## An Accurate Data Preparation Approach for the Prediction of Mortality in ACLF Patients Using the CANONIC Dataset

M.S. Garcia, Gordon Doyle, Gabriele Ranco, Vicente Arroyo, Marco Pavesi, Elizabet Garcia, Faouzi Saliba, Rafael Banares, Javier Fernandez, Banwari Agarwal, Rajeshwar P. Mookerjee and Rajiv Jalan



The incidence of chronic liver disease has increased in Europe and can lead to Acute on Chronic Liver Failure (ACLF) which is associated with high levels of mortality due to multisystem organ failure. The characteristics of the ACLF patients can change very rapidly within a short period of time. Continuous assessment of their recovery status is critical for clinicians to adjust and deliver effective treatment. The aim of this paper is to validate the usefulness of a data preparation approach by combining different criteria to replace missing values, balance target-class variables, select useful patient characteristics and optimise hyperparameters of machine learning models for the prediction of ACLF associated mortality rates. A key step in the data preparation is a feature selection Mutual Information (MI) based multivariate approach to build smaller, and yet equally and in some cases more informative, subsets of patient characteristics than those frequently proposed for the prediction of mortality, from patients with ACLF in the CANONIC dataset. The usefulness of the data preparation approach proposed to predict mortality was evaluated by training the XGBoost and Logistic Regression models with the prepared data. Evaluations of the models trained using a test set provided evidence of an overall high accuracy in the prediction of the mortality rates of patients for days after their diagnosis, and in some cases even higher when reduced and more informative subsets of patient characteristics were found.



Sum number\_of\_characteristics\_used



0.136

0.191

0.02

3.8E-4

0.012

7E-04

4E-05

P-Inter-Conf