First Year Project - Project 3

Medical imaging

Group 8

Ida Maria Zachariassen
idza@itu.dk

Sabrina Fonseca Pereira
sabf@itu.dk

Magnus Sverdrup
magsv@itu.dk

Rasmus Bondo Hansen
rabh@itu.dk

Ruben Oliver Jonsman
rubj@itu.dk

April 23, 2021

1 Introduction

The project presented in this report was developed with the purpose of researching and analysing the visual identification and classification of potential melanoma cancerous lesions through imaging. This project also aims to investigate how a model based on the characteristics of melanoma lesions performs when classifying keratosis lesions. This was done to gain insight and knowledge about extracting features from medical imaging and using these features in simple classifiers. Lastly the report is an evaluation on whether visual classification of skin lesions is a reliable source for cancer detection. Based on our analysis of various models for classifying features of skin lesions, the following research question was formulated: "How do models trained on melanoma lesions perform when classifying seborrheic keratosis lesions?"

2 Data

The initial raw data came from an ISIC 2017 challenge, provided by Veronika Cheplygina, Associate Professor at ITU. The data consists of 150 images of lesions. Each image is stored as a JPG file with a corresponding masks stored as a PNG file. To link the masks and images there were provided a ground truth csv file. The csv file labels each image to an id and as having melanoma, seborrheic keratosis or being healthy. To improve the accuracy of our machine learning models, an external data set was found at the ISIC Challenge archive [1]. This consists of 2000 new images and masks as well as a ground truth csv file.

To first validate the raw data, we checked that each lesion id was unique and that each id had an image and a mask connected to it. No missing data or non-connected lesion ids were found, thus no extra cleaning needed. Then we checked the quality of the pictures and the segmentation files, these were good as well.

To easily manipulate the csv file, it was loaded into a pandas data frame. Since the main focus for the project was to predict the skin cancer type melanoma, the data frame was filtered to only contain rows with healthy or melanoma cancerous lesions. After calculating each of the features, they were inserted into the data frame. The features based on the ABCs of lesions being:

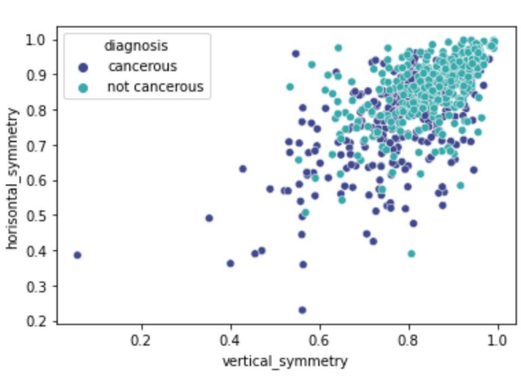
2.1 Asymmetry

Cancerous lesions tend to not have perfect symmetric shapes, therefore if you draw a line through the lesion, the two halves will not match. To measure this feature on a lesion, we have come up with an algorithm that turns and flips the mask of a given lesion and calculates how much mass lies outsides of the original boundaries. This is done two times, both for the vertical symmetry and the horizontal symmetry. What we found for this feature was that cancerous lesions on average tend to be slightly less symmetric than healthy lesions.

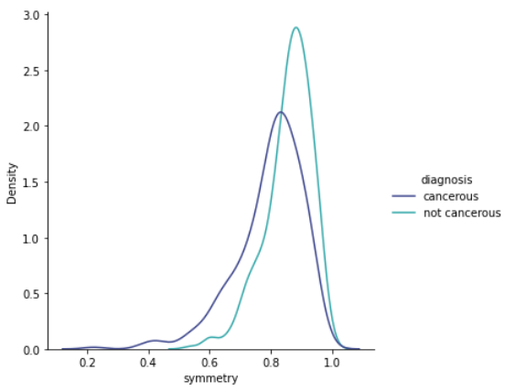
2.2 Border

To identify skin lesions with edges that are irregular, ragged, notched, or blurred we calculated the area, perimeter, radius mean and radius standard deviation.

