Final Assignment

Aniela Marta Ciecierska
anci@itu.dk
Francisco Gonçalves Medeiros Lemos Moreno
fmor@itu.dk
Jakub Piotr Gąsior
jaga@itu.dk
Jonas Drøivoldsmo Lesund
jles@itu.dl
Michaela Macejovska
mimac@itu.dk

Abstract

This paper showcases the route that the members of the group G took in order to explore the capabilities of a machine to identify melanoma, as well as their findings, through different programming algorithms and human evaluations. It contains a description of the process, auxiliary graphs and a conclusion of the results found.

1 Introduction

Artificial intelligence is becoming an undeniable force in technological advancements. Its true capacity is yet to be comprehended, and for this reason, people disagree on what it can or cannot do, as well as what it should or should not do. One of the fields where AI might become a usual tool is medicine. This report will focus on its possible presence in this field, more specifically for detection and diagnosis of melanoma. Melanoma is a form of skin cancer that develops from melanocytes, a type of skin cells. It occurs when these cells grow uncontrollably until it eventually becomes malign. This phenomenon can be identified based on different factors. In the report, the group will study the theory of if an algorithm can be trained to correctly detect and diagnose melanoma based on its asymmetry, color variability and the presence of a blue-white veil.

2 Data

2.1 Source

The PAD-UFES-20 dataset was created in collaboration with the Dermatological and Surgical Assistance Program (PAD) at the Federal University of Espírito Santo (UFES) in Brazil. This program offers free treatment for various skin lesions, focusing on those unable to afford private care. The dataset comprises 2,298 samples representing six distinct types of skin lesions. Each sample includes a clinical image and up to 22 clinical attributes, such as patient age, lesion location, Fitzpatrick skin type, and lesion diameter. Approximately 58% of the dataset's samples are biopsy-confirmed. The images in the dataset vary in size due to being captured by different smartphones and are stored in .png format. Metadata accompanying each lesion contains up to 26 features, available in a CSV file where each row represents a lesion and each column represents a metadata feature. Overall, the dataset comprises data from 1,373 patients, 1,641 skin lesions, and 2,298 images, with each image/sample linked to the corresponding patient and lesion in the metadata.

2.2 Data Cleaning

Upon analysis, the images were classified into two distinct categories: low quality and high quality. Upon further examination, it was evident that within the Low quality subset, some images still retained identifiable lesions. Consequently, the Low quality category was subdivided into usable and unusable segments. Notably, four pictures from the data set were deemed unusable (`PAT_246_377_159.png`; `PAT_153_233_45.png`; `PAT_356_4511_960.png`; `PAT_1618_2771_628.png`) due to the quality being too low to identify any feature relevant for this report.

3 Methodology

3.1 Image Reduction

The images from the data set include a large area that is not relevant for this research. For this reason, the group conducted a process of segmentation masks to crop the image down to a thin frame around the identifiable lesions. Not only did this process allow for a more adequate identification of

the RGB channels of the lesion, but also promoted a swifter process and run time overall.

3.2 Feature Extraction

The group created different functions in order to extract specific features from the images. The features chosen for the tests were based on the medical suggestions that high color variability, low symmetry and presence of blue-white veils could indicate melanoma. Additionally, these factors are relatively more evident to the human eye, when compared to the other possibilities, which contributed to the decision of the group for which features to work with.

Color: This code loads an image and its corresponding mask, calculates the variance of colors within the lesion area defined by the mask. Then, it computes a color score based on the sum of variances across all color channels. Finally, it categorizes the score into the ascending levels of 1, 2, 3 or 4.

Symmetry: The code use here computes the symmetry score of a lesion based on its shape and orientation. It starts by doubling the size of the input image with a black background, then, it finds the longest diameter of the lesion in the mask. It aligns the longest diamter vertically by rotating the mask, then cropping it to only include the left and right sides of the lesion. It calculates the pixel differences between the left and the right halves vertically and horizontally, determining the fraction of similarity for both cases. Finally, it places each score into, again, four ascending levels (1, 2, 3 or 4).

Blue-White Veil: For the third test, the objective is to identify the occurrence of a blue or white veil, which can be indicative of melanoma. The code converts the image to the HSV color space, and defines ranges for detecting blue and white colors in it. It creates masks for the blue and white colors separately. Then, it combines the to identify areas with a blue-white veil, and it calculates the ratio of its area to the area of the whole lesion. Finally, it returns a binary score where 1 represents the presence of a blue-white veil and 0 the absence.

3.3 Group Test

For all three of the above mentioned tests, the members of the group conducted an individual rating of each image into the same categories as the code did. The average of these results was used to compare the interpretation of the images by the human eye and by a computer.

3.4 Comparison between Computer and Human

For each of the tests, the results of the computer and the mean of the group members were compiled into plots. This mean is then approximated to its closest rating (1, 2, 3 or 4). If this value matches the computer test, it creates a 'yes', otherwise, a 'no'.

