Machine Learning Analysis of Proteomics Data for Early Diagnosis

Dmitry DEVETYAROV 1

Computer Learning Research Centre, Royal Holloway, University of London, UK

Abstract. Ovarian cancer (OC) can usually clinically be diagnosed only at the late stages of the disease. We show that the combination of the biomarker antigen CA125 with proteomic mass spectra data and applying newly developed machine learning algorithms, may establish a diagnosis of OC very early.

Keywords. machine learning, proteomic mass spectra, conformal and Mondrian predictors, early diagnosis, ovarian cancer

It has been shown recently that the use of modern mass spectrometry (MS) techniques allow us to identify some potential biomarkers in serum proteome that can improve the diagnostic of OC [1]. We use intensities of MS peaks extracted using matrix assisted laser desorption/ionization time-of-flight MS data acquisition as additional information to the biomarker CA125 and try to make diagnosis predictions in early stages.

This research describes recently developed machine learning techniques called “hedging prediction” [2] and their application to early diagnostic of OC. Conformal predictors allow us to assign confidence to each prediction – the measure indicating how reliable the prediction is. Mondrian predictors help us treat sensitivity and specificity in a different way: e.g., considering a misclassification of an OC sample more serious than misclassification of a healthy sample.

The main advantages of these techniques are that they assess the confidence for each individual patient and provide valid results [2]: for a given significance level $\epsilon > 0$ (the probability of error we are ready to tolerate), these predictions make mistakes with frequency at most $\epsilon$ in the long run.

As a result, the analysis demonstrates that mass spectrometry peaks contain statistically significant information for early prediction of OC.


1 Corresponding Author: Dmitry Devetyarov, Computer Learning Research Centre, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, United Kingdom; E-mail: dmitry@cs.rhul.ac.uk.