

# Assessment of portal hypertension severity using machine learning models in patients with compensated cirrhosis

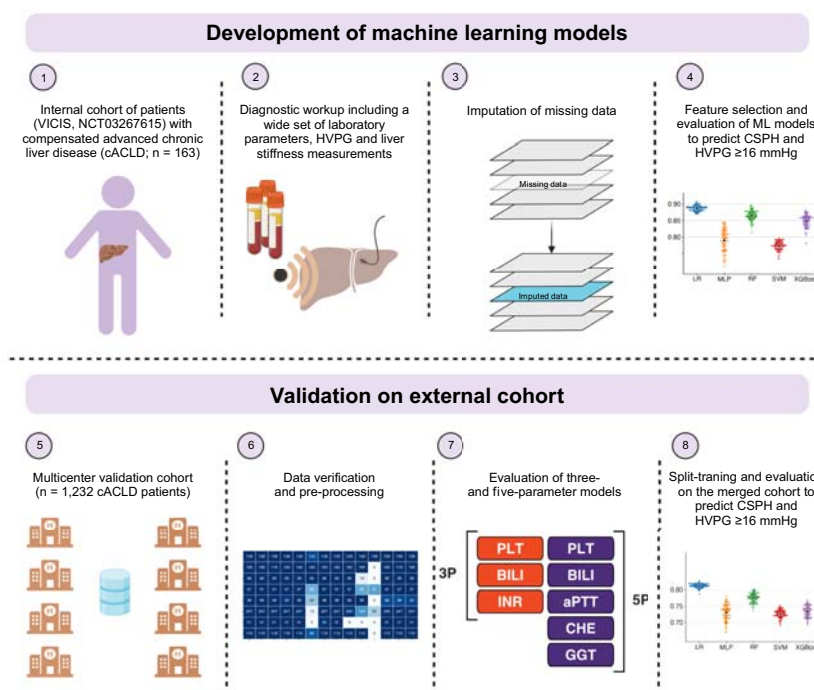
## Authors

Jiří Reiniš, Oleksandr Petrenko, Benedikt Simbrunner, ..., Stefan Kubicek, Mattias Mandorfer, Thomas Reiberger

## Correspondence

thomas.reiberger@meduniwien.ac.at (T. Reiberger).

## Graphical abstract



## Highlights

- Models that can non-invasively assess portal hypertension severity are an unmet clinical need.
- Machine learning models trained on 3/5 laboratory parameters enabled non-invasive assessment of portal hypertension severity.
- These models could predict portal pressures of  $\geq 10$  mmHg or  $\geq 16$  mmHg in individuals with compensated cirrhosis.
- An online tool based on these models has been made available and can be used for portal hypertension risk stratification.

## Impact and implications

We used machine learning models based on widely available laboratory parameters to develop a non-invasive model to predict the severity of portal hypertension in individuals with compensated cirrhosis, who currently require invasive measurement of hepatic venous pressure gradient. We validated our findings in a large multicentre cohort of individuals with advanced chronic liver disease (cACLD) of any cause. Finally, we provide a readily available online calculator, based on 3 (platelet count, bilirubin, international normalised ratio) or 5 (platelet count, bilirubin, activated partial thromboplastin time, gamma-glutamyltransferase, choline-esterase) widely available laboratory parameters, that clinicians can use to predict the likelihood of their patients with cACLD having clinically significant or severe portal hypertension.

# Assessment of portal hypertension severity using machine learning models in patients with compensated cirrhosis

Jiří Reiniš<sup>1,†</sup>, Oleksandr Petrenko<sup>1,2,3,4,†</sup>, Benedikt Simbrunner<sup>1,2,3,4</sup>, Benedikt S. Hofer<sup>1,2,3,4</sup>, Filippo Schepis<sup>5</sup>, Marco Scoppettuolo<sup>5</sup>, Dario Saltini<sup>5</sup>, Federica Indulti<sup>5</sup>, Tomas Guasconi<sup>5</sup>, Agustin Albillos<sup>6</sup>, Luis Téllez<sup>6</sup>, Cándid Villanueva<sup>7</sup>, Anna Brujats<sup>7</sup>, Juan Carlos Garcia-Pagan<sup>8</sup>, Valeria Perez-Campuzano<sup>8</sup>, Virginia Hernández-Gea<sup>8</sup>, Pierre-Emmanuel Rautou<sup>9</sup>, Lucile Moga<sup>9</sup>, Thomas Vanwolleghem<sup>10,11</sup>, Wilhelmus J. Kwanten<sup>10,11</sup>, Sven Francque<sup>10,11</sup>, Jonel Trebicka<sup>12,13,14</sup>, Wenyi Gu<sup>12,13</sup>, Philip G. Ferstl<sup>12,13</sup>, Lise Lotte Gluud<sup>15</sup>, Flemming Bendtsen<sup>15</sup>, Søren Møller<sup>16</sup>, Stefan Kubicek<sup>1</sup>, Mattias Mandorfer<sup>3,4</sup>, Thomas Reiberger<sup>1,2,3,4,\*</sup>

Journal of Hepatology 2023. vol. 78 | 390–400



**Background & Aims:** In individuals with compensated advanced chronic liver disease (cACLD), the severity of portal hypertension (PH) determines the risk of decompensation. Invasive measurement of the hepatic venous pressure gradient (HVPG) is the diagnostic gold standard for PH. We evaluated the utility of machine learning models (MLMs) based on standard laboratory parameters to predict the severity of PH in individuals with cACLD.

**Methods:** A detailed laboratory workup of individuals with cACLD recruited from the Vienna cohort (NCT03267615) was utilised to predict clinically significant portal hypertension (CSPH, *i.e.*, HVPG  $\geq 10$  mmHg) and severe PH (*i.e.*, HVPG  $\geq 16$  mmHg). The MLMs were then evaluated in individual external datasets and optimised in the merged cohort.

**Results:** Among 1,232 participants with cACLD, the prevalence of CSPH/severe PH was similar in the Vienna ( $n = 163$ , 67.4%/35.0%) and validation ( $n = 1,069$ , 70.3%/34.7%) cohorts. The MLMs were based on 3 (3P: platelet count, bilirubin, international normalised ratio) or 5 (5P: +cholinesterase, +gamma-glutamyl transferase, +activated partial thromboplastin time replacing international normalised ratio) laboratory parameters. The MLMs performed robustly in the Vienna cohort. 5P-MLM had the best AUCs for CSPH (0.813) and severe PH (0.887) and compared favourably to liver stiffness measurement (AUC: 0.808). Their performance in external validation datasets was heterogeneous (AUCs: 0.589–0.887). Training on the merged cohort optimised model performance for CSPH (AUCs for 3P and 5P: 0.775 and 0.789, respectively) and severe PH (0.737 and 0.828, respectively).

**Conclusions:** Internally trained MLMs reliably predicted PH severity in the Vienna cACLD cohort but exhibited heterogeneous results on external validation. The proposed 3P/5P online tool can reliably identify individuals with CSPH or severe PH, who are thus at risk of hepatic decompensation.

© 2022 The Authors. Published by Elsevier B.V. on behalf of European Association for the Study of the Liver. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Introduction

Cirrhosis, *i.e.*, advanced chronic liver disease (ACLD), is most often caused by chronic viral hepatitis, alcohol-related liver disease, and non-alcoholic fatty liver disease.<sup>1</sup> Decompensation – which is clinically defined by the occurrence of complications such as variceal bleeding or ascites<sup>2</sup> – is mainly driven by portal hypertension (PH).<sup>3</sup> Invasive measurement of the hepatic venous pressure gradient (HVPG) represents the current gold standard method to assess PH severity.<sup>4,5</sup> In individuals with compensated ACLD (cACLD), decompensation occurs almost exclusively after clinically significant portal hypertension (CSPH, *i.e.*, HVPG  $\geq 10$  mmHg) has developed.<sup>6</sup> However, an HVPG  $\geq 16$  mmHg, *i.e.*, severe PH, is linked to a substantially increased risk of decompensation.<sup>7</sup> In previous studies, this cut-off was shown to be an independent predictor

of the first decompensation.<sup>8,9</sup> Besides, HVPG  $\geq 16$  mmHg has been linked to worse outcomes after extrahepatic surgery in a cohort of compensated individuals, and its use in risk stratification has improved the identification of those at high risk following variceal bleeding episodes.<sup>10,11</sup> Altogether, HVPG  $\geq 16$  mmHg can be used to categorise those at the highest risk of decompensation.

While HVPG measurement establishes the diagnosis of severe PH most reliably, non-invasive methods are needed to identify cACLD at the highest risk of decompensation and for interventional pharmaceutical trials.<sup>12</sup>

The ANTICIPATE multicentre study, including 542 individuals with cACLD, showed that liver stiffness and platelet count are valuable non-invasive parameters to predict the severity of PH.<sup>13</sup> In a convolutional neural network-based

Keywords: hepatic venous pressure gradient; machine learning; non-invasive testing.

Received 13 February 2022; received in revised form 5 September 2022; accepted 12 September 2022; available online 22 September 2022

\* Corresponding author. Address: Division of Gastroenterology and Hepatology, Department of Internal Medicine III, Medical University of Vienna, Währinger Gürtel 18–20, A-1090 Vienna, Austria. Tel.: +43140400-47410, fax: +43140400-47350.

E-mail address: [thomas.reiberger@meduniwien.ac.at](mailto:thomas.reiberger@meduniwien.ac.at) (T. Reiberger).

