UAB
Universitat Autònoma
de Barcelona

# Diagnosis of Pulmonary Lesions in 3D Scans

The goal of this project is to develop and validate a machine learning (ML) system that can accurately characterize the pathology (malignancy) of pulmonary masses (nodules) in computed tomography (CT) scans.

The initial step is to segment the lesions in a volume of interest (VOI). Subsequently, since malignancy of lesions is associated with both texture and shape, we will use classic feature extractors to obtain texture and shape descriptors from the segmented lesions [1]. These descriptors will then serve as input to supervised classifiers to predict malignancy.

The primary objective is to obtain a comprehensive understanding of unsupervised and supervised approaches, and to identify how they can be applied in different scenarios to get the most of data for the task at hand. The project is organized in the following milestones:

1. ***Analysis of Unsupervised Techniques for Lesion Segmentation***. Comparison between Otsu intensity thresholding and k-means over a feature space. Discuss advantages and disadvantages. Influence of pre and post processing. Discuss how to fix parameters. Perform qualitative and quantitative with hypothesis tests analysis. Propose a strategy for pulmonary lesion segmentation.

2. ***Analysis of Supervised and Unsupervised Techniques for Lesion Classification***. Comparison between different supervised learning and combinations with feature selection/dimensionality reduction techniques. Discuss advantages and disadvantages for the different classification methods (e.g. SVM vs Logistic Regression vs Random Forests), features selection/dimensionality reduction techniques and feature spaces (e.g. Gabor, DoG, etc). Comparison between supervised and unsupervised methods. Evaluate the models in terms of fairness and robustness in order to detect any bias in models by using appropriate evaluation strategies and metrics. Perform a complete qualitative and quantitative with hypothesis tests analysis. Provide explanations for the results of your models taking into account the most important features. Propose a strategy for pulmonary lesion classification.

To achieve the goal of this project, challenges with specific objectives will be presented in the practice classes, promoting a variety of practical methods to be applied. These challenges are designed to help build the project incrementally, providing opportunities for learning and growth. A key aspect is being able to explain how the behavior observed in each component affects the overall performance of the machine learning (ML) system. This understanding can help improve the system and identify areas for optimization.

**Reference**

[1] VAN GRIETHUYSEN, Joost JM, et al. Computational radiomics system to decode the radiographic phenotype. Cancer research, 2017, vol. 77, no 21, p. e104-e107.