Project: LUNG CT SCANS PREPROCESSING
Students:
Henrique Caetano Anraki – 8643412
Ítalo Tobler Silva – 8551910
Marcus Vinícius dos Santos Araújo – 9005871

Contextualization

Ancillary methods for diagnosis are widely discussed, given their potential to improve the discovery and treatment of diseases. Although the replacement of the specialist is not feasible, it is necessary to explore techniques with better accuracy, since the specialist's perception contributes to the final diagnosis.

Especially in this application, the dataset is composed of 3258 regions of interest (ROIS, located in the 'folder' data), with 64x64 pixels extracted by specialists. Among the classes we have Healthy, Consolidation, Emphysema, Thickening, Honeycombing and Ground Glass.

Our proposal consists of a preprocessing module for the classification of medical images, using characteristic extraction techniques and image filters.

Methodology

Before even discussing the details of our technique it is necessary to explain that we use K-NN, a classifier that chooses the classified element that is closest to the unknown object to attach a label, together the Euclidean distance. We apply the classifier 10 times for each random selection of training set (10 fold cross validation). Although it is possible to improve the results with the variation of these approaches, we decided that it is not up to the Image Processing proposal, so we maintained the classifier and measure of dissimilarity in all the experiments.

Initially, we used the PCA to analyze the class dispersion, since each ROI is expressed by the color histogram. We noted that the Healthy class is intertwined among the others (Appendix), which explains the confusion that diminishes the quality of the State of the Art classificators.

Diagnostics of the area are determined primarily by color and shape, so we focused on methods that prioritize these two attributes. At first, the focus was only the color histogram, and it had accuracy of approximately 76%. In an attempt to highlight the colors, we applied the logarithmic function to highlight the organic composition of the image, which presented a slight improvement (approximately 2%). The breakpoint was the application of the color histogram with identification of internal or external pixels to the figure that had an accuracy of 86% when combined with the logarithmic function.

In the end our pre-processing module applies a simple filter (log function) to the ROIs and extracts the adapted color histogram. A simple idea with satisfactory results.

Final remarks

1) The training set was drawn from the own data set by lottery. Due to the randomness, we obtained non-addictive results, but lacked on reproduction capability, because no experiment would be the same as the other. For that reason, we ran the module

several times and calculated the average with standard deviation (output of the classification module).

2) Since the Healthy class confuses the classifier, we advise that the method only be applied to classes that represent diseases with the specialist having the final decision. Thus, we increase the assertiveness and decrease the work of the professionals involved.

Appendix

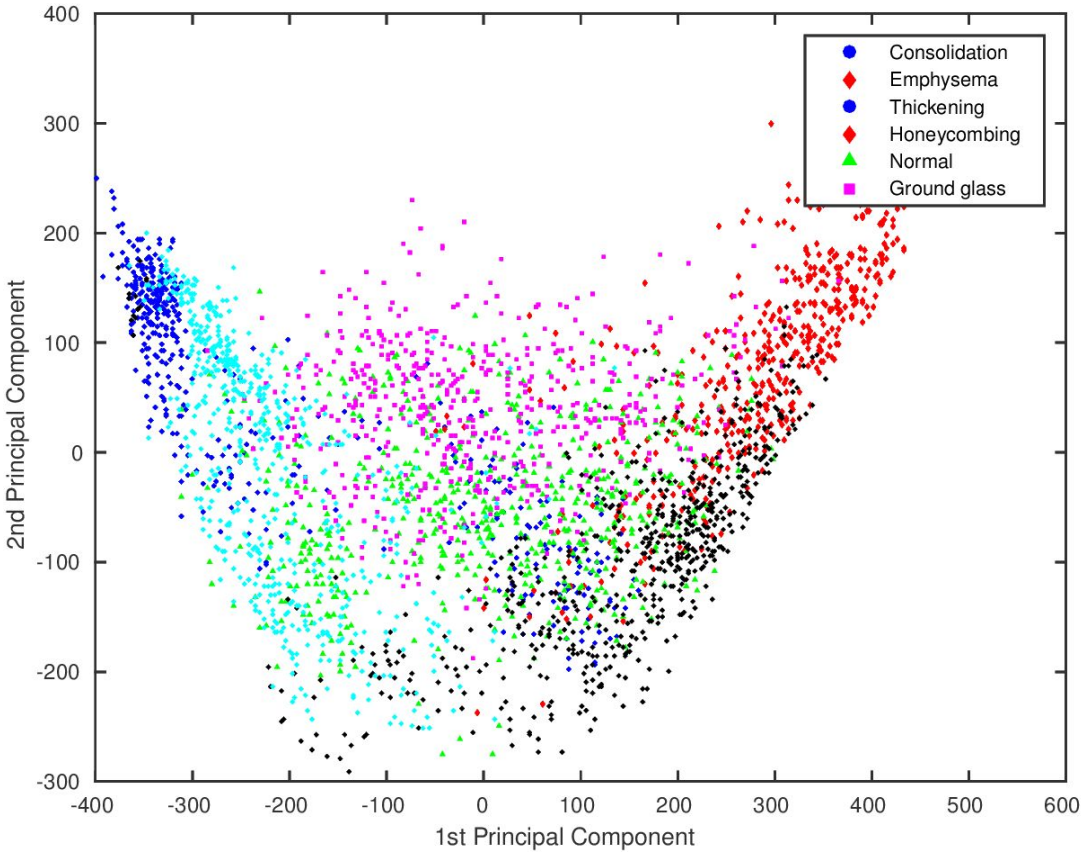

Figure 1: PCA image