<sup>†</sup> These authors contributed equally to the study.

<https://doi.org/10.1016/j.jhep.2022.09.012>



study, a machine learning model (MLM) predicted HVPG based on histological features of picrosirius red-stained liver biopsies from 218 individuals with non-alcoholic steatohepatitis.<sup>14</sup> The diagnostic accuracy was further improved by the addition of the enhanced liver fibrosis (ELF) score, platelet count, aspartate aminotransferase, and bilirubin. Still, this MLM approach was based on data from invasive liver biopsy. In another prospective study on 202 participants undergoing liver biopsy and HVPG measurement, artificial neural networks were used as a classification prediction model to identify individuals with cirrhosis, CSPH, and oesophageal varices.<sup>15</sup> Nevertheless, this approach, based on several routine serum markers, did not surpass liver stiffness measurement as a solitary predictor.

In our multicentre study, we assessed the capability of different MLMs – based on non-invasive readouts only – to predict the risk of severe PH and CSPH, defined by HVPG  $\geq 16$  mmHg and  $\geq 10$  mmHg, respectively, in individuals with cACLD of different liver disease aetiologies.

## Patients and methods

### Study design, participant selection, and recorded parameters

In this study, records from individuals who had previously been recruited by corresponding centres were used. The research protocols were approved by respective institutional committees in accordance with the 1975 Helsinki Declaration's ethical principles, and participation required a signed informed consent form.

The internal Vienna cohort included 163 individuals with cACLD recruited between 2017 and 2021 in the prospective Vienna Cirrhosis Study (VICIS, NCT03267615). Specifically, the inclusion criteria included *all* the following criteria: (i) cACLD of any liver disease aetiology defined by HVPG  $\geq 6$  mmHg and (ii) written informed consent from participants to have their records processed for research purposes. In addition, exclusion criteria were any of: (i) previous or current decompensation (ascites, overt hepatic encephalopathy, or variceal bleeding); (ii) invalid or unreliable HVPG measurement; or (iii) hepatic or extrahepatic malignancies. The external validation was performed by including an additional 1,069 individuals with cACLD from international collaborators that run large-scale hepatic haemodynamic laboratories across Europe, categorised together as the external cohort, under the same inclusion and exclusion criteria. Participants from the Vienna cohort were prospectively recruited from VICIS and provided written informed consent. Individuals from the other centres were retrospectively included and did not sign a separate written informed consent for this study.

Individuals with alcohol-related liver disease and viral hepatitis were included regardless of their aetiological status (*i.e.*, abstinent/consuming alcohol, viral suppression/viraemic) at the moment of HVPG measurement. However, this information was utilised for subset analysis to evaluate whether this variable affects model performance.

The HVPG measurement was performed by a respective centre's trained physicians, using local standard operating procedures and according to accepted practices.<sup>5</sup> The vast majority of HVPG measurements represented baseline assessments with participants being beta-blocker-naïve. In those receiving beta-blockers, the medication was paused in

most centres for 48 h (Frankfurt and Paris 24 h) prior to HVPG measurements. Only 28 patients from the Antwerp and Barcelona Hospital Clinic cohorts (2.27% of the merged cohort) underwent HVPG measurements while on beta-blocker therapy.

In addition to the result of the HVPG measurement, participant demographics, disease activity, and clinical and biochemical parameters were recorded for the internal Vienna cohort. The laboratory parameters used for the 5P and 3P models were obtained on the same day of HVPG measurement in 73.5% of participants. In the remaining 26.5% of participants, the median time span between laboratory tests and HVPG was 18 days, with no measurements more than 6 months apart. All 124 clinical, hemodynamic, and biochemical parameters of the training cohort used for feature selection are listed in [Table S1](#).

### Data pre-processing and exploration

Python (version 3.9.6) and R (v4.1.1) were used to carry out the analyses. The pandas library (v1.3.1) was used for data manipulation.<sup>16</sup> Cohort demographics and predicted parameters were evaluated using the Tidyverse package (v1.3.1)<sup>17</sup> in R. Python's scikit-learn library (v0.24.2) was used for scaling, imputations, and model development.<sup>18</sup> Statistical tests were performed using Python's statannot package (version 0.2.3). The xgboost library (v1.4.0) in Python was utilised for training the XGBoost model.<sup>19</sup> Matplotlib (version 3.4.2) and seaborn (version 0.11.1) libraries were used to create the figures.<sup>20</sup>

Parameters with missing entries in more than one-third of the Vienna cohort were excluded from the analysis ([Table S1](#)). The remaining 85 variables had a total of 782 (7.2%) missing values. Recursive feature elimination (RFE) was performed to select optimal parameters for prediction. Prior to RFE, we used the k-nearest neighbours algorithm to impute the missing values, with the number of neighbours set to five. Imputation was performed on standardised parameters, obtained by subtracting the mean and scaling to unit variance. No imputation was applied to the external datasets.

### Development of MLMs

For the development of MLMs, all pre-processed parameters of the Vienna cohort ( $n = 85$ ) were utilised. MLMs were used to detect individuals with HVPG  $\geq 16$  mmHg and HVPG  $\geq 10$  mmHg, as an additional threshold. For the prediction of severe PH as a binary classification task, we introduced "less severe PH" (HVPG  $< 16$  mmHg) and "high-risk" (HVPG  $\geq 16$  mmHg) class labels. A total of five classification models were used: logistic regression (LR); multilayer perceptron (MLP); random forest (RF); support vector machine (SVM); XGBoost.

### Feature selection for MLMs: recursive feature elimination

We used RFE for feature selection in the Vienna cohort to identify three and five optimal variables for the classification prediction model that can be used to identify individuals with HVPG  $\geq 16$  mmHg who are consequently at high risk of decompensation.<sup>21</sup> Our primary analysis incorporated 52 widely available parameters. RFE eliminated the worst ranking parameter per iteration from each scenario's initial set. LR was

## Assessment of portal hypertension severity

used as a model for RFE to score the parameters. In addition, we selected features from a broader set of 85 clinical variables, which also included non-standard laboratory values not considered to be widely available, such as the von Willebrand factor,<sup>22</sup> interleukin-6<sup>23</sup> or the ELF test.<sup>24</sup>

### Robust estimation of classification performance using 5-fold cross-validation

Since prediction performance might vary greatly for different train-validation splits, repeated 5-fold cross-validation (CV) was used for robust estimation of model performance. In each CV, the dataset is randomly split into five subsets (folds). Prediction performance is assessed five times in an iterative manner: one-fold is used for validation, and the four remaining folds to train the model. For each CV, a single AUROC (also referred to as AUC within this frame of reference) was calculated as the mean AUC across the five individual folds. The robust final score of the model is the mean value of AUC means across 100 CVs.

### External model validation

We applied three and five-parameter classification MLMs to the combined external patient datasets for validation. First, we trained MLMs in the Vienna cohort and then we evaluated their predictive performance for HVPG  $\geq 10$  mmHg and HVPG  $\geq 16$  mmHg in external cohorts. Consequently, we performed both training and validation using a pooled (merged internal + external cohorts) dataset, following training-test splits. Considering the unbalanced nature of the datasets, or unequal distribution of participants above and below the HVPG cut-off, we applied re-sampling in a separate analysis in the merged external cohort to address whether balancing the classes could improve predictive power, resulting in an equal number of participants with high-risk and less-severe PH (Fig. S5).

### Extraction of LR coefficients

Final LR models were trained on the entire set of participants for whom the variables were available. Coefficients were extracted and used to develop a publicly available calculator to determine the probability of severe PH. We additionally report the threshold producing the highest Youden's J statistic.<sup>25</sup>

### Prediction of HVPG value with regression

The Lasso regression model was trained on the entire set of participants for direct prediction of numerical HVPG values. Its coefficients were then extracted for the online tool.

## Results

### Patient characteristics

The study population included 1,232 individuals with cACLD combined from the VICIS cohort (hereafter referred to as Vienna;  $n = 163$ ) and the external cohort ( $n = 1,069$ ) from 7 participating sites (Fig. 1A,B). The main aetiologies in the Vienna cohort were viral hepatitis (30%), alcohol-related liver disease (ALD: 23.3%), and non-alcoholic fatty liver disease (NAFLD: 16.6%). The most prevalent aetiologies in the pooled external cohort were similar: viral hepatitis 34.3%, NAFLD 25.2%, and ALD 20% (Fig. 1C, Table 1). Of note, the proportions of individuals with CSPH or severe PH in the Vienna

(67.4%/35.0%) and external (70.3%/34.7%) cohorts were similar. In all but one of the contributing individual centre datasets, the number of individuals with less severe PH (HVPG  $< 16$  mmHg) was higher than the number with high-risk PH (HVPG  $\geq 16$  mmHg). While this means that the dataset was unbalanced, class balance correction for HVPG was not applied for proposed models except when explicitly mentioned, thus reflecting a real-world clinical setting.

The availability of clinical parameters was distinct between external datasets (Fig. 1B; Fig. S1). The external datasets were used to either validate the models with selected parameters or in CVs, in which all datasets were pooled and then divided several times for training and validation.

