28th ECCMID EUROPEAN CONGRESS OF CLINICAL MICROBIOLOGY AND INFECTIOUS DISEASES
Madrid, Spain
21 – 24 April 2018

P1941 Supervised machine learning for the prediction of bacteraemia using routinely collected blood science data

Timothy Rawson*[1], Bernard Hernandez[1], Oliver Blandy[1], Luke Moore[1 2], Pau Herrero[1], Christofer Toumazou[1], Shiranee Sriskandan[1], Pantelis Georgiou[1], Alison H. Holmes[1]

[1]Imperial College London, [2]Chelsea & Westminster Hospital NHS Trust

Background: Supervised machine learning (SML) offers a method of combining computer algorithms with statistics to facilitate enhanced learning. Free of the limitations of human thinking, machines can rapidly identify patterns buried within data, augmenting the utility of routinely collected data to inform clinical practice. We evaluated a SML using routinely collected pathology data to predict the likelihood of blood stream infection (BSI) at initial presentation with infection.

Materials/methods: A SML tool was trained against 205,958 patient profiles containing 6 common haematology/biochemistry markers linked to either microbiology or no microbiology data. These variables were identified through review of the literature and their availability at the time of presentation. Ten-fold cross validation was performed. Data from a cohort study of patients with confirmed *E.coli* BSI were collected and matched to control patients presenting with other medical conditions. The SML tool was used to predict BSI in the test case using variables available on presentation. Area-under the receiver operator characteristics (ROC) curve was calculated and the distribution of likelihood predictions analysed using SPSS 22.0 software.

Results: Cross validation demonstrated an ability for the SML tool to predict positive blood culture with ROC=0.89. The sensitivity was 85% and specificity 92%. The test cohort contained 76 patients with confirmed *E.coli* BSI and 20 control patients who presented with no infection (both inflammatory and non-inflammatory). The median age (range) of patients in the *E.coli* BSI cohort was 66(19-94) years of age. The majority of patients were male (39/76;51%). For the non-infection cases, median age was 63 (23-89) years**,** the majority were male (11/20;55%)**,** and diagnosis was broad with 14/20 (70%) medical, 5/20 (25%) surgical, and 1/20 (5%) palliative patients. Distribution of probabilities for infected vs. non-infected cohorts was significantly different ($p<0.01$) with median (IQR) probabilities of 1.000 (0.996-1.000) for infected individuals vs. 0.381 (0.049-0.519) in non-infected. The ROC for prediction of BSI was 0.74 in this cohort.

Conclusions: SML may improve our ability to predict BSI at the time of presentation using a small number of routinely available variables. The performance of this tool for enhancing decision making during infection management is now under evaluation.