Abstract 66 – Paper ID: 119

Developing a Machine Learning-Based Cardiometabolic Disease Model for Predicting Liver Disease

Miranda Moirangthem[1], Hillul Chutia[2], Nagamani Selvaraman[2,3], Romi Wahengbam[1,3]

[1]Biological Sciences and Technology Division, CSIR–North-East Institute of Science and Technology, Jorhat, Assam–785006, India

[2]Advanced Computation and Data Sciences Division, CSIR–North-East Institute of Science and Technology, Jorhat, Assam–785006, India

[3]Faculty of Biological Sciences, Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, Uttar Pradesh–201002, India

Email: mirandarangthem@gmail.com

Abstract

Cardiometabolic diseases, which are a leading cause of global mortality, are interconnected metabolic and cardiovascular disorders that include diabetes, MASLD and ischemic heart diseases. Predicting disease may help in its early diagnosis and treatment. Cohort studies are crucial in cardiometabolic disease research as it can give significant insight into disease demographics, prevalence and its prediction. Here, we utilise the data of a national longitudinal cohort study to investigate and predict liver disease. Clinical and anthropometric data of Phenome India Cohort ($n = 207$) were analysed and divided into subgroups based on the status of hepatic steatosis and fibrosis. Sixteen key metadata, including liver enzyme, renal, FibroScan and anthropometric parameters were used for initial model development, and eight parameters were identified using forward and recursive feature selection. Seven machine learning (ML) algorithms, namely Random Forest, XGBoost, CatBoost, SVM, Logistic Regression, Naïve Bayes, and Neural Network, were trained on the new parameters, and data was split into training (75%) and testing (25%) sets. Models using all 16 features tended to overfit, achieving perfect performance on the training set but lower generalisation on the testing set. Feature reduction to eight resulted in a simpler model with similar performance. SVM provided the most desirable test performance among the seven algorithms achieving balance between sensitivity and specificity (accuracy 0.738, sensitivity 0.857, specificity 0.500, F1-score 0.814, ROC-AUC 0.724; 5-fold cross-validated accuracy 0.710 and ROC-AUC 0.741). Adjusting the decision threshold between 0.55 and 0.80 led to lower sensitivity at lower thresholds and high sensitivity at higher thresholds. The application of ML algorithms to clinical metadata can help in the prediction of liver disease.

Keywords: Cardiometabolic disease, machine learning, cohort study, liver disease, SVM, clinical metadata

ISBN: 978-81-989164-2-6; DOI: 10.21467/abstracts.8.1