### Selection of widely available parameters for assessment of PH severity

Using RFE on the Vienna cohort, we identified the most suitable three and five-parameter sets for MLMs (Fig. 2A, 2B) for PH risk prediction. The resulting three-parameter (3P) model consisted of platelet count (PLT), total serum bilirubin (BILI), and international normalised ratio (INR), while the five-parameter (5P) model additionally included cholinesterase and gamma-glutamyltransferase (GGT), with activated partial thromboplastin time (aPTT) replacing INR. RFE-prioritised parameters were among those with the highest Spearman correlation with HVPG (Fig. 2A).

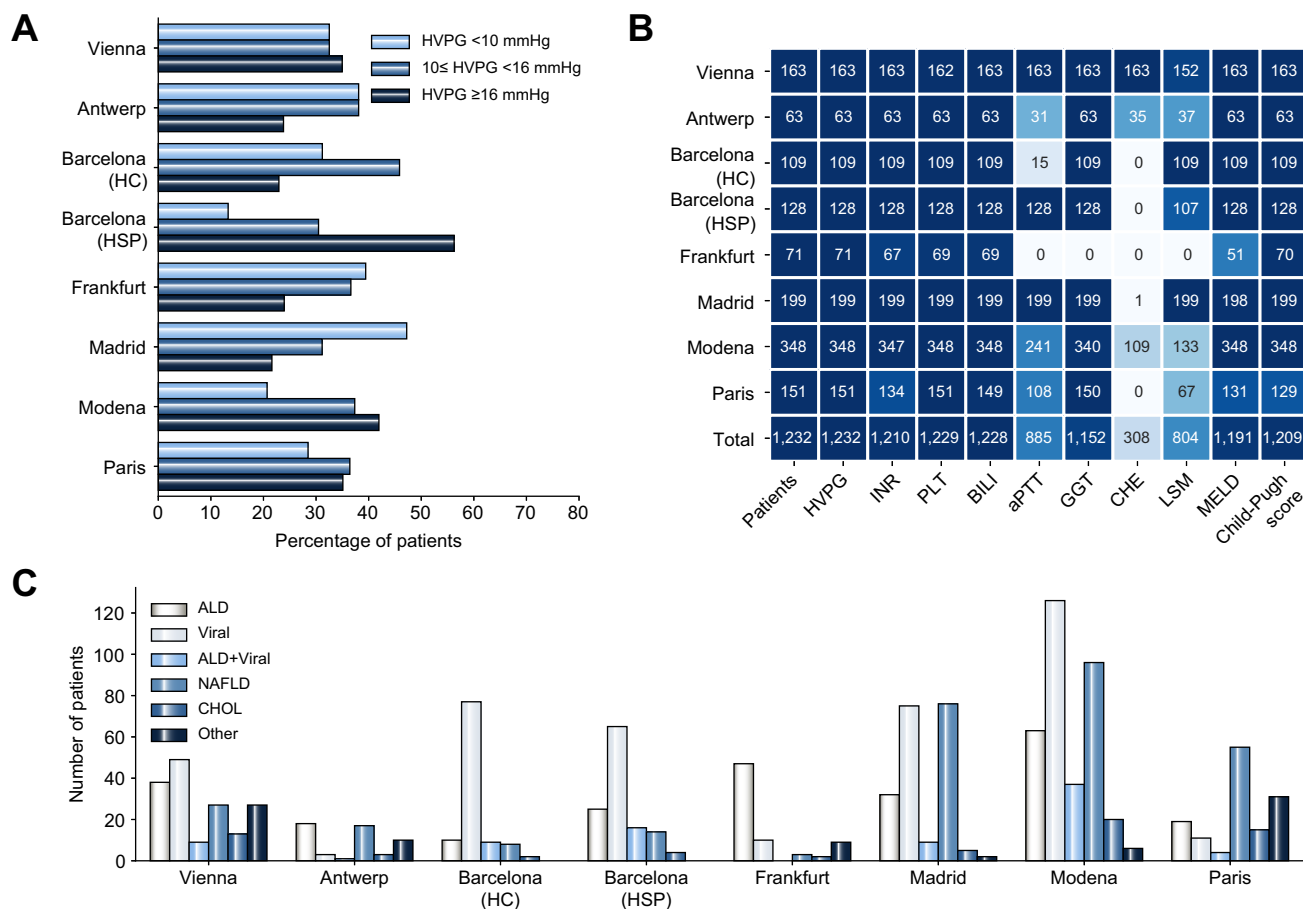
The chosen 5P and 3P parameter sets were then used in multiple MLMs to investigate how well they predicted severe PH and CSPH in internal CVs. The best performing models were LR, RF, XGBoost, SVM, and MLP (Fig. 2C-F). Both 5P and 3P models outperformed liver stiffness measurement (LSM) alone for the prediction of severe PH (Fig. 2G). All 5P and 3P MLMs achieved AUC values above 0.739 for the prediction of severe PH, reaching 0.887 and 0.813, respectively, with LR.

Surprisingly, 5P and 3P outperformed models trained on all parameters (Table S2). Moreover, this performance was superior to employing a broad range of 52 laboratory variables (lowest AUC: 0.663 with MLP; 0.778 with LR) (Table S2, Fig. S2). For CSPH prediction (Fig. 2C-D), LR performed best, with AUC scores of 0.813 (5P) and 0.784 (3P).

These "internal" results support our hypothesis that MLMs derived from widely accessible laboratory markers can accurately predict CSPH and severe PH in individuals with cACLD.

### Performance of the Vienna-trained MLMs in external validation

The 5P and 3P models trained on the Vienna cohort were applied to external cohorts for validation (Table 2). Due to the partial unavailability of cholinesterase, the 5P model could only be applied to two external centres (Antwerp and Modena). For the prediction of CSPH (*i.e.*, of HVPG  $\geq 10$  mmHg), the internally trained 5P model showed reliable performance only in the Modena dataset (LR: AUC = 0.691; Table 2A). The validation of internally trained 3P MLMs for the prediction of CSPH showed more robust performance, reaching AUCs of  $> 0.8$  in three datasets (Table 2B), with the best AUCs observed in the Madrid (LR: AUC = 0.859) and the Barcelona Hospital Sant Pau cohort (LR: AUC = 0.838), which even outperformed the training cohort prediction (LR: AUC = 0.794).



**Fig. 1. Characteristics of patient datasets.** (A) Distribution of individuals without clinically significant portal hypertension (*i.e.*, with HVPG <10 mmHg) vs. with clinically significant portal hypertension (*i.e.*, with HVPG ≥10 mmHg) vs. with severe portal hypertension (*i.e.*, HVPG ≥16 mmHg) across the datasets. (B) Availability of single variables across the different cohort datasets. The blue colour shade corresponds to the percentage of patients within the cohort with available values for each parameter ranging from dark (complete, 100%) to white (absent, 0%). (C) Patient count per aetiology in the respective datasets. ALD, alcohol-related liver disease; aPTT, activated partial thromboplastin time; BILI, serum bilirubin; CHE, cholinesterase; CHOL, cholestatic disease; GGT, gamma-glutamyltransferase; HVPG, hepatic venous pressure gradient; INR, international normalised ratio; LSM, liver stiffness measurement; MELD, model for end-stage liver disease; NAFLD, non-alcoholic fatty liver disease; PLT, platelet count.

When applied to the HVPG ≥16 mmHg threshold, the internally trained 5P model reached an AUC of 0.694 with RF in the Modena dataset and a lower performance of 0.521 with LR in the Antwerp dataset (Table 2C). Although internally trained 3P MLMs reached an AUC of >0.85, the performance was heterogeneous (Antwerp: the best AUC of 0.61 with XGBoost, Barcelona-HSP: 0.6 with LR) (Table 2D). Consistent with the previous threshold, the MLMs performed even better in Madrid than in the original Vienna training cohort (AUC: 0.881 with LR; 0.867 with MLP). Of note, MLP usually performed poorly for the prediction of CSPH but resulted in AUC >0.65 for HVPG ≥16 mmHg in most of the centres.

These findings support that our MLMs, trained on one specific dataset, can then be used for PH risk prediction in new patient cohorts, although the heterogeneity of performance is noted. To develop a more robust approach, we therefore set out to train models on the combined (merged) cohort.

### Evaluation of the overall performance of the MLMs in the merged cACLD cohort

Subsequently, we assessed the prediction performance of both the 5P and 3P models in the merged cohort (combined from all

study datasets) using repeated CVs (Fig. 3). For prediction of CSPH (HVPG ≥10 mmHg, Fig. 3A-C), LR performed better than other models (AUC = 0.773 with 3P, AUC = 0.754 with 5P). For severe PH prediction (HVPG ≥16 mmHg threshold), the LR models again performed best (5P AUC = 0.812, 3P AUC = 0.735), followed by RF in 5P (AUC = 0.776) or MLP in 3P (AUC = 0.726) (Fig. 3D-E).

In the single-dataset resolution, validation results were slightly improved and less heterogenous compared with the internally trained setting. For both CSPH and severe PH prediction, AUC scores were always above 0.625 in all cohorts, except for the Antwerp cohort, where they hovered around 0.5 (Table 3).

The 3P LR model performed similarly to LSM alone to predict CSPH and severe PH in the merged cohort (Fig. 3C,F), and 5P MLMs showed better performance than LSM for HVPG ≥16 mmHg (Table 4). The combination of 3P and LSM reached AUC of 0.858 for CSPH, and 5P+LSM achieved 0.901 for HVPG ≥16 mmHg. Notably, the highest performance was achieved by 3P+LSM, trained on the merged dataset, with AUC of 0.929 in the Modena dataset (Table S3). Both models, when applied with LSM, resulted in better prediction (Table 4, Fig. S3).