[1] ISIC, *ISIC Challenge Data sets* Accessed: 20/04-2021, Retrieved from: https://challenge.isic-archive.com/data



(a) Scatterplot of Horizontal and Vertical Symmetry



(b) Plot of Density to Symmetry

Figure 1: Plots of Horizontal and Vertical Symmetry and Density to Symmetry

2.2.1 Area & perimeter

Segmentation masks were used to calculate the area and perimeter of each lesion. The area was calculated as the sum of all white pixels in the mask. For the perimeter, the mask was reduced by a few pixels and subtracted from the original mask. This left a thin line of white pixels, the sum of these was then used as the perimeter.

2.2.2 Compactness & Radius standard deviation

To calculate the compactness we used the following formula:

$$Compactness = \frac{perimeter^2}{4\pi * area}$$

Compactness is used to identify the roundness of a lesion, the higher the value the less round a skin lesion is. But this does not necessarily tells us about how jagged the outline of the lesion is. To identify this, we implemented the radius standard deviation feature. This takes all the radial distances in the lesion, and calculates the standard deviation. To compute this, we first find the center of the lesion. From there we use the perimeter and calculate the distance between the center and each pixel in the perimeter. From that list we calculate the standard deviation between all values found.

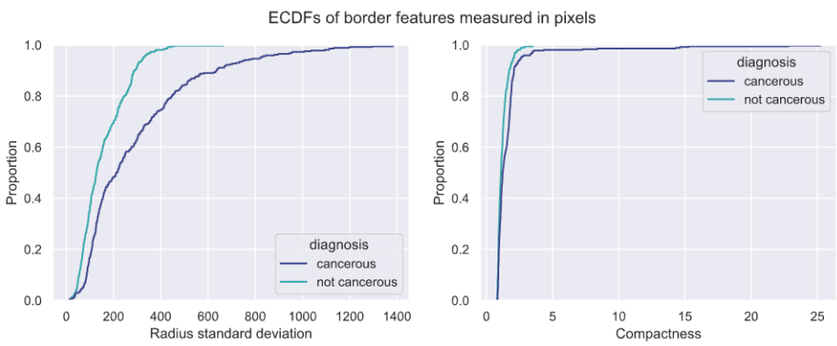
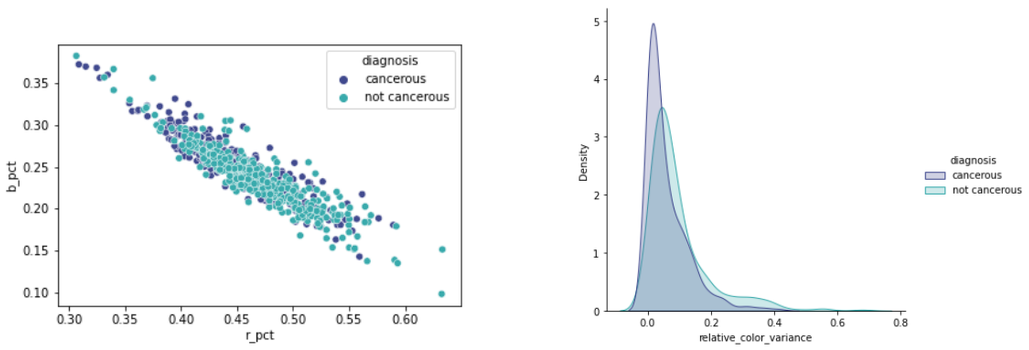


Figure 2: Comparison of features radius standard deviation and compactness

When comparing the empirical distributions of compactness and radius standard deviation values, we can clearly see how the radius standard deviation has a much clear division between cancerous and non cancerous lesions in our data set and will probably be useful when creating a classifier.

2.3 Color

When characterising a lesion as healthy or cancerous, the color plays a big role. The color is measured to check for a possible uneven distribution of the R, G and B values or big presence of blue and green values rather than red. To do so, the image files of the lesion were masked using the segmentation files to only show the lesion, leaving the rest black. All RGB-values of the lesion, excluding the black pixels, were found to make further calculations. The relative color variance was found as the number of unique RGB-values in relation to the total number of pixels. This color variance was found to be bigger for the cancerous lesions than for the healthy lesions. The standard deviation of the R, G and B-values were calculated to get the amount of variation of each of these, as well as the average of these. Further individual calculations of the R, G and B-values were made to get the range and percentage of each, thinking the range and percentage of each would tend to be more uneven for melanoma lesions than for healthy lesions.



