Big data and omics: bioinformatics to support personalized medicine

Luigi CHIRICOSTA - *IRCCS Centro Neurolesi "Bonino-Pulejo"*

Currently methods used by clinicians to diagnose neurodegenerative disorders, such as Alzheimer's disease (AD), are based on healthcare records and neuroimaging techniques. Nevertheless, they are usually too invasive or lack either sensitivity or specificity. The purpose of this work is the realization of a machine learning model able to perform an early diagnosis of AD in minimal invasive way with high accuracy level. In this line, different models were trained on transcriptomics data obtained through an 8 × 60 K microarray from blood samples of 90 AD and 90 non-AD individuals. For each model, the combination of several hyperparameters were tested. Support vector machines resulted the best model achieving an accuracy of 89% as well as a Receiver Operating Characteristic score of 93%. Furthermore, we focus our attention on the biological role of the features used by the model to obtain its scores. Thus, Gene Ontology enrichment analysis were computed using Fisher's Exact text taking in consideration all genes differentially expressed between the non-AD and AD transcriptomic profiles. The "mitochondrial translation" biological process was linked to the features selected by the model suggesting the oxidative stress the main discriminator pattern.