Table 1. Patient characteristics.

| Variable                          | Cohort           |                  |                  |
|-----------------------------------|------------------|------------------|------------------|
|                                   | Vienna           | External         | Combined         |
| Participants, n                   | 163              | 1,069            | 1,232            |
| Age, years (median, IQR)          | 59 (51–67)       | 61 (54–69)       | 61 (54–69)       |
| Sex, n (%)                        |                  |                  |                  |
| Female                            | 53 (32.5%)       | 390 (36.5%)      | 443 (36%)        |
| Male                              | 110 (67.5%)      | 679 (63.5%)      | 789 (64%)        |
| Aetiology, n (%)                  |                  |                  |                  |
| ALD                               | 38 (23.3%)       | 214 (20%)        | 252 (20.4%)      |
| Viral hepatitis                   | 49 (30%)         | 367 (34.3%)      | 416 (33.8%)      |
| ALD + viral hepatitis             | 9 (5.5%)         | 76 (7.1%)        | 85 (6.9%)        |
| NAFLD                             | 27 (16.6%)       | 269 (25.2%)      | 296 (24.0%)      |
| Cholestatic                       | 13 (8%)          | 51 (4.8%)        | 64 (5.2%)        |
| Other                             | 27 (16.6%)       | 92 (8.6%)        | 119 (9.7%)       |
| HVPG, median (IQR)                | 12 (8–17.5)      | 13 (9–17)        | 13 (9–17)        |
| VCTE-LSM, kPa median (IQR)        | 22.4 (15.3–35)   | 21.3 (14.9–32.5) | 21.5 (15.0–33.2) |
| HVPG <16 mmHg                     | 17.8 (13.8–26.8) | 18.0 (13.7–25.9) | 17.9 (13.8–26.0) |
| HVPG ≥16 mmHg                     | 38.9 (23.9–59.9) | 32.8 (22.8–46.4) | 33.8 (22.9–47.2) |
| MELD, median (IQR)                | 9 (8–11)         | 9 (7–10)         | 9 (7–10)         |
| HVPG <16 mmHg                     | 8 (7–10)         | 8 (7–9)          | 8 (7–9)          |
| HVPG ≥16 mmHg                     | 10 (9–14)        | 10 (8.5–12)      | 10 (9–12)        |
| Platelet count, G/L, median (IQR) | 104 (78–143)     | 108 (74–152)     | 107 (75–151)     |
| HVPG <16 mmHg                     | 122 (95–167)     | 122 (86–170)     | 134 (87–169)     |
| HVPG ≥16 mmHg                     | 81 (64–104)      | 83 (62–114)      | 83 (62–111)      |
| Bilirubin, mg/dl, median (IQR)    | 0.87 (0.64–1.31) | 0.96 (0.67–1.46) | 0.95 (0.66–1.44) |
| HVPG <16 mmHg                     | 0.76 (0.59–1.03) | 0.86 (0.59–1.29) | 0.84 (0.59–1.25) |
| HVPG ≥16 mmHg                     | 1.22 (0.92–1.86) | 1.16 (0.84–1.78) | 1.17 (0.85–1.79) |
| INR, median (IQR)                 | 1.20 (1.10–1.40) | 1.15 (1.07–1.27) | 1.16 (1.07–1.30) |
| HVPG <16 mmHg                     | 1.20 (1.10–1.30) | 1.11 (1.04–1.22) | 1.12 (1.04–1.23) |
| HVPG ≥16 mmHg                     | 1.40 (1.20–1.50) | 1.23 (1.13–1.35) | 1.25 (1.13–1.39) |
| GGT, U/L median (IQR)             | 102 (53–206)     | 121 (45–145)     | 82 (46–151)      |
| HVPG <16 mmHg                     | 76 (45–168)      | 79.5 (45–145)    | 79 (45–149)      |
| HVPG ≥16 mmHg                     | 132 (83–235)     | 84 (47–146)      | 91.5 (49–156)    |
| aPTT, s, median (IQR)             | 38.0 (35.0–41.2) | 35.0 (31.3–39.0) | 35.4 (32.0–39.5) |
| HVPG <16 mmHg                     | 36.6 (33.5–39.0) | 34.0 (30.5–37.4) | 34.5 (31.0–38.0) |
| HVPG ≥16 mmHg                     | 40.1 (38.3–46.4) | 36.8 (33.5–41.0) | 37.8 (34.0–41.8) |
| Cholinesterase, kU/L median (IQR) | 5.5 (4.2–7.1)    | 5.6 (4.0–7.2)    | 5.5 (4.0–7.1)    |
| HVPG <16 mmHg                     | 6.3 (4.8–7.5)    | 5.9 (4.3–7.6)    | 6.1 (4.6–7.5)    |
| HVPG ≥16 mmHg                     | 4.4 (3.2–5.7)    | 4.5 (3.8–6.3)    | 4.5 (3.5–6.0)    |

The IQR is presented as values Q1 to Q3.

ALD, alcohol-related liver disease; aPTT, activated partial thromboplastin time; GGT, gamma-glutamyl transferase; HVPG, hepatic venous pressure gradient; NAFLD, non-alcoholic fatty liver disease; VCTE-LSM, liver stiffness measurement by vibration-controlled transient elastography.

### Contribution of disease activity and aetiology to model performance

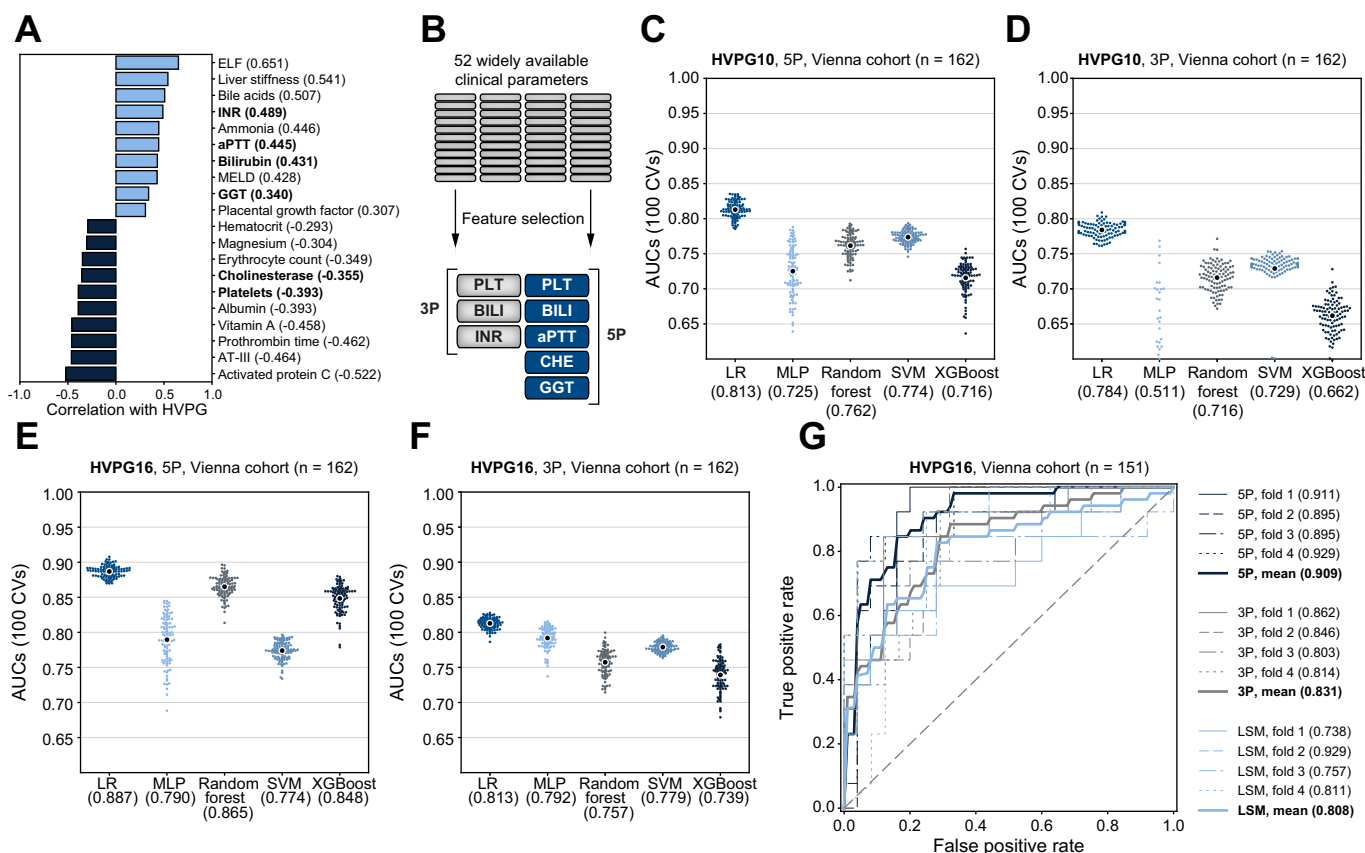
We analysed whether splitting patients according to their disease activity in the merged cohort could improve predictive performance. To this end, we made subsets of participants with ALD (abstinent/consuming alcohol) and viral hepatitis (suppressed/viraemic) (Table S4, Fig. S4). We observed that for the HVPG ≥16 mmHg threshold, the performance of the 3P LR model was in line or better with inactive disease (ALD abstinent AUC = 0.775; viral suppressed AUC = 0.759). The results were consistent and slightly better for the HVPG ≥10 mmHg prediction (Table S4).

