Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predicting Immunotherapy Response in Patients with Hepatocellular Carcinoma from Clinical and Textual Features Using AI Techniques

Anwaar Saeed
UPMC Hillman Cancer Center
Thant Hoe
Curenetics
Kilsi Kobani
Curenetics
Rachelle Atchina
Curenetics
Maren Buk
Curenetics
Virginia Yuk-chun Lam
Curenetics
Meghana Singh
UPMC Hillman Cancer Center
Yuming Shi
UPMC Hillman Cancer Center
Alireza Tojjari
UPMC Hillman Cancer Center
Vaishnavi Balaji
Curenetics
Lakshya Sharma
Curenetics
Yuxi Zhang
Curenetics
Sola Adeleke
sola@curenetics.io

Curenetics

Research Article

Keywords: hepatocellular carcinoma, immunotherapy, machine learning, natural language processing, predictive modeling, immune checkpoint inhibitors

Posted Date: May 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9557672/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Immunotherapy (IO) improves survival in advanced hepatocellular carcinoma (HCC), yet under 30% of patients respond to treatment. Existing biomarkers have shown limited predictive accuracy. Machine learning (ML) techniques could be used to develop prediction models that support personalised treatment. We developed and evaluated machine learning models that predict response to IO in patients with HCC. We retrospectively analyzed data from 298 patients with HCC treated with immunotherapy at the University of Pittsburgh Medical Center (UPMC) Hillman Cancer Center between December 2014 and December 2023. Of the 298 patients, 215 (71%) had stable disease and 87 (29%) had progression. The best-performing model was a late-fusion model with text and tabular features (AUC 0.78, Precision: 0.76, Recall 0.70). Good performance was retained even when restricted to tabular variables (AUC 0.71, Precision: 0.70, Recall 0.70). Key predictors of response included lower alpha-fetoprotein (AFP), liver function tests within normal range (AST, ALT, ALP, albumin, bilirubin), higher total protein and lower grade of ECOG performance status. 142 patients had first-line IO treatment with atezolizumab and bevacizumab (Atezo/Bev) and 57 patients had durvalumab and tremelimumab (Durva/Treme). The model performed well for both treatment groups (Atezo/Bev: AUC 0.78, Durva/Treme: AUC 0.69). This study demonstrates that ML models integrating both clinical and NLP-derived features can accurately predict IO response in patients with HCC. Future work will externally validate these results on larger datasets, with the aim of developing generalizable and clinically useful predictive models.

Introduction

The incidence of liver cancer increased by 75% between 1990 and 2015 [1]. The World Health Organization predicts that more than 1.3 million people will die from liver cancer in 2040 [2]. Hepatocellular carcinoma (HCC) is the most common primary malignancy of the liver, accounting for approximately 85–90% of cases [3]. The main risk factors for HCC include chronic hepatitis B or C virus infection, alcohol misuse, obesity, type 2 diabetes, and smoking [4]. Most patients with HCC are diagnosed at an advanced stage, at which point surgical resection, ablation and liver transplantation are no longer feasible [5]. Concurrent underlying liver disease often limited treatment options due to deranged liver function [6], consequently, survival for advanced HCC remains poor, with a 5-year survival rate of less than 20% [7].

Immune checkpoint inhibitors (ICIs) have emerged as a transformative treatment modality in oncology, leading to improved survival outcomes in several malignancies [8]. In HCC, the combination of atezolizumab, a programmed death-ligand 1 (PD-L1) inhibitor, and bevacizumab, an anti-angiogenic agent (Atezo/Bev), and the combination of durvalumab, a PD-L1 inhibitor, and tremelimumab, a cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) inhibitor (Durva/Treme), has demonstrated superiority over sorafenib in patients with unresectable HCC and has since become the first-line standard-of-care treatment for advanced HCC (aHCC) [9] [10]. Although treatment with Atezo/Bev has resulted in significant improvements in overall survival, 20% of patients are refractory to Atezo/Bev, with only 20–30% radiological response rate [11]. Substantial adverse effects including gastrointestinal bleeding,

hypertension as well as immune-related adverse events which can be debilitating, irreversible and life-threatening [12]. Currently, the most widely used approach for predicting response to immune checkpoint blockade is PD-L1 expression assessed by immunohistochemistry (IHC) [13] However, PD-L1 IHC has demonstrated limited predictive accuracy across tumor types, including HCC. Response rates among

PD-L1–positive patients remain low (26%), while responses are occasionally observed in PD-L1–negative patients (18%) [14]. In addition, PD-L1 testing requires invasive tissue sampling, which may be challenging in patients with advanced disease and carries procedural risks. Given the toxicity profile, high cost, and limited efficacy of immunotherapy in unselected patient populations, there is an urgent unmet need to identify accurate and robust predictive tools to identify patients who are mostly likely to have favorable treatment response.

