Machine learning model to predict response to short duration direct-acting antiviral therapy for hepatitis C virus infection

Joanne M Carson[1], Sebastiano Barbieri[2], Gregory J Dore[1], Gail V Matthews[1], Marianne Martinello [1]
[1] The Kirby Institute, UNSW, Sydney, Australia; [2] Centre for Big Data Research in Health, UNSW, Sydney, Australia

Introduction

Despite increasing use of pangenotypic direct-acting antivirals (DAA) to treat hepatitis C virus (HCV) in Australia (8-12 week durations), national discontinuation rates continue to rise. However, evidence suggests a significant proportion of people will achieve sustained virological response (SVR; cure) with shorter treatment durations. Shortening treatment could reduce costs, optimize adherence and support treatment expansion particularly among key populations including people who inject drugs, and people in prison. This analysis aimed to develop a machine learning model to predict response to short duration DAA treatment.

Methods

Data from clinical trials evaluating response to short duration DAA therapy (4-6 weeks; REACT, TARGET3D, DARE-C II, STRIVE) and data of individuals discontinuing standard duration DAAs from other clinical trials were collated. Due to a relatively small sample size and treatment failures representing a minority of cases, an adaptive synthetic sampling approach for imbalanced learning (ADASYN) was used to generate synthetic treatment failure samples for training based on their difficulty level (i.e., how hard they were for the model to learn to classify). The methodology computes density distribution within the k nearest neighbors of a given sample. Minority instances that were surrounded by many majority instances (i.e., harder to learn or distinguish) received more synthetic examples. An XGBoost machine learning model was developed using baseline demographic, clinical and laboratory data. Nested cross-validation was undertaken to assess model performance and optimize hyperparameters. A SHapley Additive exPlanations (SHAP) plot was used to visualize the impact of each feature on the model output including magnitude and direction.

Results

Data of 206 individuals receiving short duration DAAs in clinical trials (median age 45 years; 95% male; 71% white; 33% HIV; 44% injecting drug use [IDU] past month) was extracted, of whom 45 (22%) had treatment failure. DAA regimens taken were sofosbuvir-velpatasvir (SOF-VEL; 46%), glecaprevir-pibrentasvir (GLE-PIB; 27%), ombitasvir-paritaprevir+ritonavir+dasabuvir (PrOD; 15%) and sofosbuvir+ ribarvirin (SOF+RBV; 8%). Comparing those with SVR vs. treatment failure, 47% vs. 36% had acute HCV (<6 months), median HCV RNA was 342,505 vs. 3,719,143 IU/mL (5.5 vs 6.6 $\log_{10}$ IU/mL), and median liver stiffness was 5.9 vs 7.1 kPa. Implementation of ADASYN to generate synthetic failure data increased the sample size (n=282; 50% SVR, 50% treatment failure). The overall accuracy of the model was 82.5% (standard deviation [SD] 9.5%). Model specificity for treatment failure was high (90.5%; SD 9.7%), but sensitivity was lower with greater variance (59.4%; SD 26.5%). Performance metrics for the model are displayed in **Table 1.** The area under receiver operating characteristic (AUROC) curve indicating model capacity to distinguish between SVR and treatment failure cases was 85%. The area precision-sensitivity curve reflecting the trade-off between accuracy of predictions and capacity to detect all treatment failures was 65%. Model features and SHAP values are displayed in **Figure 2**. Features that had the greatest impact on the prediction of short duration treatment failure vs. SVR were baseline HCV RNA (higher vs. lower), alanine aminotransferase (ALT; lower vs. higher – potentially due to elevated ALT during acute HCV infection), DAA regimen (SOF+RBV vs. Other DAAs), and Fibroscan score (higher vs. lower).

	Precision, % [SD]	Sensitivity, % [SD]	Specificity, % [SD]	F1 Score, % [SD]
Average	83.9 [9.4]	82.5 [9.5]	82.5 [9.5]	82.1 [9.7]
Treatment failure	63.2 [21.3]	59.4 [26.5]	90.5 [9.7]	57.4 [21.3]
SVR	88.0 [7.8]	90.5 [9.7]	59.4 [26.5]	88.8 [6.5]

Table 1. Performance metrics. Sensitivity is proportion of predicted treatment failures relative to the total number of treatment failures. Specificity is the proportion of predicted treatment failures relative to the total SVR outcomes. Precision (positive predictive value) is the proportion of treatment failure classifications that were true treatment failures. F1-Score is the harmonic mean between sensitivity and precision.

This study used machine learning to predict response to short duration DAA therapy (4-6 weeks) using baseline demographic, clinical and treatment characteristics. The model had high specificity for predicting treatment failure, but suboptimal sensitivity. A larger more diverse sample is needed to improve performance and generalizability.