Aetiology-wise, the highest performance was achieved in the NAFLD subset (n = 286, AUC = 0.808). The lowest results were observed in the subset with cholestasis, which was relatively small (n = 63, AUC = 0.666) (Table S4). Nevertheless, a re-

analysis of the whole cohort without cholestasis led to only minor improvements (AUC = 0.742 vs. 0.737) (Table S4). In conclusion, the MLMs can be applied to patient datasets without accounting for the underlying aetiology or disease activity.

### Balancing datasets does not significantly improve the predictive power

Unbalanced datasets are often obtained in clinical studies, where a particular selection bias is present. Thus, we explored whether correcting for the HVPG class imbalance might enhance predictive performance. We made a subset with an equalised number of participants with and without severe PH in the merged dataset and re-evaluated the predictive performance of our models (Fig. S5). Nevertheless, this did not improve performance compared with the unbalanced dataset



**Fig. 2. Selection of features for machine learning models and CV on the internal cohort.** (A) Parameters with the highest positive and negative Spearman correlation coefficients with HVPG are shown. (B) Among 52 clinically established laboratory parameters, a final number of three (for the 3P machine learning models) and five variables (for the 5P models) were selected. (C, D) Performance of the 5P and 3P models to discriminate participants with vs. without clinically significant portal hypertension (i.e., HVPG  $\geq 10$  mmHg). For each model, mean AUCs from individual CVs are shown and mean AUC across all CVs is reported. (E, F) Performance of the 5P and 3P models to discriminate individuals with vs. without HVPG  $\geq 16$  mmHg. (G) The performance of 3P and 5P models, compared to liver stiffness measurement in the internal cohort. The AUC is reported across 4-fold CVs. 3P, 3 parameter; 5P, 5 parameter; aPTT, activated partial thromboplastin time; BILI, serum bilirubin; CHE, cholinesterase; CVs, cross-validations; GGT, gamma-glutamyltransferase; HVPG, hepatic venous pressure gradient; INR, international normalised ratio; LSM, liver stiffness measurement; MELD, model for end-stage liver disease; MLP, multilayer perceptron; PLT, platelet count; SVM, support vector machine.

(Fig. 3D-E). As such, we do not consider balancing datasets an essential data pre-processing step for developing MLMs to predict severe PH in individuals with cACLD.

### The formula for prediction of severe PH with LR

Based on coefficients of the 3P LR MLM, the formula for prediction of severe PH for the final 3P model is:

$$HVPG_{\text{probability}} = \sigma(\beta_0 + \beta_1 \times \text{platelet count} + \beta_2 \times \text{serum bilirubin} + \beta_3 \times \text{INR})$$

For the 5P model:

$$HVPG_{\text{probability}} = \sigma(\beta_0 + \beta_1 \times \text{platelet count} + \beta_2 \times \text{serum bilirubin} + \beta_3 \times \text{aPTT} + \beta_4 \times \text{CHE} + \beta_5 \times \text{GGT})$$

where  $\sigma$  is a sigmoid function calculated as  $\sigma(x) = \frac{1}{1+e^{-x}}$ ,  $e$  = Euler's

number, platelet count units are  $10^9/L$ , serum bilirubin units are mg/dl, aPTT units are seconds, GGT is in U/L, and CHE is provided in kU/L.

The extracted coefficients  $\beta_n$  both for HVPG  $\geq 10$  and  $\geq 16$  mmHg predictions are available in the supplementary materials (Table S5). Using them, we developed an online probability calculation tool (available for both HVPG thresholds).<sup>26</sup>

Youden's J statistic defined the optimal cut-off point for the final 3P model as 0.332 (HVPG  $\geq 16$  mmHg) and 0.663 (HVPG  $\geq 10$  mmHg) (Table S6).

### Discussion

In this study, an array of laboratory and instrumental records of 163 individuals with cACLD from the Vienna cohort were initially analysed to identify the most robust parameters for predicting HVPG 16 mmHg, to develop MLMs and to validate them on external cACLD datasets.

**Table 2. Performance of the internally trained machine learning models for the assessment of portal hypertension severity in the different single cohorts.**

| Model  | Datasets |              |                                |                                  |              |              |              |              |
|--|----------|--------------|--------------------------------|----------------------------------|--------------|--------------|--------------|--------------|
|  | Vienna   | Antwerp      | Barcelona -<br>Hospital Clinic | Barcelona -<br>Hospital Sant Pau | Frankfurt    | Madrid       | Modena       | Paris        |
| <b>Performance of the 5P models (PLT, BILI, aPTT, CHE, GGT) trained on the Vienna cohort for prediction of HVPG ≥10 mmHg</b> |          |              |                                |                                  |              |              |              |              |
| Logistic regression  | 0.843    | 0.269        |                                |                                  |              |              | <b>0.691</b> |              |
| MLP  | 0.997    | 0.321        |                                |                                  |              |              | 0.647        |              |
| Random forest  | 1.000    | 0.256        |                                |                                  |              |              | 0.679        |              |
| SVM  | 0.795    | <b>0.474</b> |                                |                                  |              |              | 0.675        |              |
| XGBoost  | 1.000    | 0.218        |                                |                                  |              |              | 0.599        |              |
| <b>Performance of the 3P models (PLT, BILI, INR) trained on the Vienna cohort for prediction of HVPG ≥10 mmHg</b>            |          |              |                                |                                  |              |              |              |              |
| Logistic regression  | 0.794    | 0.441        | 0.743                          | <b>0.838</b>                     | <b>0.618</b> | <b>0.859</b> | 0.696        | <b>0.807</b> |
| MLP  | 0.307    | 0.375        | 0.338                          | 0.204                            | 0.457        | 0.204        | 0.394        | 0.340        |
| Random forest  | 1.000    | 0.521        | 0.729                          | 0.800                            | <b>0.618</b> | 0.842        | 0.668        | 0.779        |
| SVM  | 0.740    | <b>0.560</b> | 0.698                          | 0.785                            | 0.554        | 0.845        | 0.653        | 0.740        |
| XGBoost  | 1.000    | 0.476        | <b>0.744</b>                   | 0.743                            | 0.575        | 0.808        | <b>0.681</b> | 0.792        |
| <b>Performance of the 5P models (PLT, BILI, aPTT, CHE, GGT) trained on the Vienna cohort for prediction of HVPG ≥16 mmHg</b> |          |              |                                |                                  |              |              |              |              |
| Logistic regression  | 0.902    | <b>0.521</b> |                                |                                  |              |              | 0.690        |              |
| MLP  | 0.813    | 0.500        |                                |                                  |              |              | 0.693        |              |
| Random forest  | 1.000    | 0.438        |                                |                                  |              |              | <b>0.694</b> |              |
| SVM  | 0.812    | 0.500        |                                |                                  |              |              | 0.673        |              |
| XGBoost  | 1.000    | 0.438        |                                |                                  |              |              | 0.644        |              |
| <b>Performance of the 3P models (PLT, BILI, INR) trained on the Vienna cohort for prediction of HVPG ≥16 mmHg</b>            |          |              |                                |                                  |              |              |              |              |
| Logistic regression  | 0.824    | 0.535        | 0.683                          | <b>0.599</b>                     | 0.637        | <b>0.881</b> | <b>0.686</b> | 0.720        |
| MLP  | 0.790    | 0.640        | <b>0.677</b>                   | 0.576                            | 0.583        | 0.867        | 0.668        | 0.689        |
| Random forest  | 1.000    | 0.597        | 0.593                          | 0.592                            | <b>0.681</b> | 0.835        | 0.670        | <b>0.763</b> |
| SVM  | 0.783    | <b>0.701</b> | 0.674                          | 0.543                            | 0.548        | 0.853        | 0.653        | 0.658        |
| XGBoost  | 1.000    | 0.610        | 0.561                          | 0.554                            | 0.632        | 0.808        | 0.663        | 0.741        |

The machine learning models' performance is reported as mean AUC values from 100 cross-validations in the single cohorts. Bold values highlight the best performing model in a specific dataset.

3P, 3 parameter; 5P, 5 parameter; aPTT, activated partial thromboplastin time; BILI, serum bilirubin; CHE, cholinesterase; GGT, gamma-glutamyltransferase; HVPG, hepatic venous pressure gradient; INR, international normalised ratio; LSM, liver stiffness measurement; MLP, multilayer perceptron; PLT, platelet count; SVM, support vector machine.

Our key finding is that MLMs trained on three (PLT, BILI, and INR) and five (aPTT instead of INR, and the addition of cholinesterase and GGT) routine laboratory parameters, following a train-validation split on the internal cohort, accurately predict CSPH and HVPG 16 mmHg. These MLMs enable the identification of individuals with cACLD who are at high risk of hepatic decompensation independently of liver disease aetiology.

We developed and employed models for binary classification and regression prediction tasks. Notably, 5P and 3P MLMs performed even better than those using all available parameters, which further facilitates their wide applicability (Fig. S2).

