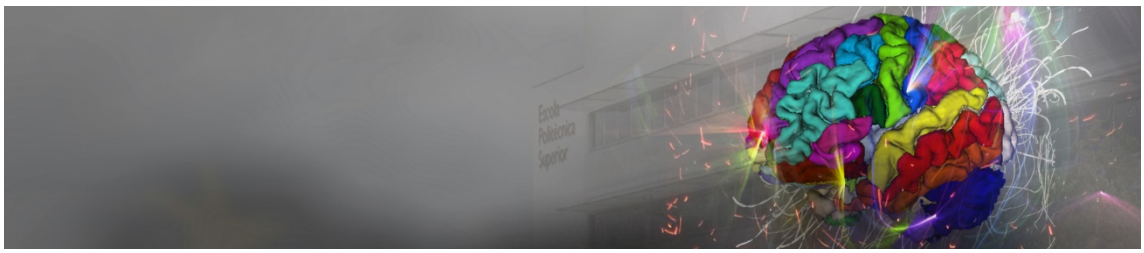
Computer Aided Diagnosis

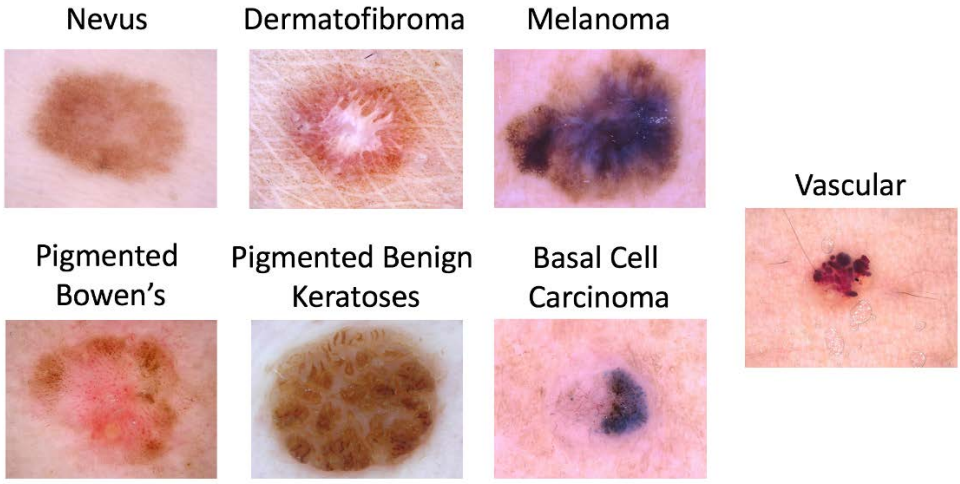
The First Project:
October is a classical month



Introduction

In this coursework, the goal is to develop a computer aided algorithm for the diagnosis in dermoscopic images. This final coursework is a mini-project where you can choose the strategy to solve the problem.

The input data are dermoscopic images in JPEG format. The distribution of disease states represents a modified real world setting whereby there are more benign lesions than malignant lesions, but an over-representation of malignancies. The lesion images come from the HAM10000 Dataset (ViDIR Group, Medical University of Vienna), the BCN_20000 Dataset (Hospital Clínic de Barcelona) and the MSK Dataset (ISBI 2017), hence images were acquired with a variety of dermatoscope types and from different anatomic sites. Images were acquired from a historical sample of patients that presented for skin cancer screening from several different institutions.



We divided this project into two main challenges with a similar structure. Three different sets of images will be given: training, validation, and testing. For the training and validation images, you will have the corresponding ground truth, while for the testing set it will not be available. We will use the results in this testing set to independently evaluate your algorithms (this dataset will be given during the last week of the coursework).

For analysing and reporting the results of your approaches, you should use the validation set and the confusion matrix as a result. Note that, in this part, the validation images cannot be used for training. When evaluating the testing set, you can merge all the information (if you want).

As a suggestion, start with a simple approach and increase its complexity smoothly. You can use software already developed during all your MAIA studies or even from external sources (in this case, cite it properly). Report all the trials and experiments, analysing the improvement on each step.

- **Challenge 1**: The binary problem of classifying Nevus images vs all the others. We will give you more than 15000 images, being half of them nevus and the other half a combination of abnormal areas to train the system. The test set will be open the last week of the project.
- **Challenge 2**: A three-class problem consisting on the classification of cancers: melanoma vs basal cell carcinoma vs squamous cell carcinoma. The training set consists on more than 5000 images, being 50% melanoma, 40% basal cell carcinoma and only 10% squamous cell carcinoma (imbalanced problem).

Objectives

- **A** Information search. Teamwork.
- **B** To design, analyse and implement approaches for automatic diagnosis, i.e., to classify a given region of interest.
- **C** To test implementations at least with the provided images. To evaluate results using provided ground truth. To study problems and possible improvements. To submit a final result for the testing set (unknown ground-truth).
- **D** Documentation.

Coursework

- **A** Powerpoint (7 min) with the following sections:
  - **1** Proposal analysis.
  - **2** Design and implementation of the proposed solutions.
  - **3** Experimental section and results analysis (qualitative/ quantitative analysis, speed, etc). Results should be provided for the validation set.
  - **4** Conclusions.
- **B** Code with comments.
- **C** The diagnosis of the images with unknown ground-truth (.csv file). It should be a column vector with 0 and 1. Use 0 for nevus and 1 for others.

Coursework evaluation:

- **A** Work done during the lab sessions.
- **B** Presentation done after the coursework.
- **C** Results of the unknown dataset will be also considered.

DEADLINE: It will be the one indicated in the moodle submission link. Late submission will be penalised.