Machine learning (ML) offers a promising framework to address these challenges by enabling the integration of high-dimensional and heterogeneous clinical data. ML models are capable of capturing complex, non-linear relationships that are difficult to model using traditional statistical approaches [15]. In oncology, ML has been applied to outcome prediction, treatment response modeling, and patient stratification using diverse data sources, including clinical variables, imaging, pathology, and unstructured electronic health record data. Importantly, ML-based approaches may facilitate the development of non-invasive, data-driven predictors of immunotherapy response that extend beyond single biomarkers [16]. Our ML and explainable artificial intelligence (AI) models have shown to predict ICI response, immune related adverse events and subsequent admission in patients with melanoma [17]–[20]. A previous study demonstrated higher inter-model agreement among large language models (LLMs) compared with physician predictions in assessing immunotherapy response in HCC, albeit with moderate accuracy and sensitivity [21]. The aim of the study is to use our validated ML models to predict ICIs response among patients with aHCC treated with ICIs. It is hoped that identifying patient and clinical features associated with favorable responses to immune checkpoint inhibitors (ICIs) will support and inform clinical decision-making.

Methods

Patient Cohort

Electronic health record data was collected from 302 patients with a diagnosis of hepatocellular cancer who were being treated at the University of Pittsburgh Medical Centre (UPMC) Hillman Centre between December 2014 and December 2023. For each patient, this included demographic data (age, gender, occupation), clinical data such as laboratory measurements (e.g. liver enzymes, serum proteins, coagulation studies), tumour staging and genomic testing results, treatment plan (inclusive of different immunotherapy regimens, current and past adjuvant treatments and neoadjuvant treatments), biopsy status, clinical text in imaging (e.g. magnetic resonance imaging, computer tomography) and histopathology reports at diagnosis and at start of immunotherapy, other relevant clinical data such as

presenting complaint of the cancer diagnosis, past medical history, exposure to hepatotoxic chemicals such as alcohol, cocaine, heroin and marijuana, Child-Pugh scores and Eastern Cooperative Oncology Group (ECOG) performance status were included, contributing to a total of 77 data points. All patient data was de-identified and anonymized.

Outcome of Interest

The outcome of interest for this study was response to immunotherapy. A binary treatment response outcome was derived by categorizing radiological response categories (complete response, partial response, stable disease, progressive disease) into responder and non-responder classes [22]. Patients who had complete or partial response were classified as responders and those who had stable disease or progressive disease were classified as non-responders.

Data Preprocessing

All data processing and machine learning model development was done in Google CoLab (RRID:SCR_018009) using Python version 3.12.12 (RRID:SCR_008394). Columns with more than 80% missing values were removed [23], [24]. Irrelevant clinical identifiers, dates, free-text administrative fields, and post-treatment outcome variables that could imply response and hence introduce data leakage were excluded. Outliers were assessed visually using boxplots and removed if they were physiologically implausible and likely due to measurement or data entry errors (e.g., alpha-fetoprotein > 800,000 IU/mL).

Past medical history and comorbidities were transformed into binary features indicating the presence or absence of a condition using keyword-based extraction, this will help to identify significant past medical history such as liver cirrhosis, viral hepatitis, metabolic liver disease, hypertension, diabetes and cardiovascular disease which are risk factors of HCC.

To reduce multicollinearity and model instability, liver function tests and cancer staging variables with high correlation were excluded, and correlation matrices and variance inflation factors (VIFs) were assessed. Variables with sparse events or low prevalence were removed to ensure stable coefficient estimation. Treatment variables, including local therapies and immunotherapy regimens, were standardized and one-hot encoded based on predefined drug and procedure groupings.

Of the remaining columns, those containing free-text (‘notes', 'presenting_complaints',

'physical_examination_findings', 'ct_report_at_diagnosis', 'mri_report_at_diagnosis', 'ct_report_around_the_start_of_immunotherapy',

'mri_report_at_start_of_immunotherapy', 'histopathological_features') were combined into one column with section markers named ‘full_text’. After removal of duplicated variables, the final dataset contained

54 features. Some patient entries were also deleted due to insufficient variables, resulting in a final cohort of 298 patients.

The combined clinical free-text data was preprocessed using a custom lemmatization and tokenization pipeline. Text was first tokenized using a regular expression that captures alphanumeric terms and decimal numbers, then lemmatized with the WordNet lemmatizer. Stop words were removed using the standard scikit-learn English stop word list, with negation terms (e.g., no, not, never, none) explicitly retained to preserve clinically relevant polarity.