For severe PH, the best 5P MLM, based on LR, included PLT, BILI, aPTT, cholinesterase, GGT, and resulted in a mean AUC of 0.887 in CV in the Vienna cohort. The internally trained 5P MLMs performed reasonably in the Modena dataset with an AUC of 0.694 but less well in the Antwerp dataset (AUC = 0.521). Unfortunately, cholinesterase is not a routine parameter in other centres; hence, the validation dataset size for the 5P MLMs was limited. When trained on merged records, the 5P MLMs yielded AUCs ranging from 0.899 to 0.542. For prediction of CSPH, the internally trained models performed worse in the Antwerp dataset (AUC = 0.449) and moderately in the Modena (AUC = 0.694) and merged validation datasets (Table 3).

The best 3P MLMs included PLT, BILI, and INR. LR of the internally trained 3P model using CV on the Vienna dataset yielded an AUC of 0.813. Depending on the dataset, external validation produced heterogeneous yields. Predictive performance in the Madrid dataset exceeded that in the internal

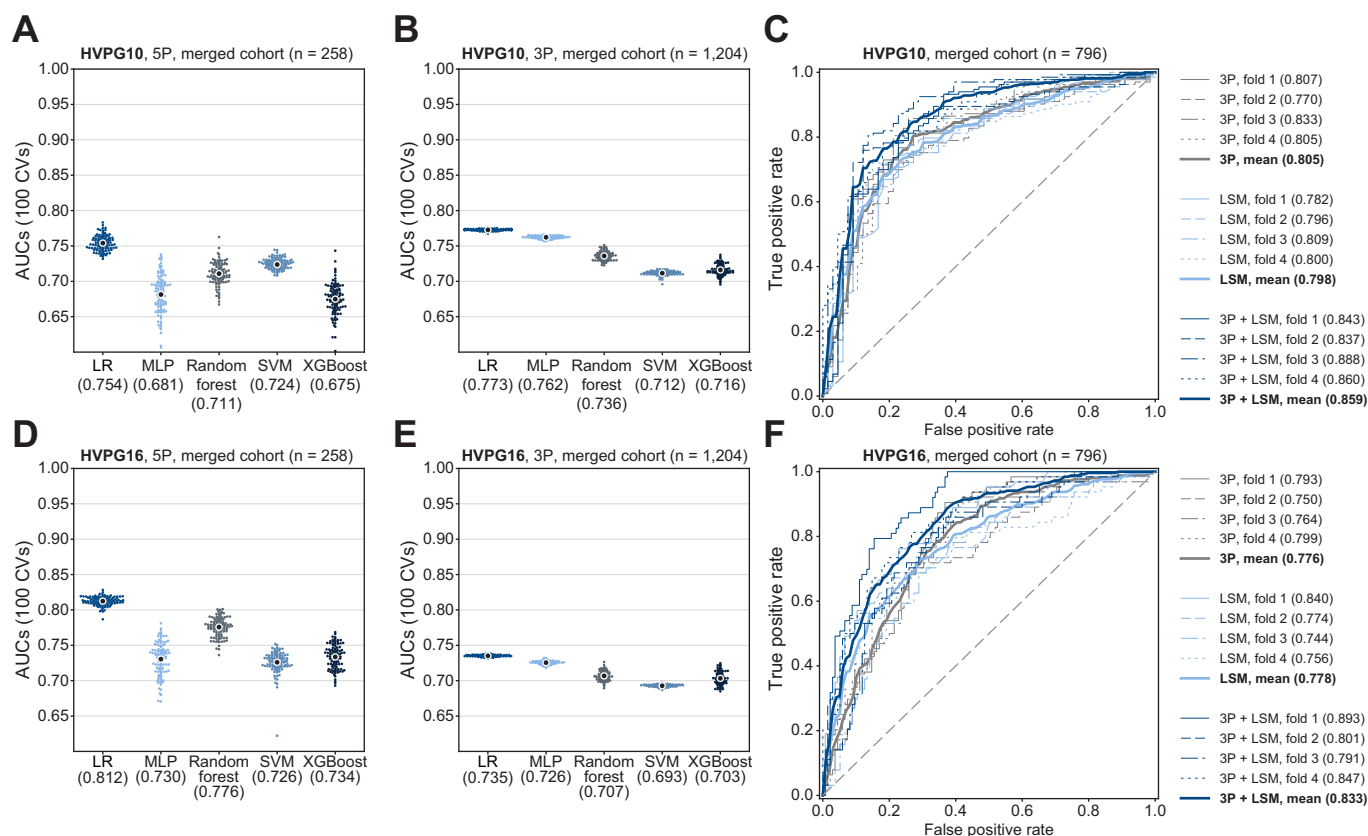
cohort (Table 2). LR showed the best performance among models in the merged cohort with the 3P set (AUC = 0.735).

Using the extracted coefficients of the discussed MLMs, we developed an online calculator that can be readily used for risk estimation in the clinical management of individuals with cACLD.<sup>26</sup>

To the best of our knowledge, there are no published studies on the prediction of severe PH using machine learning techniques that rely solely on laboratory parameters. Previously, an MLM was described for the non-invasive diagnosis of oesophageal varices in individuals with compensated cirrhosis.<sup>27</sup> This approach employed RF with widely available laboratory parameters, including PLT, BILI and INR, with demographic data, aetiology, and presence of complications. In the validation sets, the AUC ranged between 0.75 and 0.82. While this is an example of a non-invasive assessment of ACLD complications with predictive capability comparable to what we observe in some datasets, this approach cannot be directly translated to HVPG prediction. Utilising a decompensation event as a feature would introduce an association bias, as PH is one of the primary drivers of such events.

The reported performance of 5P in the internal Vienna cohort and 3P in the Madrid dataset closely aligns with the predictive capability of the previously reported neural network method based on histological readouts for CSPH (*i.e.*, HVPG 10 mmHg) prediction (AUC of 0.85 on the training set, 0.76 on the test set).<sup>14</sup> Remarkably, combining the morphological parameters with serum markers, including ELF, led to only a minor improvement in MLM prediction in that external study. In another study with 107 participants, multiple methods





**Fig. 3. Performance of the machine learning models to predict HVPG  $\geq 10$  mmHg and HVPG  $\geq 16$  mmHg in the merged cohort.** (A, B) Performance of 5P and 3P models to discriminate individuals with vs. without clinically significant portal hypertension (*i.e.*, HVPG  $\geq 10$  mmHg). (D, E) Performance of 5P and 3P to discriminate individuals with vs. without HVPG  $\geq 16$  mmHg. For each model, mean AUCs from individual CVs are shown as dots, and the mean AUC across all 100 CVs is reported. (C, F) Performance of 3P and 5P models compared to liver stiffness measurement for HVPG  $\geq 10$  mmHg and HVPG  $\geq 16$  mmHg thresholds in the merged cohort, single CV. The AUC is reported across 4-fold CVs. 3P, 3 parameter; 5P, 5 parameter; CSPH, clinically significant portal hypertension; CVs, cross-validations; HVPG, hepatic venous pressure gradient, LSM, liver stiffness measurement; MLP, multilayer perceptron; SVM, support vector machine.

were employed to identify those with CSPH.<sup>28</sup> Despite the lack of an external validation cohort and a limited number of participants for MLM training, LR, RF, and MLP were the best performing models, consistent with our findings. We also conclude that MLMs, particularly LR, RF, and MLP in both 5P and 3P settings, are useful clinical tools for HVPG prediction.

We evaluated whether our chosen features outperform LSM alone for detecting PH. In the merged cohort, 3P and 5P performance was similar to LSM (Table 4). Combining laboratory-based prediction with LSM can further improve predictive performance, with AUCs reaching 0.858 for CSPH using 3P+LSM and 0.901 for HVPG 16 mmHg using 5P+LSM. These results are superior to either LSM or MLMs alone (Table 4).

**Table 3. Performance of the machine learning models, trained on the merged cohort, for the assessment of portal hypertension severity in the different single cohorts.**

| Model  | Datasets |         |                             |                               |           |        |        |       |
|--|----------|---------|-----------------------------|-------------------------------|-----------|--------|--------|-------|
|  | Vienna   | Antwerp | Barcelona - Hospital Clinic | Barcelona - Hospital Sant Pau | Frankfurt | Madrid | Modena | Paris |
| <b>Performance of the final 5P model (PLT, BILI, aPTT, CHE, GGT) trained on the merged cohort for prediction of HVPG <math>\geq 10</math> mmHg</b> |          |         |                             |                               |           |        |        |       |
| Logistic regression  | 0.839    | 0.449   |                             |                               |           |        |        | 0.725 |
| <b>Performance of the final 5P model (PLT, BILI, aPTT, CHE, GGT) trained on the merged cohort for prediction of HVPG <math>\geq 16</math> mmHg</b> |          |         |                             |                               |           |        |        |       |
| Logistic regression  | 0.899    | 0.542   |                             |                               |           |        |        | 0.695 |
| <b>Performance of the final 3P model (PLT, BILI, INR) trained on the merged cohort for prediction of HVPG <math>\geq 10</math> mmHg</b>            |          |         |                             |                               |           |        |        |       |
| Logistic regression  | 0.802    | 0.451   | 0.739                       | 0.856                         | 0.629     | 0.873  | 0.721  | 0.796 |
| <b>Performance of the final 3P model (PLT, BILI, INR) trained on the merged cohort for prediction of HVPG <math>\geq 16</math> mmHg</b>            |          |         |                             |                               |           |        |        |       |
| Logistic regression  | 0.822    | 0.589   | 0.686                       | 0.629                         | 0.647     | 0.887  | 0.688  | 0.727 |

The performance of different machine learning models for the prediction of HVPG  $\geq 16$  mmHg is reported as mean AUC values from 100 cross-validations in the single cohorts. 3P, 3 parameter; 5P, 5 parameter; aPTT, activated partial thromboplastin time; BILI, serum bilirubin; CHE, cholinesterase; GGT, gamma-glutamyltransferase; HVPG, hepatic venous pressure gradient; INR, international normalised ratio; LSM, liver stiffness measurement; PLT, platelet count.

