The use of advanced machine learning to predict outcomes after atezolizumab plus bevacizumab for advanced hepatocellular carcinoma: a retrospective cohort study

Mathew Vithayathil, Giulia Francesca Manfredi, Antonio D'Alessio, Claudia Angela Maria Fulgenzi, Ciro Celsa, Pasquale Lombardi, Aria Torkpour, Bernardo Stefanini, Peter R Galle, Naoshi Nishida, Bernhard Scheiner, Fionnuala Crowley, Linda Wu, Celina Ang, Thomas U Marron, Wei-Fan Hsu, Chun-Yen Lin, Ryan Po-Ting Lin, Susanna V Ulahannan, Haripriya Andanamala, Noha Soror, Ahmed Kaseb, Michael Lapelusa, Yehia I Mohamed, Martin Schönlein, Johann von Felden, Kornelius Schulze, Henning Wege, Brooke Wietharn, Caterina Vivaldi, Francesca Salani, Gianluca Masi, Tiziana Pressiani, Andrea Dalbeni, Leonardo Natola, Mariarosaria Marseglia, Fabio Piscaglia, Yi-Hsiang Huang, Pei-Chang Lee, Mario Pirisi, Alessandro Parisi, Natascha Rohlen, Robert Thimme, Marianna Silletta, Emanuela Di Giacomo, Bruno Vincenzi, Luca Galbato, Arndt Vogel, Giuseppe Cabibbo, Calogero Camma, Hong Jae Chon, Matthias Pinter, Alessio Cortellini, Masatoshi Kudo, Valentina Zanuso, Lorenza Rimassa, David J Pinato, Eric O Aboagye, Rohini Sharma

Summary

Background Combination immune checkpoint inhibitors are recommended as first-line therapy for advanced hepatocellular carcinoma. However, only a third of patients respond to treatment, and improved approaches to predict response are required. Using baseline clinical data, we aimed to use advanced machine learning models to predict overall survival and progression-free survival in patients with advanced hepatocellular carcinoma receiving atezolizumab plus bevacizumab.

Methods This retrospective cohort study was conducted at 24 centres across eight countries. Patients aged 18 years and older with a histological or radiological diagnosis of advanced hepatocellular carcinoma were included; those who had received previous systemic therapy for hepatocellular carcinoma were excluded. All patients received intravenous atezolizumab 1200 mg plus bevacizumab 15 mg/kg once every 3 weeks until disease progression. Seven supervised machine learning models, in combination with 13 feature selection techniques, were trained on 44 baseline clinical variables for the prediction of overall survival and progression-free survival. The three best-performing models, combined with their optimum feature selection techniques, were used to develop ensemble machine learning models for the prediction of overall survival and progression-free survival. The primary outcomes of the study were the predictions of overall survival, progression-free survival, and immunotherapy response using advanced machine learning. *k*-means clustering was used to stratify patients into two groups: those at low risk and those at high risk of either death (in the overall survival model) or disease progression (in the progression-free survival model).

Findings 934 patients who received immunotherapy from May 1, 2018 and were followed up until Oct 1, 2023 were screened, of whom 160 were excluded and 774 were included in the final study. Patients were divided into training (n=339), internal validation (n=146) and external validation (n=289) cohorts. Support vector machine, neural network, and naive Bayes algorithms had the best performance in the prediction of overall survival; for progression-free survival, the highest-performing algorithms were ridge regression, naive Bayes, and logistic regression. In the external validation cohort, the ensemble model for the prediction of overall survival (area under the receiver operating characteristic curve 0·75 [95% CI 0·69–0·81]) significantly outperformed all eight of the tested clinical benchmark variables: Barcelona Clinic Liver Cancer (BCLC) stage (0·54 [0·48–0·61]; p<0·0001), α-fetoprotein (AFP) concentration (0·60 [0·54–0·67]; p=0·0007), albumin–bilirubin (ALBI) grade (0·64 [0·58–0·71]; p=0·0003), neutrophil-to-lymphocyte ratio (0·56 [0·49–0·62]; p<0·0001), platelet-to-lymphocyte ratio (0·51 [0·44–0·58]; p<0·0001), combined ALBI grade and BCLC stage (0·67 [0·60–0·73]; p=0·0074), and two BCLC subclassifications (0·62 [0·55–0·69]; p=0·0007 and 0·61 [0·55–0·68]; p=0·0018). The ensemble model for the prediction of progression-free survival (0·64 [0·59–0·70]) outperformed five of the eight clinical predictors: BCLC stage (0·52 [0·46–0·58]; p<0·0001), neutrophil-to-lymphocyte ratio (0·53 [0·47–0·59]; p=0·0069), platelet-to-lymphocyte ratio (0·54 [0·48–0·60]; p=0·016), and two BCLC subclassifications (0·57 [0·50–0·64]; p=0·020 and 0·55 [0·49–0·62]; p=0·0091); the model did not outperform AFP concentration (0·59 [0·53–0·64]; p=0·14), ALBI grade (0·62 [0·56–0·67]; p=0·44), or combined ALBI grade and BCLC stage (0·59 [0·53–0·66]; p=0·12). For the overall survival model, patients stratified into the low-risk group had significantly longer median overall survival (16·4 months [95% CI 14·2–21·6]) than those in the high-risk group (4·8 months [3·0–6·9]; p<0·0001); similarly, patients stratified by the progression-free survival model into the low-risk group had significantly longer median progression-free survival (8·9 months [7·3–11·1]) than those in the high-risk group (3·7 months [2·9–5·6]; p=0·0021).

Interpretation Our advanced machine learning models, which use routinely collected baseline clinical variables, are robust and externally validated and outperform established clinical biomarkers for predicting clinical outcomes with

Lancet Digit Health 2026; 8: 100952

Published Online February 12, 2026
https://doi.org/10.1016/j.landig.2025.100952

Department of Surgery and Cancer, Imperial College London, London, UK (M Vithayathil MBBS, G F Manfredi MD, A D'Alessio MD PhD, C A M Fulgenzi MD, C Celsa MD PhD, P Lombardi MD, A Torkpour BSc, B Stefanini MD, A Cortellini MD PhD, Prof D J Pinato MD PhD, Prof E O Aboagye PhD, Prof R Sharma MBBS PhD); **Department of Translational Medicine, Università del Piemonte Orientale, Novara, Italy** (G F Manfredi, A D'Alessio, Prof M Pirisi MD); **Gastroenterology and Hepatology, Department of Health Promotion, Mother and Child Care, Internal Medicine and Medical Specialties, University of Palermo, Palermo, Italy** (C Celsa, M Silletta MD PhD, E Di Giacomo MD, Prof B Vincenzi MD PhD, L Galbato MD, Prof G Cabibbo MD PhD, Prof C Camma MD PhD); **Department of Medical and Surgical Sciences, University of Bologna, Italy** (B Stefanini, Prof F Piscaglia MD PhD); **Medical Department I, University Medical Center Mainz, Mainz, Germany** (Prof P R Galle MD); **Department of Gastroenterology and Hepatology, Kindai University Faculty of Medicine, Osaka, Japan** (Prof N Nishida MD PhD,

Prof M Kudo MD PhD); **Division of Gastroenterology and Hepatology, Department of Medicine III, Medical University of Vienna, Vienna, Austria** (B Scheiner MD, Prof M Pinter MD PhD); **Brookdale Department of Geriatrics and Palliative Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA** (F Crowley MBBChBAO); **Division of Hematology/Oncology, Department of Medicine, Tisch Cancer Institute, Mount Sinai Hospital, New York, NY, USA** (F Crowley, L Wu MD, C Ang MD, Prof T U Marron MD PhD); **Division of Hematology/Oncology, Department of Medicine, Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, New York, NY, USA** (L Wu); **Center for Digestive Medicine, Department of Internal Medicine, China Medical University Hospital, Taichung, Taiwan** (W-F Hsu MD PhD); **Department of Gastroenterology and Hepatology, Chang Gung Memorial Hospital, Linkou Branch, Taoyuan, Taiwan** (Prof C-Y Lin MD PhD, R P-T Lin MD); **College of Medicine, Chang Gung University, Taoyuan, Taiwan** (Prof C-Y Lin, R P-T Lin); **Division of Hematology/Oncology, Stephenson Cancer Center, University of Oklahoma, Oklahoma City, OK, USA** (S V Ulahannan MD, H Andanamala MD, N Soror MD); **Department of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA** (Prof A Kaseb MD, M Lapelusa MD, Y I Mohamed MD); **Department of Oncology, Hematology and Bone Marrow Transplantation with Section of Pneumology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany** (M Schönlein MD); **Department of Medicine I, University Medical Center Hamburg-Eppendorf, Hamburg, Germany** (J von Felden MD, K Schulze MD, Prof H Wege MD PhD); **Department of Medicine, Division of Medical Oncology, Kansas University Cancer Center, Kansas City, KS, USA**

atezolizumab plus bevacizumab. These data-driven models could be used to stratify patients with hepatocellular carcinoma for personalised treatment strategies.