(a) Scatterplot of Red and Blue color value percentage

(b) Plot of Density to Relative color variance

Figure 3: Plots of red and blue percentage and Density to relative color variance

3 Results and discussion

The data extracted from the features was split into 3 data sets: Training data, validation data and test data. From the training data, feature selection was performed, and the training data was filtered to only include the features selected. From the extracted features in the training data, multiple Machine Learning classifier algorithms were trained. These were K-Nearest Neighbor, Decision Tree Classifier, Random Forest Classifier, Support Vector Classifier and Logistic Regression Classifier, all with the purpose of classifying a skin lesion as melanoma or benign.

The Machine Learning classifier algorithms were evaluated from their respective accuracy and ROC area under curve accuracy measurements on the validation data, the specificity and sensitivity of the model was also taken into account.

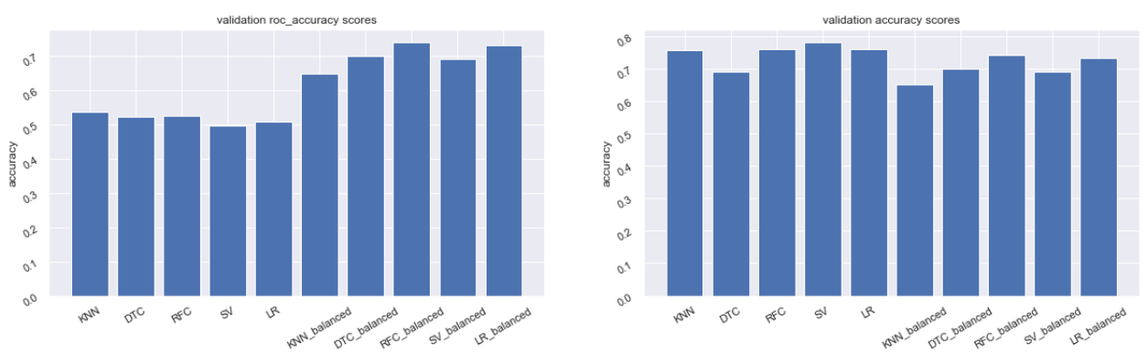


Figure 4: Accuracy and roc auc accuracy for all the trained models

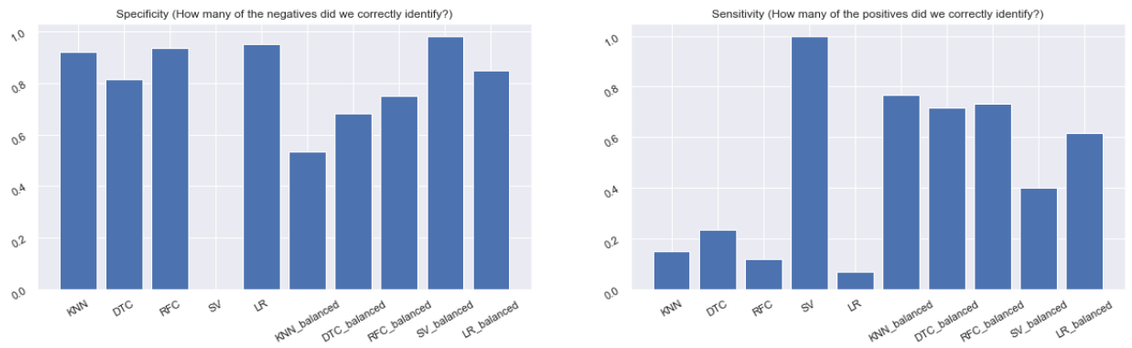


Figure 5: Specificity and sensitivity for all the trained models

The models trained on the non-balanced data were accurate, but performed badly when correctly classifying melanoma lesions. This was partly due to the data being skewed. The ratio between healthy and cancerous lesions are very large, and are causing the model to predict almost every lesion to be healthy. Thus having an high accuracy, but a low sensitivity. To increase the sensitivity of the model, under-sampling was performed, and ensured a 1:1 ratio of healthy to cancerous lesions. This should result in higher sensitivity, though possibly with the cost of lower specificity, which is deemed a fair trade.