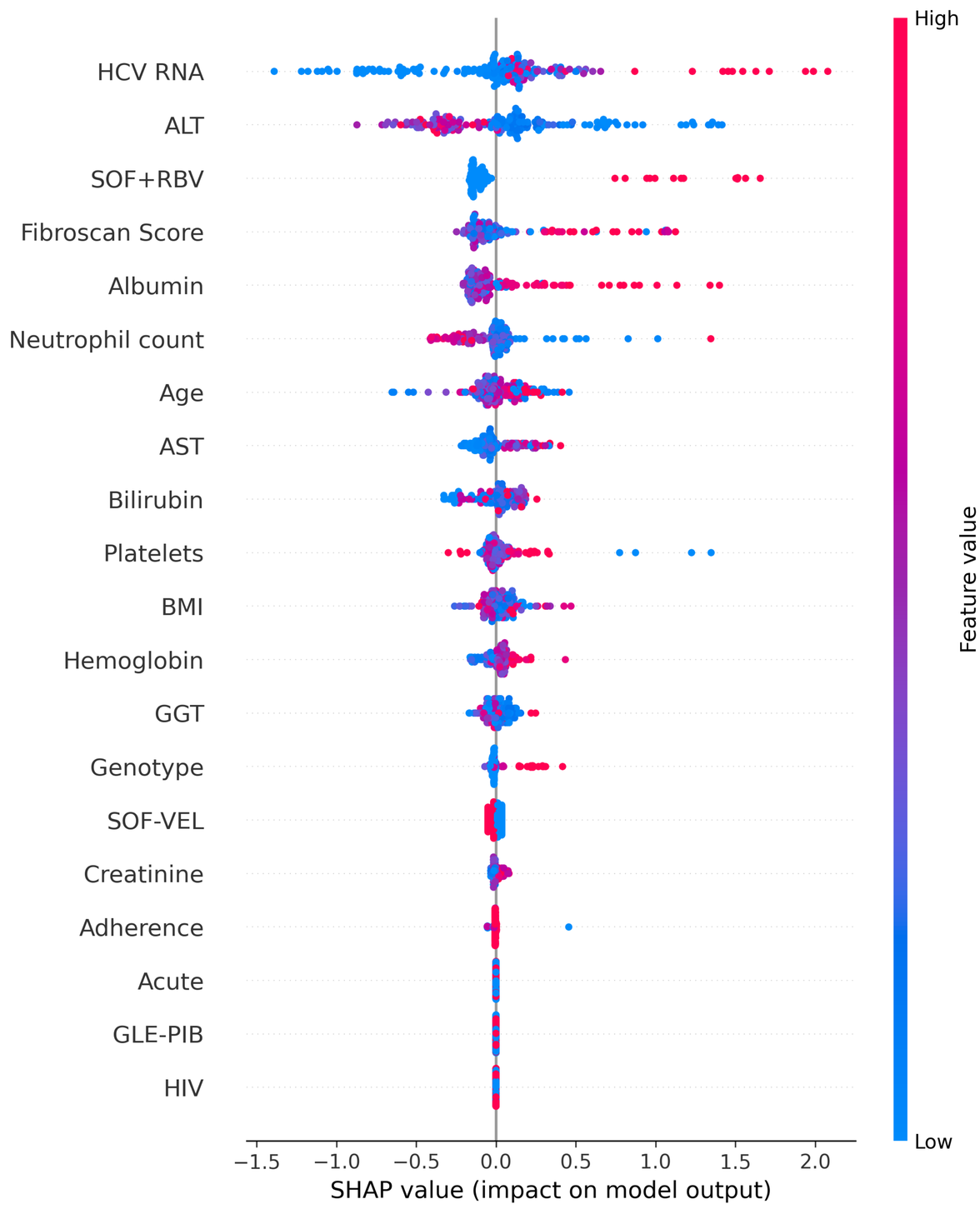

Figure 2. SHAP plot of feature impact. Features as listed in order of importance. Features pushing the prediction higher are shown in red, and those pushing the prediction lower in blue. The SHAP value quantifies the contribution of each feature to the prediction, with features sorted by the sum of SHAP value magnitudes across all samples. **Abbreviations:** SOF+RBV, sofosbuvir+ribavirin GLE-PIB, glecaprevir-pibrentasvir; SOF-VEL sofosbuvir-velpatastir; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; GGT, gamma-glutamyl transpeptidase.

Conclusion

Although limited by sample size, these findings suggest there is potential to identify individuals with high probability of HCV cure with short duration DAAs and that personalized HCV treatment delivery is possible. A larger sample with greater demographic, clinical and treatment diversity is needed to improve predictive capability, externally validate the model, assess clinical utility, and design a clinical prediction tool.

For further information about this study or if you would like to contribute data from short duration trials, please contact Joanne Carson **jcarson@kirby.unsw.edu.au**