Funding None.

Introduction

Hepatocellular carcinoma is the third leading cause of cancer-related mortality worldwide,[1] and the majority of cases are diagnosed at an advanced stage.[2] The landmark IMbrave150 trial[3] showed that combination immunotherapy with the immune checkpoint inhibitor (ICI) atezolizumab plus the anti-angiogenic drug bevacizumab significantly improved overall survival and progression-free survival in patients with advanced hepatocellular carcinoma.[3] These findings are supported by those of the HIMALAYA,[4] CheckMate-9DW,[5] and CARES-310[6] trials, which showed that other ICI-based regimens are also associated with a significant improvement in overall survival compared with tyrosine kinase inhibitors.

Despite the clear clinical benefit of combination ICI therapy, the identification of patients who could derive significant survival benefit remains elusive. Because hepatocellular carcinoma can be diagnosed radiologically,[7] molecular and genetic phenotyping of patients before treatment is not widely conducted. Real-world studies have used traditional linear model approaches to identify clinical factors—such as Child–Pugh class,[8] α-fetoprotein (AFP) concentration,[9] C-reactive protein (CRP) and AFP in Immunotherapy (CRAFITY) score,[10] albumin–bilirubin (ALBI) grade,[8] neutrophil-to-lymphocyte ratio, and platelet-to-lymphocyte ratio[11]—that could be used to predict clinical responses to combination ICI therapy. However, these studies focused on unidimensional disease features, rather than providing an overall estimation of the disease course based on multiple variables, limiting the power—and therefore utility—of these predictors in routine clinical practice. Given the risk of poor response and ICI-related toxicity, and the availability of alternative treatment strategies in advanced hepatocellular carcinoma, it is important that patients who are likely to benefit from combination ICI therapy are identified in advance of beginning treatment.

Machine learning is a key component of the artificial intelligence framework. Traditionally, clinical studies have focused on linear machine learning approaches, such as logistic regression and Cox regression, to identify predictors. However, advanced non-linear approaches can capture more complex relationships between baseline variables and clinical outcomes, allowing for more precise predictions. A frequent limitation of machine learning models is overfitting on training cohorts and diminishing performance on external validation cohorts. Feature selection techniques can be used to identify key predictor variables and have been shown to improve machine learning model performance in external validation cohorts.[12]

We aimed to use machine learning to predict clinical outcomes to atezolizumab plus bevacizumab, using readily available baseline clinical data from patients with advanced

Research in context

Evidence before this study

Randomised clinical trials have shown that atezolizumab plus bevacizumab is superior to systemic therapy for advanced hepatocellular carcinoma; however, only a third of patients respond to treatment and robust clinical biomarkers for predicting outcomes are scarce. We searched PubMed using the search string ("hepatocellular carcinoma" AND "immunotherapy" AND "biomarkers" AND "predictors"), with no language restrictions, for records published from Jan 1, 2010, to Sept 1, 2024. Our search retrieved 116 studies, the findings of which showed that markers derived from logistic regression models—such as the albumin–bilirubin grade, the neutrophil-to-lymphocyte ratio, and the platelet-to-lymphocyte ratio—had limited predictive performance when using linear models. To our knowledge, no studies to date have used advanced machine learning models with routine clinical variables to predict the response to atezolizumab plus bevacizumab in hepatocellular carcinoma.

Added value of this study

In this large, international, multicentre study, we developed and externally validated advanced machine learning models to predict clinical outcomes after immunotherapy, using standard-of-care baseline clinical data. Advance machine learning models in combination with feature selection methods outperformed established clinical biomarkers in the prediction of overall survival and progression-free survival in patients with hepatocellular carcinoma after treatment with atezolizumab plus bevacizumab.

Implications of all the available evidence

Using baseline clinical data, advanced machine learning models can identify patients who could derive benefit from atezolizumab and bevacizumab. These models could be used to stratify patients with advanced hepatocellular carcinoma for personalised treatment strategies.

hepatocellular carcinoma. Combinations of multiple machine learning algorithms with different feature selection techniques were evaluated on internal and external validation cohorts.

Methods

Study design and participants

This retrospective cohort study was conducted at 24 centres in eight countries (Austria, Germany, Italy, Japan, South Korea, Taiwan, the UK, and the USA). Patients aged 18 years and older, who received immunotherapy from May 1, 2018 and were followed up until Oct 1, 2023, were retrospectively included in the AB-real clinical study, an open registry of patients with unresectable advanced hepatocellular carcinoma treated with atezolizumab plus bevacizumab.[13] All patients had a histological or radiological diagnosis of hepatocellular carcinoma, as per the American Association for the Study of Liver Diseases criteria,[7] which was classified as either advanced disease or intermediate disease unsuitable for locoregional therapy according to the Barcelona Clinic Liver Cancer (BCLC) criteria.[14] Patients who had received previous systemic therapy for hepatocellular carcinoma were excluded. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of the Imperial College Healthcare Tissue Bank (R16008) and local ethics committees at each participating centre. Owing to the retrospective design of the study and anonymised data collection methods, informed consent was not considered necessary by the ethics review committee.

Baseline clinical parameters and data pre-processing

Baseline clinical parameters were collected for all patients before commencing combination ICI therapy (appendix p 3). As part of data pre-processing, all categorical data were converted into numerical values using one-hot encoding. For multiple level categories, the final level feature was removed. Features for which more than 25% of data were missing were excluded. The multivariate imputation by chained equations[15] method was used to impute missing values for the remaining features. After imputation, continuous features were standardised on the basis of the training cohort using Z standardisation.

Procedures and clinical outcomes

All patients received intravenous atezolizumab 1200 mg plus bevacizumab 15 mg/kg once every 3 weeks, in accordance with standard treatment protocol. Treatment continued until disease progression or intolerable ICI-related toxicity, as determined by assessment at the participating centres.

The primary outcomes of the study were the predictions of overall survival, progression-free survival, and immunotherapy response using advanced machine learning. Secondary outcomes were the performance of the machine learning models in patients with Child–Pugh class A liver dysfunction (appendix p 28) and the feature importance for the machine learning models (appendix pp 18–19).

Overall survival was defined as the time in months from the start of combination ICI therapy to death or the date of the last follow-up visit, whichever occurred earliest. Progression-free survival was defined as the time in months from the start of combination ICI therapy to progression or death, whichever occurred first. Progression was assessed radiologically, as per Response Evaluation Criteria in Solid Tumours (RECIST) version 1.1,[16] by CT or MRI at 6–12-week intervals, according to local practice. For machine learning prediction, overall survival was binarised to 0 or 1 on the basis of mortality at 12 months. Patients not known to have died and who were last seen more than 60 days before this 12-month endpoint were excluded from the analysis. Similarly, progression-free survival was binarised on the basis of disease progression or death at 6 months, in line with the median progression-free survival in the IMbrave150 study.[3] Patients not known to have died or progressed and who were last seen more than 60 days before this 6-month endpoint were excluded.

The overall response rate to combination ICI therapy was defined as the proportion of patients who had a complete response or a partial response during active ICI therapy, as per RECIST. Disease control rate was defined as the proportion of patients who had a complete response, a partial response, or stable disease during active ICI therapy. Overall response rate and disease control rate were evaluated in model-stratified risk groups (see Kaplan–Meier analysis).

Machine learning algorithms and feature selection techniques

Patients were divided into three geographical cohorts: Asia, Europe, and North America. The largest and smallest cohorts, Asia and North America, were combined. This combined cohort was then further randomly divided, in a 70:30 ratio, into a training set and an internal validation set. The European cohort was used as an independent external validation set.

Seven machine learning algorithms, based on the following linear, Bayesian, tree-based, and neural network methods, were trained on baseline data: logistic regression, naive Bayes, neural network, random forest, support vector machine, extreme gradient boosting (XGBoost), and ridge regression. Hyper-parameter optimisation was conducted on machine learning algorithms using grid-search ten-fold cross-validation. Models derived from different methods were used to minimise bias and variance, on the basis of previous studies.[12,17] Individual supervised machine learning models and hyper-parameters are shown in the appendix (p 4).