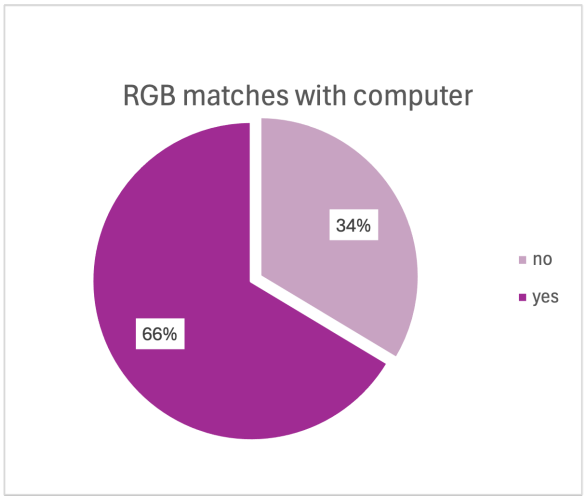

Figure 1: Matches between the computer and the human RGB tests

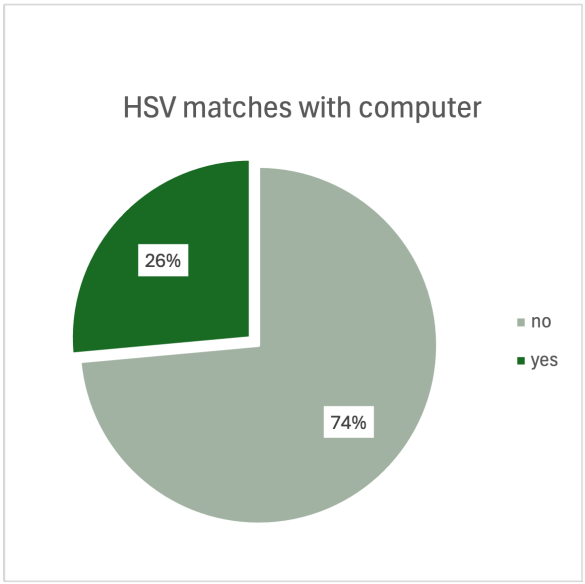

Figure 2: Matches between the computer and the human HSV tests

After analyzing both charts, the different inaccuracies of the RGB test (34%) and of the HSV

test (74%) was the reason why the group chose to use only the RGB test for the remaining of the study.

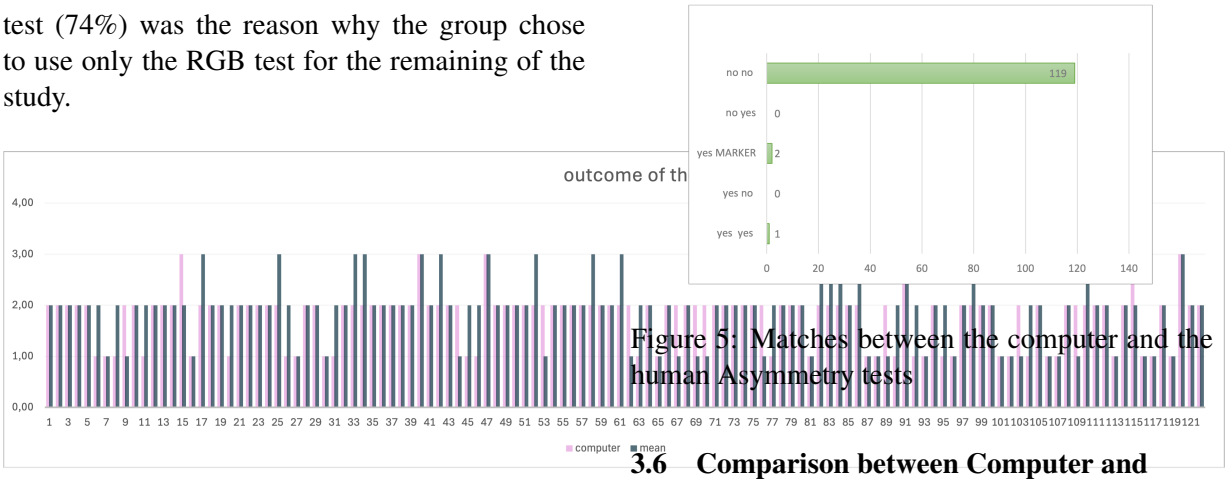

Figure 3: Differences between the computer and the group's mean RGB tests

This graph shows that although there are some instances where the mean does not match the computer, which generates a "no" as represented in Figure 1 , the deviation is usually only one point in the scale. This proves that the level of inaccuracy, although being 34%, is not as big as the graph would suggest.

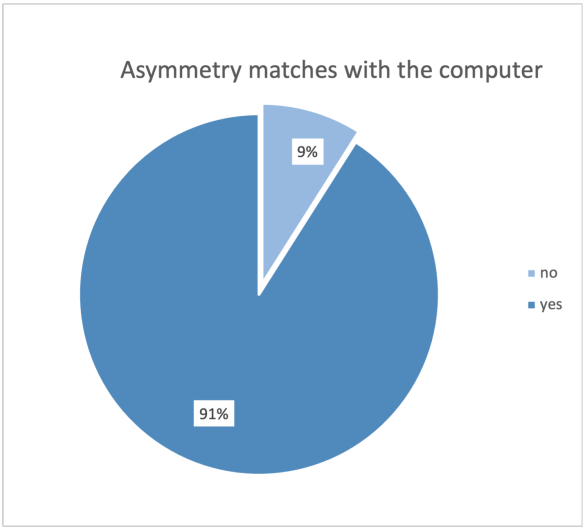

Figure 4: Matches between the computer and the human Asymmetry tests

3.5 Biopsy

The data set used for the report includes the actual diagnosis made by a biopsy to each lesion in the images. This diagnosis classifies each image based on the incidence of melanoma, where 'TRUE' and 'FALSE' represent if it is, or is not, considered melanoma, respectively.

Figure 5: Matches between the computer and the human Asymmetry tests

3.6 Comparison between Computer and Biopsy

TABLE

4 Machine Learning

.....

5 Limitations

...

6 Open Question

.....

References

Alfred V. Aho and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation and Compiling*, volume 1. Prentice-Hall, Englewood Cliffs, NJ.

American Psychological Association. 1983. *Publications Manual*. American Psychological Association, Washington, DC.

Association for Computing Machinery. 1983. *Computing Reviews*, 24(11):503–512.

Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. 1981. Alternation. *Journal of the Association for Computing Machinery*, 28(1):114–133.

Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.