Improving the detection of acute coronary syndrome using machine learning of blood biomarkers

Khaled Rjoob, MSca, Victoria McGilligan, PhDb, Raymond Bond, PhDc, Steven Watterson PhDd, Melody Chemaly PhDd, Roisin McAlister PhDb, Tiago De Melo Malaquias PhDb, Stephen Leslie, PhDf, Charles Knoery, MScf, Aleeha Iftikhar, MSca, Anne McShane, MScg, Anthony Bjourson, PhDb, Aaron Peace, PhDe.

a Faculty of Computing, Engineering & Built Environment, Ulster University, Northern Ireland, UK.
b Centre for Personalised Medicine/Northern Ireland Centre for Stratified Medicine, Ulster University, Northern Ireland, UK.
c Department of Diabetes & Cardiovascular Science, University of the Highlands and Islands, Centre for Health Science, Inverness, UK.
d Emergency Department, Letterkenny University Hospital, Donegal, Ireland
e Western Health and Social Care Trust, C-TRIC, Ulster University, Northern Ireland, UK

Background: Acute coronary syndrome (ACS) is one of the main causes of death worldwide. The 12-lead electrocardiogram (ECG) is used to help diagnose ACS, along with clinical risk factors (smoking, diabetes mellitus, hypertension, and positive family history of ACS). These methods however are associated with many limitations resulting in variable sensitivity/specificity. There is a clear need for the development of new methods to improve ACS diagnosis. The aim of this study was to use a machine learning approach to investigate an optimum panel of blood protein biomarkers capable of independently diagnosing ACS.

Methods: A hybrid feature selection (backward algorithm and filter method) and ML algorithms including two classifiers: 1) decision tree (DT) and 2) logistic regression were applied to protein biomarker data (367 proteins) collected from patients with ACS n=91 or Stable Angina n=97.

Results: Using this approach, 20 specific proteins out of 367 proteins were able to accurately distinguish between ACS and stable angina patients using logistic regression (ROC-AUC=0.8 with confidence interval (CI)[0.69,0.9], accuracy=82.5% with CI [0.72,0.92]) and a DT (ROC-AUC=0.6 with CI[0.47,0.72], accuracy=64.9% with CI[0.52,0.77]) using 5-fold cross-validation. The other 347 proteins were excluded as they did not improve the performance.

Conclusion: Logistic regression obtained a high performance compared to DT and showed promising results uncovering a panel of 20 protein biomarkers (included those associated with progressive atherosclerotic plaques, myocardial injury and inflammation) able to discriminate between patients with ACS and stable angina. Method validation is now required utilising other ML algorithms and new feature selection methods to confirm performance.