13 feature selection techniques were used on baseline data: least absolute shrinkage and selection operator (LASSO) regression, elastic net regression, recursive

(B Wietharn MD); **Trinity Health Family Medicine, Grand Rapids, MI, USA** (B Wietharn); **Division of Medical Oncology, Department of Translational Research and New Technologies in Medicine and Surgery, Pisa University, Italy** (C Vivaldi MD PhD, F Salani MD PhD, Prof G Masi MD); **Scuola Superiore Sant'Anna Pisa, Interdisciplinary Research Center, Pisa, Italy** (F Salani); **Medical Oncology and Hematology Unit, Humanitas Cancer Center, IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy** (T Pressiani MD, L Rimassa MD); **Section of General Medicine C, Medicine Department, University of Verona and University and Hospital Trust (AOUI) of Verona, Verona, Italy** (A Dalbeni MD PhD); **Liver Unit, Medicine Department, University of Verona and University and Hospital Trust (AOUI) of Verona, Verona, Italy** (A Dalbeni, L Natola MD); **Division of Internal Medicine, Hepatobiliary and Immunoallergic Diseases, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Italy** (M Marseglia MSc, Prof F Piscaglia); **Division of Gastroenterology and Hepatology, Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan** (Prof Y-H Huang MD PhD); **Department of Medical Research, Taipei Veterans General Hospital, Taipei, Taiwan** (Prof Y-H Huang, P-C Lee MD PhD); **Institute of Clinical Medicine, College of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan** (Prof Y-H Huang); **Faculty of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan** (P-C Lee); **Department of Oncology, Università Politecnica delle Marche, Azienda Ospedaliero-Universitaria delle Marche, Ancona, Italy** (A Parisi MD PhD); **Department of Medicine II (Gastroenterology, Hepatology, Endocrinology and Infectious Diseases), Faculty of Medicine, Freiburg University Medical Center, University of Freiburg, Freiburg, Germany** (N Rohlen PhD, Prof R Thimme MD PhD); **Hannover Medical School,**

Hanover, Germany (Prof A Vogel MD PhD); Division of Gastroenterology and Hepatology, Toronto General Hospital Medical Oncology, Princess Margaret Cancer Centre, Schwartz Reisman Liver Research Centre, Toronto, ON, Canada (Prof A Vogel); Division of Medical Oncology, Department of Internal Medicine, CHA Bundang Medical Center, CHA University School of Medicine, Seongnam, South Korea (Prof H J Chon MD PhD); Operative Research Unit of Oncology, Fondazione Policlinico Universitario Campus Bio-Medico, Rome, Italy (A Cortellini); Department of Medicine and Surgery, Universitá Campus Bio-Medico di Roma, Rome, Italy (A Cortellini); Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy (L Rimassa); Division of Oncology, Department of Translational Medicine, University of Piemonte Orientale, Novara, Italy (Prof D J Pinato); Phase 1 Unit, Fondazione Policlinico Universitario A Gemelli, IRCCS, Università Cattolica del Sacro Cuore, Rome, Italy (P Lombardi); Division of Oncology, ASST Melegnano e Martesana, Vizzolo Predabissi Hospital, Milan, Italy (V Zanuso MD)

Correspondence to: Prof Rohini Sharma, Department of Surgery and Cancer, Imperial College London, London W12 0HS, UK
r.sharma@imperial.ac.uk

See Online for appendix

For **MedCalc**, see https://www.medcalc.org/

feature elimination, principal component analysis, Boruta, mutual information, Pearson's correlation, Spearman's rank correlation, Kendall's rank correlation, ANOVA F-test, variance threshold, forward selection, and no feature selection. A description of each feature selection technique is given in the appendix (p 5).

All machine learning algorithms were run with each feature selection technique on the training dataset. The performance of each combination was evaluated on the internal validation set by the area under the receiver operating characteristic curve (AUC). The three machine learning models, in combination with the corresponding best-performing feature selection technique, with the highest AUC in the internal validation set for 12-month overall survival and 6-month progression-free survival were selected for the final models. An ensemble model averaging the performance of these three machine learning algorithms was used in the final models. The 95% CIs for AUC values were calculated using bootstrapping with replacement over 1000 iterations. The performance of the model was compared with predictions based on the following clinical benchmark variables: AFP concentration, BCLC stage, ALBI grade, neutrophil-to-lymphocyte ratio, platelet-to-lymphocyte ratio, combined ALBI grade and BCLC stage, and two previously reported BCLC subclassification systems[18,19] (appendix pp 6–7). A logistic regression model with grid-search ten-fold cross validation for each clinical benchmark variable was generated from the training dataset. The performance of the final ensemble model was compared with that of models using clinical benchmark data on the internal and external validation sets using DeLong's test.

To reveal the contribution of different features, we used the permutation feature importance technique[20] to calculate the feature importance for each model used in the final ensemble predictors for overall survival and progression-free survival. This calculation was run for each baseline variable with 50 iterations.

Kaplan–Meier analysis

Unsupervised machine learning using *k*-means clustering was used to stratify patients into different treatment response groups. Clusters were established from the probability estimates of the 12-month overall survival and 6-month progression-free survival ensemble machine learning models in the internal validation set. The optimal cluster numbers were determined using the knee point detection algorithm[21] from an elbow plot showing the sum of the squared Euclidean distance for each cluster number. Patients in the external validation set were stratified using clustering predictions from the *k*-means model. Kaplan–Meier analyses and log-rank tests were conducted on stratified patient groups.

Statistical analysis

We used χ^2 tests to evaluate categorical data and unpaired Student's *t*-tests to evaluate continuous data. All machine learning and feature selection analyses were conducted using the Scikit-learn package in Python 3.8.8. Receiver operating characteristic curves and Kaplan–Meier graphs were generated using MedCalc version 22.021 (MedCalc Software, Ostend, Belgium). p values less than 0·05 were considered statistically significant.

Role of the funding source

There was no funding source for this study.

Results

934 patients who received immunotherapy from May 1, 2018 and were followed up until Oct 1, 2023 were screened, of whom 160 were excluded and 774 were included in the study. Patients from Asia (n=339) and North America (n=146) were combined (n=485) and randomly divided into a training cohort (n=339) and a hold-out internal validation cohort (n=146). Patients from Europe comprised the independent external validation cohort (n=289). The baseline characteristics for each cohort are shown in table 1. The variance of baseline characteristics across different geographical cohorts was assessed using principal component analysis (appendix p 22); no clear batch effect was observed.

The median age was higher in the external validation cohort (67·9 years [IQR 62·0–75·4]) than in the internal validation cohort (66·0 years [IQR 60·0–73·0]; p=0·013). Rates of hepatitis B were lower and previous locoregional therapy were higher in the external validation cohort than in the internal validation cohort.

44 baseline clinical variables were used for machine learning prediction. The variables selected by each feature selection method are shown in the appendix (pp 8–11). The performance of the combinations of each machine learning algorithm and feature selection technique on the internal validation set is shown in the appendix (p 23). For the prediction of 12-month overall survival, the three highest-performing combinations on the internal validation set were support vector machines with Kendall's rank correlation, neural network with recursive feature elimination, and naive Bayes with Pearson's correlation. For 6-month progression-free survival, ridge regression and logistic regression with all features and naive Bayes with Pearson's correlation were the three highest-performing models. These models were used in a final ensemble model, which had AUCs in the internal validation set of 0·75 (95% CI 0·65–0·83) for overall survival and 0·70 (0·62–0·78) for progression-free survival (table 2).

The final ensemble model was evaluated against predictions based on known clinical benchmark variables. For the prediction of overall survival, the machine learning model, with an AUC of 0·75 (95% CI 0·69–0·81) in the external validation set, significantly outperformed all clinical benchmarks (BCLC stage, AFP concentration, ALBI grade, neutrophil-to-lymphocyte ratio, and platelet-to-lymphocyte ratio; figure 1, table 2), as well as the combined ALBI grade and BCLC stage and two BCLC

	Training cohort (n=339)	Internal validation cohort (n=146)	External validation cohort (n=289)	p value*
Age, years	65·2 (58·2–72·6)	66·0 (60·0–73·0)	67·9 (62·0–75·4)	0·013
Sex				
Male	279 (82%)	121 (83%)	229 (79%)	..
Female	60 (18%)	25 (17%)	60 (21%)	0·44
Risk factors for chronic liver disease				
Metabolic dysfunction-associated fatty liver disease	38 (11%)	17 (12%)	52 (18%)	0·12
Alcohol-related liver disease	103 (30%)	32 (22%)	75 (26%)	0·42
Hepatitis B infection	145 (43%)	64 (44%)	46 (16%)	<0·0001
Hepatitis C infection	96 (28%)	37 (25%)	82 (28%)	0·58
Liver disease				
Ascites	85 (25%)	30 (21%)	83 (29%)	0·086
Hepatic encephalopathy	8 (2%)	4 (3%)	5 (2%)	0·73
Child–Pugh class				
A	277 (82%)	124 (85%)	226 (78%)	0·12
B	62 (18%)	22 (15%)	63 (22%)	0·12
Maximum tumour diameter, cm	6·7 (4·4)	6·8 (4·5)	6·0 (4·0)	0·079
Macrovascular invasion	119 (35%)	46 (32%)	104 (36%)	0·41
AFP concentration >400 ng/dL	137 (40%)	52 (36%)	109 (38%)	0·75
ALBI grade				
1	96 (28%)	39 (27%)	60 (21%)	0·20
2	214 (63%)	89 (61%)	181 (63%)	0·82
3	29 (9%)	18 (12%)	48 (17%)	0·30
Extrahepatic spread	169 (50%)	71 (49%)	116 (40%)	0·11
ECOG-PS				
0	156 (46%)	60 (41%)	161 (56%)	0·0055
1	175 (52%)	82 (56%)	116 (40%)	0·0022
2	8 (2%)	4 (3%)	12 (4%)	0·55
BCLC stage				
A	6 (2%)	4 (3%)	9 (3%)	1·00
B	78 (23%)	32 (22%)	53 (18%)	0·45
C	255 (75%)	110 (75%)	227 (79%)	0·57
Previous locoregional treatment				
Resection	78 (23%)	39 (27%)	71 (25%)	0·71
Radiofrequency ablation	63 (19%)	23 (16%)	23 (8%)	0·020
Transarterial chemoembolisation	147 (43%)	60 (41%)	87 (30%)	0·029
Transarterial radioembolisation	32 (9%)	23 (16%)	8 (3%)	<0·0001
External beam radiotherapy	63 (19%)	21 (14%)	15 (5%)	0·0019

Data are median (IQR), n (%), or mean (SD). AFP=α-fetoprotein. ALBI=albumin–bilirubin. BCLC=Barcelona Clinic Liver Cancer. ECOG-PS=Eastern Cooperative Oncology Group Performance Status. *p values for the difference between internal and external validation datasets were calculated using unpaired Student's *t* tests for continuous variables and χ^2 tests for categorical data.

