MACHINE LEARNING METHODS FOR MICRO-FTIR IMAGING CLASSIFICATION OF TUMORS AND MORE

Denise Maria Zezell*

Center for Lasers and Applications – Instituto de Pesquisas Energeticas e Nucleares IPEN-CNEN

Fourier transform infrared micro-spectroscopy imaging (μ-FTIR) has emerged as one of the important tools for studying and characterizing biological materials. It is a label-free technique, relatively simple, reproducible, non-destructive to the tissue and provides accurate results. The vast amount of data and fundamental information obtained from hyperspectral images may not be readily evident. Classical statistics, through its models (parametric and non-parametric) is not able to support the increasing volume of generated data and its high dimensionality. The multivariate analysis of data presents many advantages to be explored, capable of extracting information from the infrared spectra, which go beyond the one-dimensional space, revealing characteristics or properties in the data collected from the samples. The spectral data analysis pipeline, such as the pre-processing steps and the modeling that the Biophotonics Laboratory at Ipen – Cnen, is using in the analysis of biological tissues will be discussed. Results will be presented for body fluids in the disease diagnosis, as well as thyroid, skin and breast tumors, in particular the expression of estrogen and progesterone receptors through tumor biopsies of human cell lines inoculated in mice. μ-FTIR images were collected from histological sections, and six machine learning models were applied and evaluated. The Xtreme gradient boost and Linear Discriminant Analysis showed the best accuracy results, indicating that they are potential models for breast cancer classification tasks.

Acknowledments: This work was supported by FAPESP [17/50332-0 and 21/00633-0], CAPES [Finance Code 001, CAPES/PROCAD 88881.068505/2014-01], CNPq [INCT Fotonica- 465763/2014-6, PQ- 309902/2017-7, and SISFOTON 440228/2021-2].