New Technologies - AI and Single Cell Sequencing

Machine learning models outperform risk scores in predicting hepatocellular carcinoma in patients with chronic viral hepatitis

Grace Lai-Hung WONG

Hepatocellular carcinoma (HCC) is the deadliest complications of chronic viral hepatitis. Precise HCC risk prediction facilitates appropriate surveillance strategy and reduces cancer mortality. Various novel machine-learning (ML) and artificial intelligence (AI) models are being developed and validated to predict HCC. Our team has the privilege to leverage the terrific Hospital Authority data to carry out a territory-wide study in Hong Kong from year 2000 to 2018 identified based on comprehension clinical data, namely viral markers, diagnosis codes and antiviral treatment for chronic viral hepatitis. The cohort was randomly split into training and validation cohorts in 7:3 ratio. Five state-of-the-art ML methods including logistic regression, ridge regression, AdaBoost, decision tre
e and random forest were compared for the best prediction model. We included 124,006 patients with chronic viral hepatitis and complete data to build the models. In the training cohort (n=86,804, 6,821 HCC), ridge regression (AUROC 0.842), decision tree (0.952) and random forest (0.992) performed the best. In the validation cohort (n=37,202, 2,875 HCC), ridge regression (AUROC 0.844) and random forest (0.837) maintained their accuracies and significantly higher than HCC risk scores: CU-HCC (0.672), GAG-HCC (0.745), REACH-B (0.671), PAGE-B (0.748) and REAL-B (0.712) scores. Low cut-off (0.07) of HCC ridge score (HCC-RS) achieved 90.0% sensitivity and 98.6% negative predictive values (NPV) in the validation cohort. High cut-off (0.15) of HCC-RS achieved high specificity (90.0%) and NPV (95.6%); 31.1% of patients remained indeterminate.

In conclusion, HCC-RS from ridge regression machine-learning model accurately predicts HCC in CVH patients. These machine-learning models may be developed as built-in functional keys or calculators in the electronic health systems to reduce cancer mortality. Our study is one of the most recent examples of how to apply ML and AI in HCC risk prediction. Some other AI models make good use of radiomics and clinical data through the Convolutional Neural Network (CNN)-based modelling to predict HCC risk. ML and AI models would facilitate the prediction of HCC risk, assist but not replace our clinical judgment and decision-making process.