Table 1: **Baseline characteristics**

subclassification systems (table 2). For the prediction of progression-free survival, the machine learning model significantly outperformed BCLC stage, neutrophil-to-lymphocyte ratio, and platelet-to-lymphocyte ratio (figure 1, table 2) and both BCLC subclassification systems (table 2), with an AUC of 0·64 (0·59–0·70) in the external validation cohort.

In the internal validation cohort, the median overall survival was 14·1 months (95% CI 10·0–16·1) and the median progression-free survival was 5·0 months (3·5–6·2). For each outcome, *k*-means clustering of ensemble model probabilities identified two distinct risk groups: those at low risk and those at high risk (appendix p 24; probability thresholds shown in appendix p 25). For overall survival, the median survival time was 6·8 months (3·5–10·9) for the high-risk group and 16·1 months (14·1–20·0) for the low-risk group (p<0·0001). The high-risk group also had a significantly shorter median progression-free survival time (3·4 months [2·6–5·0]) than the low-risk group (7·1 months [4·1–11·3]; p=0·0073; appendix p 26).

In the external validation cohort, the median overall survival time was 11·7 months (95% CI 10·5–14·2) and the median progression-free survival was 6·8 months (5·0–8·3). The use of *k*-means clustering for risk stratification resulted

	Training cohort (n=339)	Internal validation cohort (n=146)	External validation cohort (n=289)	
			AUC (95% CI)	p value*
Overall survival				
BCLC stage	0·59 (0·54–0·64)	0·61 (0·53–0·69)	0·54 (0·48–0·61)	<0·0001
AFP concentration	0·68 (0·61–0·74)	0·59 (0·48–0·69)	0·60 (0·54–0·67)	0·0007
ALBI grade	0·67 (0·61–0·72)	0·58 (0·49–0·67)	0·64 (0·58–0·71)	0·0003
Neutrophil-to-lymphocyte ratio	0·65 (0·58–0·71)	0·67 (0·58–0·77)	0·56 (0·49–0·62)	<0·0001
Platelet-to-lymphocyte ratio	0·58 (0·51–0·65)	0·60 (0·50–0·71)	0·51 (0·44–0·58)	<0·0001
Combined ALBI grade and BCLC stage	0·70 (0·64–0·76)	0·66 (0·56–0·75)	0·67 (0·60–0·73)	0·0074
BCLC subclassification (Giannini et al)[18]	0·73 (0·66–0·78)	0·66 (0·56–0·75)	0·62 (0·55–0·69)	0·0007
BCLC subclassification (Lin et al)[19]	0·65 (0·58–0·71)	0·57 (0·48–0·68)	0·61 (0·55–0·68)	0·0018
Ensemble model	0·86 (0·81–0·90)	0·75 (0·65–0·83)	0·75 (0·69–0·81)	Ref
Progression-free survival				
BCLC stage	0·60 (0·56–0·65)	0·55 (0·48–0·63)	0·52 (0·46–0·58)	<0·0001
AFP concentration	0·63 (0·57–0·69)	0·62 (0·54–0·72)	0·59 (0·53–0·64)	0·14
ALBI grade	0·57 (0·52–0·62)	0·54 (0·46–0·62)	0·62 (0·56–0·67)	0·44
Neutrophil-to-lymphocyte ratio	0·60 (0·54–0·66)	0·64 (0·54–0·73)	0·53 (0·47–0·59)	0·0069
Platelet-to-lymphocyte ratio	0·58 (0·52–0·64)	0·58 (0·49–0·68)	0·54 (0·48–0·60)	0·016
Combined ALBI grade and BCLC stage	0·64 (0·58–0·70)	0·56 (0·47–0·66)	0·59 (0·53–0·66)	0·12
BCLC subclassification (Giannini et al)[18]	0·64 (0·59–0·70)	0·57 (0·48–0·66)	0·57 (0·50–0·64)	0·020
BCLC subclassification (Lin et al)[19]	0·53 (0·57–0·68)	0·57 (0·48–0·65)	0·55 (0·49–0·62)	0·0091
Ensemble model	0·79 (0·74–0·83)	0·70 (0·62–0·78)	0·64 (0·59–0·70)	Ref

Data are AUC (95% CI) unless otherwise indicated. AFP=α-fetoprotein. ALBI=albumin–bilirubin. AUC=area under the receiver operating characteristic curve. BCLC=Barcelona Clinic Liver Cancer. p values comparing the AUCs from predictions using clinical benchmark data with the AUCs from the ensemble models were calculated using the DeLong test. *p values comparing the AUCs from predictions using clinical benchmark data with the AUCs from the ensemble models were calculated using the DeLong test.

Table 2: **Performance of the ensemble models compared with clinical benchmark variables in predicting overall survival and progression-free survival**

in 145 (63%) of the 229 patients for whom full follow-up data were available being stratified into the low-risk group, with a median survival time of 16·4 months (14·2–21·6), and the remaining 84 (37%) patients being stratified into the high-risk group, with a significantly shorter median survival time of 4·8 months (3·0–6·9; p<0·0001; figure 2). For progression-free survival, 155 (54%) of 289 patients were classified as low-risk and 134 (46%) as high-risk. Progression-free survival was longer in the low-risk group (8·9 months [7·3–11·1]) than in the high-risk group (3·7 months [2·9–5·6]; p=0·0021; figure 3).

The overall response rate and disease control rate after combination ICI therapy were evaluated according to risk group. The overall response rate in the external validation cohort was 27·0% (78 of 289 patients) (table 3). The low-risk group had a significantly higher overall response rate than the high-risk group in both the overall survival (34·1% [78 of 229 patients]) *vs* 19·0% [44 of 229 patients]; p=0·047) and progression-free survival (33·1% [96 of 289 patients] *vs* 20·0% [58 of 289 patients]; p=0·026) models. Similarly, the disease control rate was significantly higher in the low-risk group than in the high-risk group in the overall survival (72·0% [165 of 229 patients] *vs* 52·4% [120 of 229 patients]; p=0·011) and progression-free survival (73·5% [212 of 289 patients] *vs* 55·4% [160 of 289 patients]; p=0·0043) models (table 3).

The most important features in the each of the three models comprising the final ensemble models for overall survival and progression-free survival were assessed using the permutation feature importance technique (appendix pp 12–13). For the overall survival model, AFP concentration, bilirubin concentration, and macrovascular invasion were among the top five most important features in each of the three component models of the final ensemble model, with albumin and alkaline phosphatase (ALP) concentrations in the top five for two of the component models. AFP and bilirubin concentrations were also among the top five most important features in each of the three components of the progression-free survival ensemble model, as well as the presence of lung metastases; haemoglobin concentration and the presence of two nodules were among the top five in two of the three components of the model.

The performance of the final ensemble models was evaluated in patients with Child–Pugh class A liver dysfunction. The overall survival model significantly outperformed clinical benchmark data in the external validation cohort (appendix pp 14–15). Both overall survival and progression-free survival models stratified patients in Child–Pugh class A into risk cohorts, with the low-risk group showing a significantly longer overall survival than the high-risk group in both the internal and external validation cohorts (p<0·0001 for both; appendix p 28).

Discussion

To our knowledge, this is the first global multicentre study to show that machine learning models can accurately predict overall survival, progression-free survival, and response in patients receiving atezolizumab and

bevacizumab for advanced hepatocellular carcinoma. These models, optimised through feature selection, allow for the prediction of clinical outcome using non-invasive, readily available, baseline clinical data. Through validation in an external international patient cohort, this study illustrates how machine learning can be used to stratify patients into different treatment-response groups, allowing for a precision-based approach.

