

**P2439 Use of machine learning algorithms to optimise the empirical coverage of Gram-negative multidrug-resistant infection in high risk haematological neutropenic patients**

Carolina Garcia Vidal\*<sup>1</sup>, Pedro Puerta<sup>1</sup>, Celia Cardozo<sup>1</sup>, Miquel Angel Orellana<sup>2</sup>, Gaston Besanson<sup>3</sup>, Jaime Lagunas<sup>3</sup>, Matias Rivero<sup>3</sup>, Anna Costa<sup>3</sup>, Sandra Orozco Martin<sup>3</sup>, Francesc Marco Reverte<sup>4</sup>, Ana Del Río<sup>1</sup>, José Antonio Martínez Martínez<sup>1</sup>, José Mensa<sup>1</sup>, Laura Cozma<sup>3</sup>, Kinga Ritter<sup>3</sup>, Almudena Barreriro-Carrillo<sup>3</sup>, Montserrat Rovira<sup>5</sup>, Jordi Esteve<sup>5</sup>, Alex Soriano<sup>1</sup>

<sup>1</sup> Infectious Diseases Department, Hospital Clínic, Barcelona, <sup>2</sup> Informatics Department, Hospital Clínic, Barcelona, <sup>3</sup> Accenture Advanced Analytics – SCOA Innovation Center; Accenture Digital, Barcelona, <sup>4</sup> Microbiology Department, Hospital Clínic, Barcelona, <sup>5</sup> Hematology Department, Hospital Clínic, Barcelona

**Background:** This study was undertaken to improve risk prediction models using supervised machine learning models (ML) that optimize empirical coverage of gram-negative multidrug-resistant (MDR-GNB) infection in high risk hematologic neutropenic patients (HRFN).

**Materials/methods:** Almost 7 million structured data records from patients with HRFN were extracted directly from electronic health records (EHRs) at a tertiary hospital in Barcelona, spanning January 2008 to December 2017 (110 SCT and 60 acute leukemia intensive chemotherapies per year). These data suffered a full transformation process to generate a list of predictors that are used to train and evaluate several supervised machine learning algorithms (Random Forest, Gradient Boosting Machine, XGBoost and Logistic Regression). All methods are validated by assessing their properties out-of-sample (cross-validation). The episode is tagged as event, when predicted likelihoods are above 50%. The algorithms were applied for model development on 70% of the data set and evaluated on the remaining 30% test set.

**Results:** 3235 episodes of hematologic patients with HRFN were documented (median age: 57 [IQR 44-67] years; 56.9% males, 38% acute leukemia and 28% HSCT). Infections caused by MDR-GNB accounted for 180 (5.6%) episodes. The most frequent MDR-GNBs were MDR-*P. Aeruginosa* (53%) and ESBL Enterobacteriaceae (46%). *Table 1* presents the results of the evaluation of the different ML models used to predict which patients will need empirical antibiotic treatment that should include antibiotic coverage for MDR-GNB (cutoff = 0.5). Calibration was excellent. There is no significant difference in the results from the four models.

**Conclusions:** Prediction of patients with HRFN who require broad spectrum antibiotic for MDR-GNB infections should be done with ML algorithms. The ML algorithms can be used in production to analyze real time data from EHRs and to provide antibiotic recommendations (through an artificial intelligence process) that will help clinicians take better decisions.

**Table 1. ML models to predict the need of MDR-GNB coverage in HRFN patients.**

<b>Models</b>	<b>AUC</b>	<b>F1_Score</b>	<b>Sensitivity</b>	<b>Specificity</b>	<b>Negative Predictive Value</b>	<b>Positive Predictive Value</b>
<b>GBM</b>	0.7872	0.9705	0.4583	0.9988	0.9438	0.9778
<b>XGBoost</b>	0.7945	0.9670	0.4895	0.9886	0.9464	0.8246
<b>Random Forest</b>	0.7896	0.9711	0.4583	1.00	0.9439	1.0
<b>GLM</b>	0.7827	0.9716	0.4687	1.00	0.9449	1.0

29<sup>TH</sup> ECCMID  
13-16 APRIL 2019 AMSTERDAM, NETHERLANDS  
POWERED BY M-ANAGE.COM

