109/ACCESS.2020.2991034

Maglogiannis, I., & Doukas, C. N. (2009). Overview of Advanced Computer Vision Systems for Skin Lesions Characterization. *IEEE Transactions on Information Technology in Biomedicine*, *13*(5), 721–733. Retrieved April 7, 2022, from https://www.academia.edu/658741/Overview_of_advanced_computer_vision_systems_for_skin_lesions_characterization

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., ... Soyer, H. P. (2021). A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Scientific Data*, *8*(1), 34. https://doi.org/10.1038/s41597-021-00815-z

She, Z., Liu, Y., & Damatoa, A. (2007). Combination of features from skin pattern and ABCD analysis for lesion classification. *Skin Research and Technology*, *13*(1), 25–33. https://doi.org/10.1111/j.1600-0846.2007.00181.x

What to look for: ABCDEs of melanoma. (n.d.). Retrieved April 8, 2022, from https://www.aad.org/public/diseases/skin-cancer/find/at-risk/abcdes