Traditionally, linear-based machine learning models such as linear regression, logistic regression, and Cox regression have been used to identify predictors for outcomes of cancer therapy.[8–11] Although these approaches can highlight key predictors, they assume linear interactions between clinical variables. These models are also limited in their ability to process high-dimensionality data and are subject to high variance or overfitting.[22] Advanced machine learning models—such as random forest, XGBoost, naive Bayes, support vector machines, and neural networks—can interpret non-linear boundaries within datasets, which are potentially more representative of complex biological interactions.[23] Combining these non-linear models with feature selection methods can reduce the variance from model training in high-dimensional data, therefore improving the performance of the model on the validation set. Hindocha and colleagues[12] found that non-linear models, such as XGBoost and neural networks, showed superior performance to linear models in predicting overall survival after curative radiotherapy for non-small-cell lung cancer.[12] Deist and colleagues[24] showed that a classifier based on a random forest model outperformed linear approaches in predicting post-chemoradiotherapy survival across three cancer sites. Both of these studies combined advanced machine learning models with feature selection techniques, improving the performance of the models on validation sets.

Logistic regression and Cox regression approaches have identified markers for predicting the response to atezolizumab plus bevacizumab in patients with hepatocellular carcinoma, including AFP concentration,[9] CRAFITY score,[10] ALBI grade,[8] neutrophil-to-lymphocyte ratio, and platelet-to-lymphocyte ratio.[11] In our study, the best-performing models for predicting overall survival were support vector machine, neural network and naive Bayes, combined with Kendall's rank correlation, recursive feature elimination, and Pearson's correlation feature selection, respectively. These models can interpret non-linear relationships, and showed a significantly superior prediction compared with BCLC stage, neutrophil-to-lymphocyte ratio, platelet-to-lymphocyte ratio, and ALBI grade. Similarly, a combination of ridge regression and logistic regression with all features and naive Bayes with Pearson's correlation showed superior predictive power for progression-free survival compared with established clinical benchmarks, again using non-linear approaches. To our knowledge, only one previous study has used advanced machine learning models to predict response to ICI therapy in hepatocellular carcinoma.[25] In a cohort of 395 patients with hepatocellular carcinoma, six supervised models (logistic regression, LASSO regression, XGBoost, random forest, stochastic gradient boosting, and neural network) were used to predict 1-year mortality, on the basis of 48 baseline clinical variables, using forward and backward stepwise feature selection. LASSO regression and random forest classifiers outperformed CRAFITY score and ALBI grade on external validation, with an AUC greater than 0·90. We similarly report superior performance of machine learning models compared with ALBI grade and other clinical benchmarks for both overall survival and

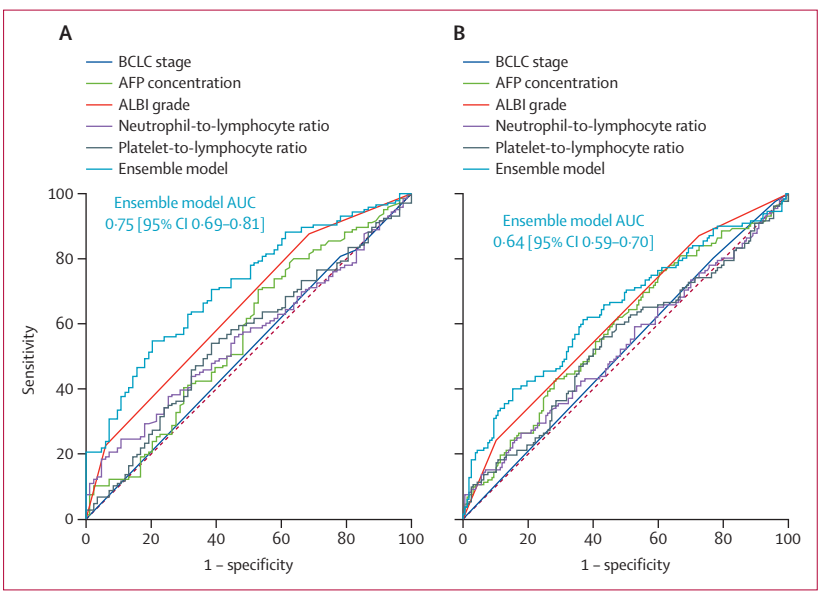

Figure 1: **Performance of clinical benchmark data and the ensemble machine learning models in predicting overall survival and progression-free survival**

Performance of existing methods using clinical benchmark data and the ensemble models in predicting overall survival (A) and progression-free survival (B) in the external validation cohort. AFP=α-fetoprotein. ALBI=albumin–bilirubin. AUC=area under the receiver operating characteristic curve. BCLC=Barcelona Clinic Liver Cancer.

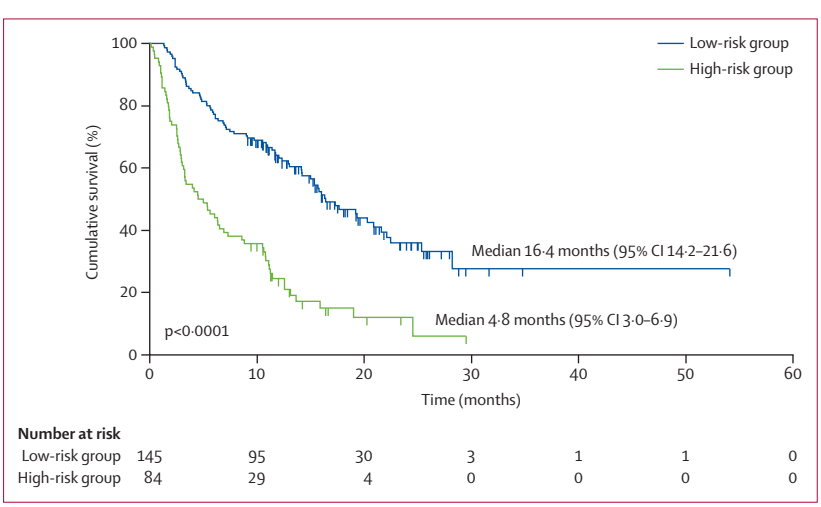

Figure 2: **Overall survival by model-stratified risk group**

Kaplan–Meier curves showing overall survival of patients in the external validation cohort, stratified into low-risk and high-risk groups by the machine learning ensemble model. The p value was calculated using a log-rank test.

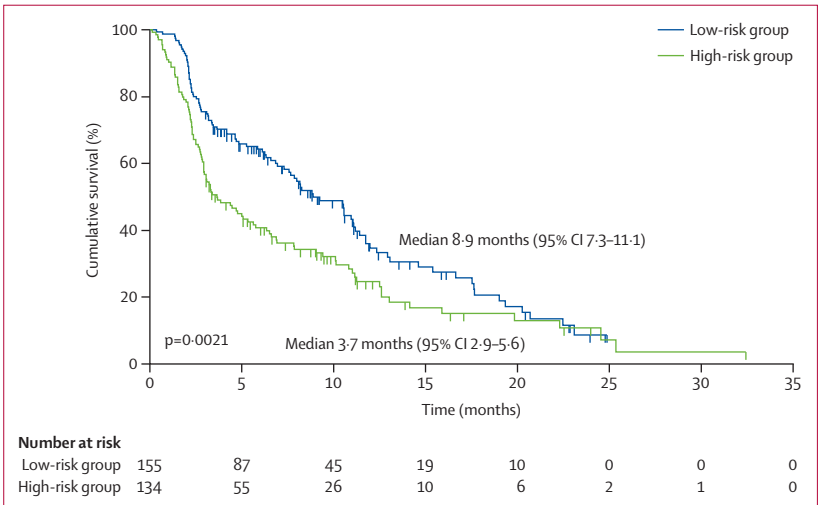

***Figure 3*: Progression-free survival by model-stratified risk group**
Kaplan–Meier curves showing progression-free survival of patients in the external validation cohort, stratified into low-risk and high-risk groups by the machine learning ensemble model. The p value was calculated using a log-rank test.

	Training cohort (n=339)	p value	Internal validation cohort (n=146)	p value	External validation cohort (n=289)	p value
Overall response rate						
All patients	28·9%	..	28·4%	..	27·0%	..
Overall survival model						
Low-risk group	35·5%	..	30·0%	..	34·1%	..
High-risk group	17·6%	0·0051	21·4%	0·51	19·0%	0·047
Progression-free survival model						
Low-risk group	38·9%	..	32·9%	..	33·1%	..
High-risk group	18·5%	0·0003	23·5%	0·61	20·0%	0·026
Disease control rate						
All patients	75·2%	..	66·0%	..	65·3%	..
Overall survival model						
Low-risk group	76·7%	..	78·3%	..	72·0%	..
High-risk group	63·5%	0·037	61·9%	0·10	52·4%	0·011
Progression-free survival model						
Low-risk group	83·0%	..	70·0%	..	73·5%	..
High-risk group	66·7%	0·0012	62·0%	0·41	55·4%	0·0043

p values comparing the overall response rate and disease control rate between low-risk and high-risk groups were calculated using χ^2 tests.