**Table 4. Performance of the final logistic regression models on the merged cohort.**

| Model    | Parameters                     | Participants with all parameters available | AUC HVPG $\geq 10$ mmHg | AUC HVPG $\geq 16$ mmHg | r2 score |
|----------|--------------------------------|--|-------------------------|-------------------------|----------|
| 5P       | PLT, BILI, aPTT, CHE, GGT      | 258  | 0.789                   | 0.828                   | 0.291    |
| 3P       | PLT, BILI, INR                 | 1,204                                      | 0.775                   | 0.737                   | 0.215    |
| 5P + LSM | PLT, BILI, aPTT, CHE, GGT, LSM | 208  | 0.832                   | 0.901                   | 0.480    |
| 3P + LSM | PLT, BILI, INR, LSM            | 796  | 0.858                   | 0.835                   | 0.431    |
| LSM-only | LSM                            | 804  | 0.799                   | 0.778                   | 0.282    |

Comparison of machine learning models for the prediction of HVPG  $\geq 10$  and  $\geq 16$  mmHg.

3P, 3 parameter; 5P, 5 parameter; aPTT, activated partial thromboplastin time; BILI, serum bilirubin; CHE, cholinesterase; GGT, gamma-glutamyltransferase; HVPG, hepatic venous pressure gradient; INR, international normalised ratio; LSM, liver stiffness measurement; PLT, platelet count.

However, we did not focus on LSM by design since it requires specialised equipment and training that may not be available in a common scenario.

Since our approach is aetiology-agnostic, the different proportions of liver disease aetiologies in the merged cohort could explain the heterogeneity in model performance on a single-dataset resolution (Fig. 1C). Surprisingly, we found that the 3P model worked best in the NAFLD subset for both HVPG thresholds (n = 286, AUC up to 0.836) (Table S4). MLMs performed worse in individuals with cholestasis (n = 63, AUC up to 0.666); however, given the low number of individuals with a cholestatic aetiology (5.2% of the merged cohort) and their distinct but variable component of presinusoidal PH, it would require a dedicated follow-up study to assess HVPG prediction models in these patients. Importantly, liver disease activity contributed to MLM performance, with inactive disease (abstinent ALD, suppressed viral hepatitis) resulting in better prediction. The performance scores, nevertheless, were comparable. To conclude, our machine learning approach can still be applied to individuals with various liver disease aetiologies (e.g., ALD, viral, NAFLD) and with distinct liver disease activity. Furthermore, well-designed and sufficiently powered datasets would likely enable the adoption of our MLMs for cholestasis and rare liver diseases.

Our study has limitations. First, the patient datasets within the validation cohort were heterogeneous. To address this, the patient records from all datasets were aggregated as a

merged dataset and used in CV, producing less prediction dispersion. Second, our study inclusion criteria comprised HVPG  $\geq 6$  mmHg, while, in a clinical setting, non-invasive assessments of PH could be indicated in individuals with lower HVPG values. Third, the reported performance may not reflect certain underrepresented aetiologies, such as cholestatic and rare liver diseases. Finally, we utilised LR within the RFE algorithm to find suitable features for MLMs, explaining why LR-based models performed better than other models – although not in all scenarios; however, the application of LR allowed us to extract coefficients and develop a simple risk prediction tool.

In conclusion, the presented 5P and 3P MLMs have promising clinical utility for the non-invasive prediction of HVPG  $\geq 10$  mmHg and HVPG  $\geq 16$  mmHg in individuals with cACLD. This approach could be used clinically to prioritise treatments aimed at preventing decompensation and for the selection of participants for clinical trials. We still consider invasive measurement of HVPG necessary for the reliable identification of individuals with cACLD and severe PH, and for the assessment of response to aetiological or PH-lowering treatments. The performance of our 5P and 3P MLMs for the prediction of CSPH and severe PH will be validated in larger cohorts by assessing decompensating events during long-term follow-up. Currently, our online risk calculator based on widely available laboratory parameters can help clinicians to assess the risk of CSPH or severe PH in their patients.<sup>26</sup>

## Affiliations

<sup>1</sup>CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria; <sup>2</sup>Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases (LBI-RUD), Vienna, Austria; <sup>3</sup>Vienna Hepatic Hemodynamic Lab (HEPEX), Division of Gastroenterology and Hepatology, Department of Internal Medicine III, Medical University of Vienna, Vienna, Austria; <sup>4</sup>Christian Doppler Laboratory for Portal Hypertension and Liver Fibrosis, Medical University of Vienna, Vienna, Austria; <sup>5</sup>Unit of Gastroenterology, Hepatic Hemodynamic Laboratory, Università degli Studi di Modena e Reggio Emilia (UNIMORE), Modena, Italy; <sup>6</sup>Department of Gastroenterology, Hospital Universitario Ramón y Cajal, IRYCIS, CIBEREHD, Universidad de Alcalá, Madrid, Spain; <sup>7</sup>Hospital de la Santa Creu i Sant Pau. Biomedical Research Institute Sant Pau (IIB Sant Pau), Universitat Autònoma de Barcelona, Barcelona, Spain; <sup>8</sup>Barcelona Hepatic Hemodynamic Laboratory, Liver Unit, Hospital Clínic, Institut de Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona; CIBEREHD (Centro de Investigación Biomédica en Red Enfermedades Hepáticas y Digestivas), Health Care Provider of the European Reference Network on Rare Liver Disorders, Barcelona, Spain; <sup>9</sup>Université de Paris, AP-HP, Hôpital Beaujon, Service d'Hépatologie, DMU DIGEST, Centre de Référence des Maladies Vasculaires du Foie, FILFOIE, ERN RARE-LIVER, Centre de recherche sur l'inflammation, Inserm, UMR 1149, Paris, France; <sup>10</sup>Department of Gastroenterology and Hepatology, Antwerp University Hospital, Antwerp, Belgium; <sup>11</sup>Laboratory of Experimental Medicine and Pediatrics (LEMP) – Gastroenterology & Hepatology, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium; <sup>12</sup>Department of Internal Medicine I, Goethe University Clinic, Frankfurt, Germany; <sup>13</sup>European Foundation for the Study of Chronic Liver Failure, EFCLIF, Barcelona, Spain; <sup>14</sup>Department of Internal Medicine B, WWU Münster, Münster, Germany; <sup>15</sup>Gastro Unit, Medical Section, Hvidovre Hospital and Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; <sup>16</sup>Department of Clinical Physiology and Nuclear Medicine, Center for Functional and Diagnostic Imaging and Research, Faculty of Health Sciences Hvidovre Hospital, University of Copenhagen, Hvidovre, Denmark

## Abbreviations

3P, three parameter; 5P, five parameter; ACLD, advanced chronic liver disease; ALD, alcohol-related liver disease; aPTT, activated partial thromboplastin time; BILI, bilirubin; cACLD, compensated ACLD; CSPH, clinically significant portal hypertension; CV, k-fold cross-validation; ELF, enhanced liver fibrosis; GGT, gamma-glutamyltransferase; HVPG, hepatic venous pressure gradient; INR, international normalised ratio; LR, logistic regression; LSM, liver stiffness measurement; MELD, model for end-stage liver disease; MLM, machine learning model; MLP, multilayer perceptron; NAFLD, non-alcoholic fatty liver disease; PH, portal hypertension; PLT, platelet count; RFE, recursive feature elimination; SVM, support vector machine; VICIS, Vienna Cirrhosis Study.

## Financial support

This study did not receive specific funding. Some authors report financial support of their research outside of this study: SF holds a senior clinical investigator fellowship from the Research Foundation Flanders (FWO) (1802154N). JT is further supported by grants from the Deutsche Forschungsgemeinschaft (SFB TRR57 to P18 and CRC1382 A09), European Union's Horizon 2020 Research and Innovation Programm (Galaxy, No. 668031, MICROB-PREDICT, No. 825694 and DECISION No.84794), and Societal Challenges - Health, Demographic Change and Wellbeing (Liverhope No. 731875). P-E.R.'s laboratory receives financial supports from the "Institut National de la Santé et de la Recherche Médicale" (ATIP Avenir), the "Agence Nationale pour la Recherche" (ANR-18-CE14-0006-01, RHU QUID-NASH, ANR-18-IDEX-0001), "Émergence, Ville de Paris", Fondation ARC and from the European Union's Horizon 2020 research and innovation programme under grant agreement No 847949 (DECISION). TR, OP, BS, BH, MM were co-supported by the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development, Boehringer-Ingelheim, and the Christian Doppler Research Association.