Training and Validation set

The dataset was randomly split into 80% for training and 20% for validation by number of patients, and stratified by outcome. Within the training set, 5-fold cross validation was performed for comparison before and after hyperparameter tuning, then against the holdout validation set.

Machine Learning Prediction Pipeline

Feature selection was performed using SelectKBest with Analysis of Variance (ANOVA) F-tests to identify and retain the top 20 predictive non-text features to reduce model complexity and dimensionality [25]. Multiple classification algorithms such as Lasso logistic regression, random forest, XGBoost, and support vector machine were implemented, with class imbalance addressed via weighting [26], [27], [28]. Hyperparameter tuning was conducted for all models and after tuning, models were evaluated using stratified train-test splits and cross-validation with ROC-AUC as the primary metric. The best-performing model was further interpreted using Shapley Additive Explanations (SHAP) values to identify feature importance across the overall cohort and within treatment-specific subgroups (Atezo/Bev and Durva/Treme) [8], [29], 30].

Statistical comparisons between treatment groups were conducted using Mann-Whitney U tests for continuous variables, Chi-squared tests for categorical variables, and t-tests for predicted response probabilities, providing interpretability and insight into subgroup treatment effects. The final optimized model was saved for downstream deployment and further analysis.

Natural Language Processing

Processed tokens were rejoined into normalized text strings and transformed into numerical feature representations using term frequency–inverse document frequency (TF-IDF) vectorization [31]. Both unigram and bigram representations were evaluated, and vocabulary size was controlled by varying the maximum number of retained features to reduce overfitting in small clinical datasets.

Logistic regression, random forest, support vector machine, and XGBoost classifiers were trained. Model selection and hyperparameter optimization were performed using five-fold cross-validated grid search.

The final model was selected based on the test set ROC-AUC performance.

Late-Fusion Strategy

A late-fusion approach was used to combine predictions from the tabular and text-based models at the probability level. Final class probabilities were computed as a weighted average of the two model outputs, with the fusion weight (α) optimized by sweeping values from 0 to 1 and selecting the value that maximized ROC-AUC on the test set.

Logistic Regression

A multivariable logistic regression model was developed to evaluate clinical predictors of immunotherapy response in patients with aHCC. The cleaned dataset was imported into R version 4.5.2 using RStudio (RRID:SCR_000432). Variables were converted to factors or retained as numeric based on their distribution and number of unique values.

A full logistic regression model with a binomial link function was initially fitted, adjusting for potential confounders. Model refinement was performed by excluding non-significant and collinear predictors, followed by refitting the final model. Adjusted odds ratios (ORs) with 95% confidence intervals were calculated by exponentiating model coefficients. Statistical significance was defined as a two-sided p-value < 0.05. Model results were summarized using tidy model outputs, and significant predictors were visualized using forest plots.

Results

A total of 298 patients with HCC receiving immunotherapy were included in the final analysis after data cleaning. The mean age was similar between groups (69.0 ± 8.0 years in responders vs 68.4 ± 10.1 years in non-responders).The majority of patients were male, in both responders (85%) and non-responders (80.2%). 167 (56.0%) achieved a response and 131 (44.0%) were classified as non-responders. Responders were more likely to have better baseline performance status, with 41.3% having an ECOG score of 0 compared with 21.4% among non-responders; while poor performance status (ECOG 2– 3) was more frequent in non-responders (22.1% vs 7.2%). Liver function differed notably between groups. A greater proportion of responders had Child–Pugh A status (83.2% vs 66.4%), and responders, on average, had higher albumin levels (3.5 ± 0.6 vs 3.2 ± 0.7 g/dL) and lower total bilirubin (1.2 ± 1.1 vs 1.8 ± 2.7 mg/dL). Markers of tumor burden and cholestasis were also lower in responders. AFP levels were substantially lower in responders (mean = 5,160.7 ± 29,463.3 ng/mL) compared with non-responders (mean =

18,646.5 ± 56,254.8 ng/mL), as were ALP levels (mean = 153.9 ± 101.0 vs 211.4 ± 146.5 U/L) (Table 1).

With respect to treatment, responders were more likely to have received Atezo/Bev (55.1% vs 38.2%), whereas the proportion receiving Durva/Treme was similar between groups (18.0% vs 20.6%).