***Table 3*: Overall response rate and disease control rate in low-risk and high-risk groups, as stratified by the overall survival and progression-free survival models**

progression-free survival. The observed higher AUC performance in the previous study could be due to the type of ICI used: all patients in that study received nivolumab, pembrolizumab, or ipilimumab, in an off-label setting without standardisation of treatment protocols, as non-first line therapy in advanced hepatocellular carcinoma. Furthermore, all patients in the training, internal validation, and external validation cohorts of the previous study were from the same centre, in comparison with our study, in which patients from multiple international centres were included.

High-dimensional data can limit the performance of machine learning models, in particular on external validation datasets, owing to high training variance. Feature selection techniques use statistical methods to reduce input features to models, which can reduce variance and computational requirements.[26] Different feature selection approaches include filter, wrapper, and embedded methods.[27] These different approaches have specific strengths and weaknesses, and the optimal feature selection technique for specific tasks can be difficult to elucidate. Often, the choice of feature selection technique is dependent on the specific dataset and the machine learning models being implemented. We adopted a systematic approach to identify the optimal combinations of feature selection and machine learning models for predicting the response to atezolizumab plus bevacizumab in the training and internal validation datasets. Through this approach, we optimised the final machine learning models specifically for the prediction of overall survival and progression-free survival. For the overall survival prediction model, we found that a reduced feature set of Kendall's rank correlation, recursive feature elimination, and Pearson's correlation optimised the performance of support vector machine, neural network, and naive Bayes models, respectively. For the prediction of progression-free survival, we found two of the top three machine learning models—ridge regression and logistic regression—performed best without feature reduction, potentially due to the small number of baseline features included in our study. To overcome the so-called curse of dimensionality, the typical requirement for optimal model performance is at least five observations per feature.[28] With 44 baseline features in our training set of 339 patients, there could be sufficient observations for building machine learning models from the entire feature set. Furthermore, the intrinsic regularisation component in ridge regression might reduce training variance. Through the use of a final ensemble model, we used the relative strengths of machine learning and feature selection combinations to improve predictive performance.

Using permutations importance, we identified the key features driving predictions of overall survival and progression-free survival. Baseline bilirubin and albumin concentrations were consistently important for the prediction of overall survival. These parameters represent the synthetic function of the liver and, as part of the ALBI score,[8] predict survival in patients with hepatocellular carcinoma across different treatment methods, including atezolizumab plus bevacizumab.[29] We found that the final ensemble model for overall survival significantly outperformed ALBI grade in the prediction of mortality; this improved performance could result from the addition of AFP and ALP concentrations and macrovascular invasion to the model. AFP concentration and macrovascular invasion are established predictors for survival in patients with hepatocellular carcinoma and in response to ICI therapy.[30] The importance of ALP concentration in the overall survival prediction model could be due to both liver-specific and

cancer-specific mechanisms. ALP is expressed at high concentrations in liver and bone tissue. Increased ALP concentrations are associated with poor prognosis in hepatocellular carcinoma,[31] reflecting cholestasis and possible direct excretion by cancer cells or undiagnosed micrometastases.[32]

For progression-free survival, we found that haemoglobin concentration and the presence of multiple tumour nodules and lung metastases were important for model prediction. A multifocal tumour and metastases in the lung could be indicative of an aggressive tumour phenotype with a poor response to treatment. A low haemoglobin concentration is associated with increased mortality in patients with hepatocellular carcinoma,[33] and is an independent biomarker for predicting response to ICI therapy.[34] Anaemia-induced hypoxia can drive tumour proliferation and suppress T-cell regulation[35] in the tissue microenvironment, decreasing sensitivity to ICI therapy. Traditional machine learning models might have failed to incorporate these different prognostic factors into single models owing to non-linear relationships between them; by contrast, the algorithms used in the current study have identified predictive baseline factors that could be used for clinical prediction.

Our machine learning models can be applied directly to clinical practice. Data on baseline clinical and laboratory variables are routinely collected across international cohorts as part of standard care. Clinical predictions for overall survival and progression-free survival can be generated from these machine learning models using a simple medical calculator interface. As more clinical scoring systems have been developed, web-based and app-based medical calculators have been increasingly used by medical practitioners[36] and integrated into electronic health systems.[37] Data on baseline clinical variables can be input to stratify patients into low-risk and high-risk groups, enabling the identification of patients who are most likely to benefit from treatment with atezolizumab plus bevacizumab. A web-based interface for the current model is currently in development and will be made available for use by clinicians. Identifying patients who are predicted not to benefit from atezolizumab plus bevacizumab before the beginning of treatment enables the consideration of alternative treatment strategies, reducing toxicity from exposure.

Our study has several strengths. First, although data on participant race and ethnicity were not available, this study involved a large international cohort of patients with considerable variation in underlying liver disease aetiology across different countries. The validation of the performance of the models across different cohorts shows their robustness and generalisability for clinical practice. Compared with other studies that used models of lower complexity, we adopted a comprehensive approach to identify optimal combinations of machine learning and feature selection techniques. Through the use of ensemble learning, the final overall survival and progression-free survival models leveraged the different strengths of individual machine learning models and reduced the effects of uncertainties within the data. The use of permutation feature importance techniques enables identification of the key clinical parameters used for prediction. Such interpretability is paramount for acceptance by clinicians and patients, overcoming the so-called black-box concerns about artificial intelligence-based clinical decision support systems.[38]

Our study has limitations. We used data from the real-world use of atezolizumab plus bevacizumab and included patients with Child–Pugh class B liver dysfunction—a population that was not included in the original IMbrave 150 study. However, subsequent studies have shown that atezolizumab plus bevacizumab can improve survival outcomes in this group.[8,39,40] When restricted to patients with Child–Pugh class A liver dysfunction, our model retained predictive performance. Given the retrospective design of the study, there is a risk of collection and selection bias. Baseline variables affecting the response to ICI therapy,[10] such as CRP and creatinine concentrations, were excluded from the model owing to missing data. However, the cohort in this study is representative of real-world clinical practice, and it is important that the model has been trained and validated using features that are widely used and available across multiple cancer centres. Although the overall survival model showed moderate predictive performance, the performance of the progression-free survival model was lower. Future work incorporating additional data, such as radiomic and genomic data, could improve model performance. Our study does not have a control group of patients receiving an alternative treatment to atezolizumab plus bevacizumab; this limits the ability to elucidate the prognostic versus treatment-specific predictive effect of machine learning models, and the generalisability of the machine learning models for other ICI-based regimens is unclear. Further analysis of the models in patient cohorts receiving different ICI-based regimens would be informative. Although the international nature of the patient population is a strength of the study, it could also lead to heterogeneity between cohorts. However, principal component analysis showed no clear batch effect accounting for variation between the North American, European, and Asian cohorts. Indeed, the consistent performance of the overall survival and progression-free survival models across the different internal and external validation patient populations shows generalisability.

In conclusion, we developed and externally validated machine learning models to predict overall survival and progression-free survival in a real-world cohort of patients receiving atezolizumab and bevacizumab for hepatocellular carcinoma. These predictive models outperformed established predictors based on widely available baseline clinical data. The adoption of data-driven approaches could allow for personalised treatment selection and the stratification of patients with advanced hepatocellular carcinoma receiving atezolizumab plus bevacizumab.

Contributors

MV, RS, and EOA were involved in study conceptualisation and design. GFM, AD'A, CAMF, CCe, PL, AT, BSt, PRG, NN, BSc, FC, LW, CA, TUM, W-FH, C-YL, RP-TL, SVU, HA, NS, AK, ML, YIM, MSc, JvF, KS, HW, BW, CV, FS, GM, TP, AD, LN, MM, FP, Y-HH, P-CL, MPir, AP, NR, RT, MSi, EDG, BV, LG, AV, GC, CCa, HJC, MPin, AC, MK, VZ, LR, and DJP were involved in investigation, acquisition of data, and data curation. MV, RS, EOA, and DJP conducted the formal analysis and interpreted the data. MV, RS, EOA, and DJP wrote the original draft of the manuscript; all authors were involved in critical revision, reviewing, and editing of the manuscript for important intellectual content. MV and EOA conducted methodology, statistical analysis, data validation, and data visualisation. MV and EOA provided computing resources and software development. GFM, AD, CAMF, CCe, RS, DJP, and EOA provided project administration and technical and material support. RS, DJP, and EOA supervised the study. MV, GFM, CCe, DJP, EOA, and RS directly accessed and verified the data. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

