Machine learning to predict coronary artery disease using proteomics biomarkers

Malika Mahoui, PhD.
School of Informatics
Indiana University – Purdue University Indianapolis

Abstract

Coronary artery disease (CAD) is the leading cause of morbidity and mortality in the United States and is greatly exacerbated by metabolic syndrome (MetS). Current techniques to diagnose CAD are invasive, expensive, and their appropriateness varies among physicians. Therefore, they cannot be used as a routine screening test to predict CAD. To assess the severity of the CAD disease, the diagnostic tests determine with various degrees of accuracy the percentage level of phenotypes such as atheroma wall coverage, stenosis, and plaque composition in the coronary arteries. These phenotypes are measured using invasive methods such as IVUS; in comparison to other phenotypes such as the insulin level that do not require such invasive methods, but at the same time, these phenotypes are less accurate for diagnosis purposes. In addition to predicting the CAD disease, there is a need to improve early screening of the disease without having to use invasive methods such as IVUS. The objective of the study described in this poster is to develop an accurate and non-invasive informatics approach to facilitate screening and monitoring of patients with CAD using a combination of plasma proteomics data and the non-invasively generated phenotypes (e.g. insulin level). This study concentrates on using machine-learning approach to predict ranges of values (e.g. low, moderate, high percentage) for the invasively generated phenotypes, with a special focus on atheroma wall coverage. The ranges of values are mapped to different stages in the CAD disease.