## Conflict of interest

JR, OP, TV, SK, WK, BH, BS, LM, WG, PF, LT, VH, MS, DS, CV, AB, FS, FI, TG, AA, VP, LLG, FB, SM declare no competing interests. MM consults for, advises, is on the speakers' bureau for, and/or received travel support from AbbVie, Collective Acumen, Gilead, and W.L. Gore & Associates. TR received grant support from AbbVie, Boehringer-Ingelheim, Gilead, MSD, Philips Healthcare, Gore; speaking honoraria from AbbVie, Gilead, Gore, Intercept, Roche, MSD; consulting/advisory board fee from AbbVie, Bayer, Boehringer-Ingelheim, Gilead, Intercept, MSD, Siemens; and travel support from Boehringer-Ingelheim, Gilead and Roche. SF has received grants from Astellas, Falk Pharma, Genfit, Gilead Sciences, GlympsBio, Janssens Pharmaceutica, Inventiva, Merck Sharp & Dome, Pfizer, Roche; has acted as a consultant for AbbVie, Actelion, Aelin Therapeutics, Aligos Therapeutics, Allergan, Astellas, Astra Zeneca, Bayer, Boehringer-Ingelheim, Bristol-Meyers Squibb, CSL-Behring, Coherus, Echoscans, Eisai, Enyo, Galapagos, Galmed, Genentech, Genfit, Gilead Sciences, Intercept, Inventiva, Janssens Pharmaceutica, Julius Clinical, Madrigal, Medimmune, Merck Sharp & Dome, NGM Bio, Novartis, Novo Nordisk, Promethera, Roche; has been a lecturer for AbbVie, Allergan, Bayer, Eisai, Genfit, Gilead Sciences, Janssens Cilag, Intercept, Inventiva, Merck Sharp & Dome, Novo Nordisk, Promethera. JT received grant support from Bayer, Falk and Gore; speaking honoraria from Gore, Falk, CSL-Behring, Grifols; consulting/advisory board fee from Alexion, Boehringer-Ingelheim, Gore, Grifols, CSL-Behring, Versantis and MSD. P-ER has received research funding from Terrafirma and acted as a consultant for Mursla and Abbelight, provided training sessions for Cook and received speaker fees from Tillots pharma. LG has received funding for research from Novo Nordisk, Sobi International, Alexion, Gilead, acted as consultant for Novo Nordisk and Pfizer, and received speakers fees from Novo Nordisk, Sobi, and Gedeon Richter. JCGP advisory for GORE and Cook and unrestricted grant from Mallinckrodt.

Please refer to the accompanying ICMJE disclosure forms for further details.

## Authors' contributions

JR, OP, BS, TR, MM, SK contributed to the study concept and design. JR and OP performed the computational and statistical analysis. BS, FS, MS, DS, FI, TG, AA, LT, CV, AB, JCGP, VP, VH, P-ER, LM, TV, WK, SF, JT, WG, PGF, LLG, FB, and SM organised and provided datasets with the relevant parameters according to the study criteria. OP, JR and TR prepared the manuscript draft. All authors contributed intellectual content, edited, and approved the final draft of the manuscript.

## Data availability statement

Participating centres provided depersonalised patient datasets for this study. They are available from the corresponding author on request. The code is publicly available at <https://github.com/reinisj/HVPG16>. The online calculator for predicting severe portal hypertension probability is available at <https://liver.at/vlsg/HVPG-Calculator/>.

## Acknowledgments

The research group of TR (OP, BS, BSH) was co-supported by the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development, the Christian Doppler Research Association, and Boehringer-Ingelheim, which is gratefully acknowledged. Elements of the graphical abstract were created with [BioRender.com](https://www.biorender.com).

## Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhep.2022.09.012>.

## References

*Author names in bold designate shared co-first authorship*

- [1] Pimpin L, Cortez-Pinto H, Negro F, Corbould E, Lazarus JV, Webber L, et al. Burden of liver disease in Europe: epidemiology and analysis of risk factors to identify prevention policies. *J Hepatol* 2018;69:718–735.
- [2] D'Amico G, Garcia-Tsao G, Pagliaro L. Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. *J Hepatol* 2006;44:217–231.
- [3] Ripoll C, Groszmann R, Garcia-Tsao G, Grace N, Burroughs A, Planas R, et al. Hepatic venous pressure gradient predicts clinical decompensation in patients with compensated cirrhosis. *Gastroenterology* 2007;133:481–488.
- [4] de Franchis R, Bosch J, Garcia-Tsao G, Reiberger T, Ripoll C, Abraldes JG, et al. Baveno VII - renewing consensus in portal hypertension. *J Hepatol* 2021;0. <https://doi.org/10.1016/j.jhep.2021.12.022>.
- [5] Reiberger T, Schwabl P, Trauner M, Peck-Radosavljevic M, Mandorfer M. Measurement of the hepatic venous pressure gradient and transjugular liver biopsy. *J Vis Exp* 2020. <https://doi.org/10.3791/58819>.
- [6] Sauerbruch T, Schierwagen R, Trebicka J. Managing portal hypertension in patients with liver cirrhosis. *F1000Res* 2018;7. <https://doi.org/10.12688/f1000research.13943.1>.
- [7] Mandorfer M, Kozbial K, Schwabl P, Chromy D, Semmler G, Stättermayer AF, et al. Changes in hepatic venous pressure gradient predict hepatic decompensation in patients who achieved sustained virologic response to interferon-free therapy. *Hepatology* 2020;71:1023–1036.
- [8] Berzigotti A, Rossi V, Tiani C, Pierpaoli L, Zappoli P, Riili A, et al. Prognostic value of a single HVPG measurement and Doppler-ultrasound evaluation in patients with cirrhosis and portal hypertension. *J Gastroenterol* 2011;46:687–695.
- [9] Turco L, Garcia-Tsao G, Magnani I, Bianchini M, Costetti M, Caporali C, et al. Cardiopulmonary hemodynamics and C-reactive protein as prognostic indicators in compensated and decompensated cirrhosis. *J Hepatol* 2018;68:949–958.
- [10] Reverter E, Cirera I, Albillos A, Debernardi-Venon W, Abraldes JG, Llop E, et al. The prognostic role of hepatic venous pressure gradient in cirrhotic patients undergoing elective extrahepatic surgery. *J Hepatol* 2019;71:942–950.
- [11] La Mura V, Garcia-Guix M, Berzigotti A, Abraldes JG, Garcia-Pagan JC, Villanueva C, et al. A prognostic strategy based on stage of cirrhosis and HVPG to improve risk stratification after variceal bleeding. *Hepatology* 2020;72:1353–1365.
- [12] Abraldes JG, Trebicka J, Chalasani N, D'Amico G, Rockey DC, Shah VH, et al. Prioritisation of therapeutic targets and trial design in cirrhotic portal hypertension. *Hepatology* 2019;69:1287–1299.
- [13] Abraldes JG, Bureau Christophe, Stefanescu H, Augustin S, Ney M, Blasco H, et al. Noninvasive tools and risk of clinically significant portal hypertension and varices in compensated cirrhosis: the "Anticipate" study. *Hepatology* 2016;64:2173–2184.
- [14] Bosch J, Chung C, Carrasco-Zevallos OM, Harrison SA, Abdelmalek MF, Shiffman ML, et al. A machine learning approach to liver histological

- evaluation predicts clinically significant portal hypertension in NASH cirrhosis. *Hepatology* 2021;74:3146–3160.
- [15] Procopet B, Cristea VM, Robic MA, Grigorescu M, Agachi PS, Metivier S, et al. Serum tests, liver stiffness and artificial neural networks for diagnosing cirrhosis and portal hypertension. *Dig Liver Dis* 2015;47:411–416.
- [16] Reback J, jbrockmendel, McKinney W, Van den Bossche J, Augspurger T, Cloud P, et al. *pandas-dev/pandas: Pandas 1.3.5*. 2021. <https://doi.org/10.5281/zenodo.5774815>.
- [17] Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. *J Open Source Softw* 2019;4:1686.
- [18] Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, et al. API design for machine learning software: experiences from the scikit-learn project. *ArXiv [CsLG]*; 2013.
- [19] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *ArXiv [CsLG]*; 2016.
- [20] Waskom M. *seaborn: statistical data visualisation*. *J Open Source Softw* 2021;6:3021.
- [21] Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn* 2002;46:389–422.
- [22] Györi GP, Pereyra D, Rumpf B, Hackl H, Köditz C, Ortmayr G, et al. The von Willebrand factor facilitates Model for End-Stage Liver Disease-independent risk stratification on the waiting list for liver transplantation. *Hepatology* 2020;72:584–594.
- [23] Costa D, Simbrunner B, Jachs M, Hartl L, Bauer D, Patenostro R, et al. Systemic inflammation increases across distinct stages of advanced chronic liver disease and correlates with decompensation and mortality. *J Hepatol* 2021;74:819–828.
- [24] Simbrunner B, Marculescu R, Scheiner B, Schwabl P, Bucsecs T, Stadlmann A, et al. Non-invasive detection of portal hypertension by enhanced liver fibrosis score in patients with different aetiologies of advanced chronic liver disease. *Liver Int* 2020;40:1713–1724.
- [25] Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32–35.
- [26] Petrenko O. *HVPG16 online calculator*. Vienna liver study groups. 2022. <https://liver.at/vlsg/HVPG-Calculator/>. [Accessed 25 August 2022].
- [27] Dong TS, Kalani A, Aby ES, Le L, Luu K, Hauer M, et al. Machine learning-based development and validation of a scoring system for screening high-risk esophageal varices. *Clin Gastroenterol Hepatol* 2019;17:1894–1901.e1.
- [28] Marozas M, Zykus R, Sakalauskas A, Kupcinskis L, Lukoševičius A. Noninvasive evaluation of portal hypertension using a supervised learning technique. *J Healthc Eng* 2017;2017:6183714.