MV received a grant from Cancer Research UK. GFM received support for attending meetings and travel from Roche. AD'A received educational support for congress attendance from Roche; consultancy fees from Roche and Chugai Pharmaceutical; and speaker fees from Roche, AstraZeneca, Eisai, and Chugai Pharmaceutical. CAMF received lecture fees from Eisai and AstraZeneca. CCe received travel support from Roche and speaker fees from Eisai and AstraZeneca. PL received a grant from a European Society for Medical Oncology fellowship. AT received a grant from Cancer Research UK. PRG received consultancy fees from Bayer, Boston Scientific, AstraZeneca, Adaptimmune, Bristol-Myers Squibb (BMS), Eisai, MSD, Sirtex, Lilly, Roche, and Guerbet; speaker fees for Bayer, BMS, Adaptimmune, Sirtex, Lilly, and Roche; and travel support from Bayer and Roche; and is on the advisory board for AstraZeneca, Eisai, BMS, and Roche. BSc received grant support from AstraZeneca, Eisai, and Ipsen; speaker honoraria from AstraZeneca and Eisai; and travel support from AstraZeneca, Ipsen, and Roche; and sits on the advisory board of AstraZeneca. FC received travel support from Onclive and the New York Celtic Medical Society. LW received a Winn Career Development Award and speaker fees from Curio Science and is on the advisory board for Partner Therapeutics. TUM received grants from Regeneron, Genentech, BMS, Merck, and Boehringer Ingelheim; received consulting fees from Avammune Therapeutics and Ono; and is on the advisory and/or data safety monitoring boards for Rockefeller University, AbbVie, Celldex, Regeneron, Merck, EMD Serono, BMS, Boehringer Ingelheim, Atara, AstraZeneca, Genentech, Chimeric Therapeutics, Glenmark, Simcere, Surface/Coherus, G1 Therapeutics, NGM Bio, DBV Technologies, Arcus, Fate Therapeutics, and Astellas Pharma. SVU received grants from AbbVie, Adlai Nortye, ArQule, AstraZeneca, Atreca, Boehringer Ingelheim, BMS, Celgene, Ciclomed, Erasca, Evelo Biosciences, Exelixis, ExThera Medical, G1 Therapeutics, GlaxoSmithKline (GSK), IGM Biosciences, Incyte, Isofol, Klus Pharma, Macrogenics, Merck & Co, Mersana Therapeutics, OncoMed Pharmaceuticals, Pfizer, Qurgen, Regeneron, Revolution Medicines, Synermore Biologics, Takeda, Tarveda Therapeutics, Tesaro, Tempest Therapeutics, and Vigeo Therapeutics to their institution; is on the advisory board for Elsai, AstraZeneca, and IGM Biosciences; and is on the data safety monitoring committee for the Targeted Agent & Profiling Utilization Registry (TAPUR) study by the American Society of Clinical Oncology. AK received grants from Roche/Genentech, BMS, Merck, AstraZeneca, Eisai, Exelixis, Adaptimmune, and Tvardi Therapeutics; consulting fees from Roche/Genentech, BMS, Merck, AstraZeneca, Eisai, and Exelixis; and speaker fees from Roche/Genentech, BMS, Merck, AstraZeneca, Eisai, and Exelixis. JvF has received speaking fees from AstraZeneca and is on the advisory board for Roche. KS received speaker fees from Roche and is on the advisory board for Roche, MSD, BMS, and AstraZeneca. CV received speaker fees from Roche, AstraZeneca, MSD, and Eisai; received travel support from Roche, AstraZeneca, and MSD; and is on the advisory board for AstraZeneca and MSD. FS received speaker fees from Eisai, MSD, AstraZeneca, and Daiichi Sankyo; travel support from LEO Pharma and Roche; and support at the University of Pisa from European Union - NextGenerationEU through the Italian Ministry of University and Research. TP received consulting fees from AstraZeneca and travel support from Roche. FP received speaker fees from AstraZeneca, Bracco, Esaote, Eisai, GE Healthcare Technologies, Gilead Sciences, Ipsen, MSD, Roche, Samsung, and Siemens Healthineers; is on the advisory board of AstraZeneca, BMS, Eisai, MSD, Nerviano Medical Sciences, Roche, and Siemens Healthineers; and has an unpaid role as President Elect of the Italian Society of Ultrasound of the Governing Board of the International Contrast Ultrasound Society. Y-HH received grants from Gilead Sciences and AstraZeneca and speaker fees from Roche, AstraZeneca, and Gilead Sciences. AP received consulting fees from AstraZeneca, Amgen, and MSD and travel support from Merck, Daiichi Sankyo, and Accord Healthcare. NR received funding from the German Research Foundation. MSi received consulting fees from AstraZeneca; speaker fees from MSD, Roche, and AstraZeneca; and travel support from AstraZeneca, MSD, and Roche. AV has received consulting fees and speaker fees from, and is on advisory boards for, Roche, AstraZeneca, Bohringer Ingelheim, Ipsen, Incyte, Cogent Technology, Eisai, Zymeworks, Biologix, BMS, Terumo, Elevar Therapeutics, Servier, MSD, Taiho Oncology, Jazz Pharmaceuticals, Medivir, AbbVie, Tyra Biosciences, Dr Falk Pharma, Janssen Pharmaceuticals, and Lilly; and has received travel support from Roche, MSD, and Astellas. CCa received speaker fees from Eisai, AstraZeneca, Roche, and Gilead Sciences. HJC received grants from Roche, BeiGene, Boryung, HK inno.N, Dong-A ST, and Hanmi Pharmaceutical; consulting fees from Roche, Bayer, BMS, AstraZeneca, MSD, Ono Pharma, Eisai, Sanofi, Aptamer Sciences, BeiGene, and Servier; speaker fees from Roche, BMS, Eisai, Sanofi, AstraZeneca, Bayer, Servier, and Dong-A ST; and travel support from Medivir. MPin received speaker honoraria from AstraZeneca, Bayer, BMS, Eisai, Lilly, MSD, and Roche; consulting fees for AstraZeneca, Bayer, BMS, Eisai, Ipsen, Lilly, MSD, and Roche; grants from AstraZeneca, Ipsen, BMS, Eisai, and Roche; and travel support from Bayer, BMS, Ipsen, and Roche; and has an unpaid role on the scientific committee and governing board of the European Association for the Study of the Liver. AC received grants from the International Association for the Study of Lung Cancer; consulting fees from MSD, OncoC4, Roche, Regeneron, BMS, Amgen, Daiichi Sankyo, AstraZeneca, Access Infinity, Ardelis Health, Alpha Sight, Capvision, Techspert, Atheneum, and Alira Health; speaker fees from Roche, Sanofi/Regeneron, AstraZeneca, MSD, BMS, Janssen Pharmaceuticals, and Pierre Fabre Pharmaceuticals; and travel support from Sanofi, MSD, and Roche. MK received grants from Otsuka Pharmaceutical, Taiho Oncology, Chugai Pharmaceutical, GE Healthcare Japan, and Eisai; consulting fees from Chugai Pharmaceutical, F Hoffmann-La Roche, Eisai, AstraZeneca, and MSD; and speaker fees from Chugai Pharmaceutical, Eisai, and AstraZeneca. LR received consulting fees from, and is on advisory boards for, AbbVie, AstraZeneca, Basilea Pharmaceutica, Bayer, BMS, Eisai, Elevar Therapeutics, Exelixis, Genenta Science, Jiangsu Hengrui, Incyte, Ipsen, Jazz Pharmaceuticals, MSD, Nerviano Medical Sciences, Roche, Servier, Taiho Oncology, and Zymeworks; received lecture fees from AstraZeneca, Bayer, BMS, Guerbet, Incyte, Ipsen, Roche, and Servier; received travel expenses from AstraZeneca and Servier; received research grants (to their institution) from AbbVie, AstraZeneca, BeiGene, Exelixis, FibroGen, Incyte, Ipsen, Jazz Pharmaceuticals, MSD, Nerviano Medical Sciences, Roche, Servier, Taiho Oncology, TransThera Sciences, and Zymeworks; and has an unpaid leadership role as the Head of External Relations for the International Liver Cancer Association (ILCA), a second role with the ILCA as a member of the executive committee and the governing board, is Chair of the European Organisation for Research and Treatment of Cancer's Gastrointestinal Tract Cancer Group – Hepatopancreatobiliary and Neuroendocrine Tumours Task Force, and is a Special Expert for the International Trials Europe on the Hepatobiliary Task Force of the National Cancer Institute's Gastrointestinal Steering Committee. DJP received grants from BMS, GSK, and MSD; consulting fees from Mina Therapeutics, Eisai, Roche, Avammune, Da Volterra, Mursla Bio, H3B, Ipsen, Boston

Scientific, Starpharma, Exact Sciences, and AstraZeneca; speaker fees from Roche, BMS, Eisai, and Boston Scientific, and travel support from Roche, BMS and MSD; and is on the advisory board for Mina Therapeutics, EISAI, Roche, Avammune, Da Volterra, Mursla Bio, H3B, Ipsen, LIfT Biosciences, Exact Sciences, and AstraZeneca. RS received speaker fees from Eisai and Roche; travel support from Novartis; and research funding to their institution from Novartis, Boston Scientific, and Terumo Europe; and is on an appraisal board for the National Institute for Health and Care Excellence. All other authors declare no competing interests.