With the new balanced data set, we performed a new feature selection and redid all the training on our classifiers. The new models with optimized features and parameters, results as expected in higher sensitivity, but much lower specificity. The balanced models are able to predict 60% of lesions that are melanoma cancerous to be cancerous. On the other hand the balanced models are only able to predict 50% of lesion that are healthy to be healthy. Thereby the success-rate for the trained Machine Learning models is in general too low. The non-balanced models can be really good at prediction either healthy or cancerous lesions, but not both. Therefore we decided to go for the Random Forest Classifier trained model on the balanced data. Here we have the best accuracy measurements on the validation data, and in this case sensitivity is valued higher than specificity.

The low accuracy of the models could be caused by the way the features was extracted from the images. It is apparent that the symmetry and color variance of a lesion, can be calculated easily without any reference measurement unit such as cm. For border variation, area and perimeter it is apparent that these can not reliably be calculated without a reference measurement unit. Due to their dependence on the resolution of the picture, these measurements are calculated on the basis of pixels. The total number of pixels in the lesion will therefore change due to multiple factors such as different cameras and the distance between camera and lesion.

3.1 Open Question

To further explore the classification of lesions based on characteristics, we went on to use the developed Machine Learning model on images of lesions diagnosed with seborrheic keratosis. By first removing all the cases with seborrheic keratosis, we create a world in which only two types of lesions exist: healthy and cancerous melanoma. In reality, people get both melanoma and keratosis, and they should somehow be warned whether their lesion is dangerous or not. What does the machine learning model designed to predict melanoma, predict when fed images of seborrheic keratosis?

To investigate this we created a new data frame only containing the lesions diagnosed with keratosis and recalculated all of our features. By then running our keratosis data set in the balanced Random Forest Classifier model, we found that it predicted 60% of the keratosis lesions to be melanoma. Thereby classifying it as cancerous, even though keratosis is neither healthy nor cancerous. A keratosis lesion can go without treatment, thus classifying keratosis to be healthy would be correct.

However, what we are designing here is software for an app. The purpose of this app is not to identify cancer with 100% accuracy, but more as a tool to increase focus on the potential dangers in skin lesions. And wrongly labeling a keratosis lesion as being cancer, will not harm anyone. Since it is best to get treatment for keratosis, some people might visit a doctor because of a suspicion of cancer. The doctor will then diagnose the lesion as being keratosis, and the person will get the right treatment. This is not saying that everyone should go and get checked for every disease. However, we are dealing with cancer and similar skin lesions, and awareness about the different types could save lives.

4 Limitations

First and foremost, more data would be required to get improved results, as more data provides more information and therefore makes estimation based hereof more precise. Some of the data was excluded from analysis to get an equal amount of healthy and cancerous lesions. We went from having 1372 healthy lesions and 374 melanoma lesions, to having 374 lesions of each class.

Another limitation is that we have only worked with pictures of white people. Thus the accuracy of the models will probably decrease when trying to predict on people with different skin color.

One big limitation is with the provided images, as there seem to be no common setting. It is not possible for us to know the actual size of the lesions, since there are no scalar present in all the pictures. The lighting between the pictures vary as well. Some pictures seem to have a slight blue tint, which can cause confusion when measuring the color features. These two factors could be solved with an uniform indicator of size and lighting.

Lastly, the characteristic E in ABCDE for Evolution is out of reach for measurement. As there is only one picture per lesion, there is no way to compare a lesion at one point in time to another. Therefore making it impossible to determine growth - which can be a great indicator of melanoma.

For this project only simple classifiers were taken in use. Further exploration of the possibility of visual lesion diagnosis could be done using a Neural Network and some Deep Learning.

5 Conclusion

Given the provided data, this report shows a slight, yet not ideal possibility of diagnosing skin lesions through imaging. By measuring the characteristics of the skin lesions, some models were successful in identifying non-cancerous lesions, while others did a decent job at identifying cancerous lesions. No model was good at both. The chosen model was Random Forest Classifier on the balanced data. This model was able to diagnose some of the lesions correctly but had a tendency to diagnose healthy over melanoma. Especially diagnosing keratosis skin lesions based on the classifiers for melanoma is not a reliable method.

Diagnosing melanoma skin lesions using a smartphone app should not replace consulting a dermatologist. An app would however be a good addition to raise awareness of tracking one's skin lesions, but should not be instead of a certified doctor.

6 Closure statement

In the week leading up to the submission day, the ITU GitHub enterprise seemed to be down for the whole group. A new repository was therefore created using private accounts, causing the gitlog to be split into two.