Project 3: Image Analysis

Veronika Cheplygina

No Institute Given

1 Overview: image analysis

In this project you will learn to measure features in images of skin lesions, and predict the diagnosis (for example, melanoma) from these features in an automatic way. You will likely:

- Implement methods to measure "handcrafted" features
- Predict the lesion diagnosis using simple classifiers, based on the features
- Perform experiments to test different parts of your method

2 Requirements

Similar to the last projects, you must use Python and work in Github. All Python packages are allowed, but comment your code. You can find starter code and data here:

`https://github.com/vcheplygina/fyp2021p3`

3 Assignment

A dermatologist in your city asks if you can investigate whether some characteristics of skin lesions can be reliably measured with a smartphone app. The characteristics the dermatologist is especially interested in are: asymmetry, border and color.

Your goal is first to measure two or more of these characteristics in a set of skin lesion images. You can choose which one(s) you want to implement, and you may also implement other features, if you want.

Then you should try to assess how good the measurements are, for example, by predicting the diagnosis of the skin lesions based on these features. You can focus on prediction of the following class:

- Groups 1 to 8: melanoma
- Groups 9 to 16: keratosis

4 Task 0: explore the data

Go through the data (csv file, images, segmentations) that you have available to understand what's available to you, and write a brief description. Decide if this data is sufficient, or if cleaning is needed. For example, what do you do with the images that are malignant (cancer), but not of the class you want to focus on? Are there images of low quality? Etc. You are allowed to search for and add other public dataset, to this set of images.

5 Task 1: implement two features

Choose one of the ABC features and implement a function to measure it for one image. While you are doing this, you might want to create "toy" images where you already know the results, for example a circle should be less asymmetric than an ellipse, etc.

Once you are satisfied with your implementations, run them on all your images, and examine the feature distributions for each class, for example using scatter plots. Do you see differences between the classes?

6 Task 2: predict the diagnosis

Split your data so that you are have training data and hold-out test data. Use the training data to train different classifiers and investigate their parameters. Once you made a choice, evaluate your classifier on the hold-out test data. Think of different metrics you can use, and different ways to present your results.

7 Task 3: open question

Use the data to formulate, motivate, answer, and discuss another research question of your choice. You may use other data or features here, that we did not cover.

8 Hand-in

The rules are almost the same as with the other projects, except the numbers of figures.

You must hand in:

- gitlog.txt: Your repo's git log, e.g. by running: git log ¿ gitlog.txt.
- code.zip: One zip file containing one .py scripts of your commented code that runs fully without errors using the three raw data files, and reproduces your findings. Do not include the raw data files here. If your code is making use of external data sets or .py scripts, you can include them (small files) or set up your script to download them from an external storage (big files).

– report.pdf: A project report.

The project report must be between 3 and 5 pages long including figures (with 11 pt font size and about 1.5 cm margins), and should consist of the following sections:

– 1. Introduction: Here you provide the context and motivation for the problem. What are your research questions, and why does your research provide value?
– 2. Data: Here you describe all your data tables/sets, and briefly summarize how you obtained and cleaned/transformed them..
– 3. Results and discussion: Here you provide the technical results and a discussion of your findings over the data.
– 4. Limitations: Here you give an account on the major short-coming(s) of your methodology/data.
– 5. Concluding remarks and future work: Here you provide a couple of sentences summarizing the results of the project, why your report is relevant for your country's health, and indicate how the methods, data, and analysis could be improved or extended.
– 6. Disclosure statement (optional): Here you may state if there were any serious unequal workloads among group members

Given the nature of the image data, there will be no limit on the number of figures/tables, however do take care in deciding what you want to illustrate and how.

References