Data sharing

All data and analysis methods, including source code, will be made available upon request to the corresponding author.

Acknowledgments

MV and RS are supported by a grant from Cancer Research UK. FS received support at University of Pisa from European Union - NextGenerationEU through the Italian Ministry of University and Research under PNNR - M4C2-I1.3 Project PE_00000019 "HEAL ITALIA" CUP I53C22001440006. PL acknowledges support from the European Society for Medical Oncology fellowship.

References

1 Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021; **71:** 209–49.

2 Park JW, Chen M, Colombo M, et al. Global patterns of hepatocellular carcinoma management from diagnosis to death: the BRIDGE study. *Liver Int* 2015; **35:** 2155–66.

3 Finn RS, Qin S, Ikeda M, et al. Atezolizumab plus bevacizumab in unresectable hepatocellular carcinoma. *N Engl J Med* 2020; **382:** 1894–905.

4 Abou-Alfa GK, Chan SL, Kudo M, et al. Phase 3 randomized, open-label, multicenter study of tremelimumab (T) and durvalumab (D) as first-line therapy in patients (pts) with unresectable hepatocellular carcinoma (uHCC): HIMALAYA. *J Clin Oncol* 2022; **40:** 379.

5 Galle PR, Decaens T, Kudo M, et al. Nivolumab (NIVO) plus ipilimumab (IPI) vs lenvatinib (LEN) or sorafenib (SOR) as first-line treatment for unresectable hepatocellular carcinoma (uHCC): first results from CheckMate 9DW. *J Clin Oncol* 2024; **42:** LBA4008–4008.

6 Qin S, Chan SL, Gu S, et al. Camrelizumab plus rivoceranib versus sorafenib as first-line therapy for unresectable hepatocellular carcinoma (CARES-310): a randomised, open-label, international phase 3 study. *Lancet* 2023; **402:** 1133–46.

7 Marrero JA, Kulik LM, Sirlin CB, et al. Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the Study of Liver Diseases. *Hepatology* 2018; **68:** 723–50.

8 Storandt MH, Zemla TJ, Patell K, et al. Atezolizumab plus bevacizumab as first-line systemic therapy for hepatocellular carcinoma: a multi-institutional cohort study. *Oncologist* 2024; **29:** 986–96.

9 Teng W, Lin C-C, Ho M-M, et al. Alpha-fetoprotein response at different time-points is associated with efficacy of nivolumab monotherapy for unresectable hepatocellular carcinoma. *Am J Cancer Res* 2021; **11:** 2319–30.

10 Scheiner B, Pomej K, Kirstein MM, et al. Prognosis of patients with hepatocellular carcinoma treated with immunotherapy—development and validation of the CRAFITY score. *J Hepatol* 2022; **76:** 353–63.

11 Dharmapuri S, Özbek U, Jethra H, et al. Baseline neutrophil–lymphocyte ratio and platelet–lymphocyte ratio appear predictive of immune treatment related toxicity in hepatocellular carcinoma. *World J Gastrointest Oncol* 2023; **15:** 1900–12.

12 Hindocha S, Charlton TG, Linton-Reid K, et al. A comparison of machine learning methods for predicting recurrence and death after curative-intent radiotherapy for non-small cell lung cancer: development and validation of multivariable clinical prediction models. *EBioMedicine* 2022; **77:** 103911.

13 Fulgenzi CAM, Cheon J, D'Alessio A, et al. Reproducible safety and efficacy of atezolizumab plus bevacizumab for HCC in clinical practice: results of the AB-real study. *Eur J Cancer* 2022; **175:** 204–13.

14 Reig M, Forner A, Rimola J, et al. BCLC strategy for prognosis prediction and treatment recommendation: the 2022 update. *J Hepatol* 2022; **76:** 681–93.

15 van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw* 2011; **45:** 1–67.

16 Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; **45:** 228–47.

17 Murugesu S, Linton-Reid K, Braun E, et al. Predicting outcomes of expectant and medical management in early pregnancy miscarriage using machine learning to develop and validate multivariable clinical prediction models. *BMC Pregnancy Childbirth* 2025; **25:** 225.

18 Giannini EG, Bucci L, Garuti F, et al. Patients with advanced hepatocellular carcinoma need a personalized management: a lesson from clinical practice. *Hepatology* 2018; **67:** 1784–96.

19 Lin C-W, Chen Y-S, Lo G-H, et al. Resubclassification and clinical management for Barcelona Clinic Liver Cancer Stage C hepatocellular carcinoma. *Hepatol Int* 2021; **15:** 946–56.

20 Breiman L. Random forests. *Mach Learn* 2001; **45:** 5–32.

21 Satopaa V, Albrecht J, Irwin D, Raghavan B. Finding a 'kneedle' in a haystack: detecting knee points in system behavior. In: 2011 31st International Conference on Distributed Computing Systems Workshops. IEEE, 2011: 166–71.

22 Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction, 2nd edn. Springer, 2009.

23 Lo-Thong-Viramoutou O, Charton P, Cadet XF, et al. Non-linearity of metabolic pathways critically influences the choice of machine learning model. *Front Artif Intell* 2022; **5:** 744755.

24 Deist TM, Dankers FJWM, Valdes G, et al. Machine learning algorithms for outcome prediction in (chemo)radiotherapy: an empirical comparison of classifiers. *Med Phys* 2018; **45:** 3449–59.

25 Lui TKL, Cheung KS, Leung WK. Machine learning models in the prediction of 1-year mortality in patients with advanced hepatocellular cancer on immunotherapy: a proof-of-concept study. *Hepatol Int* 2022; **16:** 879–91.

26 Yu L, Liu H. Efficient feature selection via analysis of relevance and redundancy. *J Mach Learn Res* 2004; **5:** 1205–24.

27 Pudjihartono N, Fadason T, Kempa-Liehr AW, O'Sullivan JM. A review of feature selection methods for machine learning-based disease risk prediction. *Front Bioinform* 2022; **2:** 927312.

28 Koutroumbas K, Theodoridis S. Pattern recognition, 4th edn. Elsevier, 2008.

29 Kudo M, Finn RS, Cheng A-L, et al. Albumin–bilirubin grade analyses of atezolizumab plus bevacizumab versus sorafenib in patients with unresectable hepatocellular carcinoma: a post hoc analysis of the phase III IMbrave150 study. *Liver Cancer* 2023; **12:** 479–93.

30 Kai M, Hikita H, Kazuki M, et al. Clinical factors associated with the therapeutic efficacy of atezolizumab plus bevacizumab in patients with unresectable hepatocellular carcinoma: a multicenter prospective observational study. *PLoS One* 2024; **19:** e0294590.

31 Li Q, Lyu Z, Wang L, Li F, Yang Z, Ren W. Albumin-to-alkaline phosphatase ratio associates with good prognosis of hepatitis b virus-positive HCC patients. *OncoTargets Ther* 2020; **13:** 2377–84.

32 Mori K, Janisch F, Parizi MK, et al. Prognostic value of alkaline phosphatase in hormone-sensitive prostate cancer: a systematic review and meta-analysis. *Int J Clin Oncol* 2020; **25:** 247–57.

33 Meischl T, Balcar L, Park Y-R, et al. Anaemia is independently associated with mortality in patients with hepatocellular carcinoma. *ESMO Open* 2024; **9:** 103593.

34 Huang Y, Wei S, Jiang N, et al. The prognostic impact of decreased pretreatment haemoglobin level on the survival of patients with lung cancer: a systematic review and meta-analysis. *BMC Cancer* 2018; **18:** 1235.

35 Zhao L, He R, Long H, et al. Late-stage tumors induce anemia and immunosuppressive extramedullary erythroid progenitor cells. *Nat Med* 2018; **24:** 1536–44.

36 Green TA, Whitt S, Belden JL, Erdelez S, Shyu C-R. Medical calculators: prevalence, and barriers to use. *Comput Methods Programs Biomed* 2019; **179:** 105002.

37 Chandra S, Agarwal D, Hanson A, et al. The use of an electronic medical record based automatic calculation tool to quantify risk of unplanned readmission to the intensive care unit: a validation study. *J Crit Care* 2011; **26:** 634.e9–15.

38 Stiglic G, Kocbek P, Fijacko N, Zitnik M, Verbert K, Cilar L. Interpretability of machine learning-based prediction models in healthcare. *WIREs Data Mining Knowl Discov* 2020; **10:** e1379.

39 D'Alessio A, Fulgenzi CAM, Nishida N, et al. Preliminary evidence of safety and tolerability of atezolizumab plus bevacizumab in patients with hepatocellular carcinoma and Child–Pugh A and B cirrhosis: a real-world study. *Hepatology* 2022; **76:** 1000–12.

40 Fulgenzi CAM, Scheiner B, D'Alessio A, et al. Immunotherapy vs best supportive care for patients with hepatocellular cancer with Child–Pugh B dysfunction. *JAMA Oncol* 2024; **10:** 1253–58.