Variable	Overall (n = 298)	Non-response (n = 131)	Response (n = 167)	SMD*	P-value
Age, mean (SD†)	68.7 (8.9)	68.4 (10.1)	69.0 (8.0)	0.06	0.611
Sex, n (%)					
Female	51 (17.1)	26 (19.8)	25 (15.0)	0.129	0.34
Male	247 (82.9)	105 (80.2)	142 (85.0)		
ECOG‡ performance status, n (%)				0.553	< 0.001
0	97 (32.6)	28 (21.4)	69 (41.3)		
1	160 (53.7)	74 (56.5)	86 (51.5)		
2–3	41 (13.8)	29 (22.1)	12 (7.2)		
Child–Pugh score, n (%)				0.428	0.001
A	226 (75.8)	87 (66.4)	139 (83.2)		
B–C	72 (24.2)	44 (33.6)	28 (16.8)		
Albumin, mean (SD)	3.4 (0.6)	3.2 (0.7)	3.5 (0.6)	0.421	< 0.001
Total bilirubin, mean (SD)	1.5 (2.0)	1.8 (2.7)	1.2 (1.1)	-0.334	0.007
AFP§, mean (SD)	11089.0 (43764.5)	18646.5 (56254.8)	5160.7 (29463.3)	-0.3	0.014
ALP¶, mean (SD)	179.2 (126.1)	211.4 (146.5)	153.9 (101.0)	-0.457	< 0.001
Treatment regimen, n (%)				0.344	0.005
Atezo/Bev	142 (47.7)	50 (38.2)	92 (55.1)		
Durva/Treme 57 (19.1) 27 (20.6) 30 (18.0)					

*Standardised mean difference, †Standard deviation, ‡Eastern Cooperative Oncology Group, §Alpha-fetoprotein, ¶Alkaline phosphatase

Table 1: Baseline Characteristics of Patients by Response Status

Multivariable Logistic Regression Analysis

The multivariable logistic regression model demonstrated improved fit compared with the null model (residual deviance = 224.6 vs. null deviance = 267.5; AIC = 270.6).

Performance status was strongly associated with treatment response. Patients with poorer functional status (performance status 2–3) had significantly lower chance of achieving response compared with those with performance status 0 (OR = 0.21, 95% CI: 0.06–0.70, p = 0.011). A weaker trend was observed for performance status 1, though this did not reach statistical significance. Meeting the Milan criteria for liver transplantation was 3 times more likely to respond to ICIs (OR = 3.90, 95% CI: 1.01–15.0, p = 0.049).

AFP and disease stage at diagnosis, showed borderline associations (AFP: $p = 0.067$; stage III vs. stage I: $p = 0.053$), but did not reach statistical significance. No statistically significant associations were observed for age, sex, liver function measures (albumin, bilirubin, INR), Child-Pugh class, viral hepatitis status, cirrhosis history, or treatment regimen (Durva/Treme vs. Atezo/Bev).

Machine Learning Pipeline

Among the tabular models, the random forest classifier achieved the highest discriminative performance on the held-out test set, with a ROC-AUC of 0.71, outperforming support vector machines (0.69), XGBoost (0.68), and logistic regression (0.57) (Fig. 2a). Feature importance analysis using SHAP values identified key clinical variables contributing to model predictions (Fig. 3). Key predictors of response included lower alpha-fetoprotein (AFP), liver function tests within normal range (AST, ALT, ALP, albumin, bilirubin), higher total protein and lower grade of ECOG performance status. 142 patients had first-line IO treatment with Atezo/Bev and 57 patients had durvalumab and tremelimumab Durva/Treme.

For the text-based TF-IDF models, the support vector machine achieved the best performance following tuning, with a test ROC-AUC of 0.77, followed by logistic regression (0.75), random forest (0.66), and XGBoost (0.60). The most informative textual features included numeric tokens and clinically relevant imaging and disease-related terms, such as ‘arterial’, ‘carcinoma’, ‘hepatic’, ‘enhancement’, and ‘cirrhotic’, reflecting the importance of radiologic and diagnostic language in model discrimination.

Trained on a) Tabular Features Only, and b) TF-IDF Text Features Only

Comparison of Model Performance on Treatment Regimens

Subgroup analysis showed that model performance for patients receiving Atezo/Bev (test

AUC-ROC 0.78) was slightly superior to those receiving Durva/Treme (test AUC-ROC

0.69). There was no statistically significant difference among categorical and continuous variables between the two treatment groups ($p > 0.05$).

The model's predicted risk scores did not differ significantly by treatment regimen. The predicted mean response for the Atezo/Bev group was 0.548 and 0.531 for the Durva/Treme group with a t-test statistic of 0.336 ($p = 0.739$).

Late fusion of the best-performing tabular (random forest) and text-based (SVM) models further improved predictive performance. The optimal fusion weight favored the text modality, with 90% contribution from the text model and 10% from the tabular model, yielding a combined ROC-AUC of 0.78.

Discussion

In this study, we demonstrated how our advanced machine learning models can accurately predict immunotherapy response between responders and non-responders in patients with aHCC treated with ICIs with high predicted probability of AUC of 0.78 in our best predicted model. We also determined the key predictors of treatment efficiency and progression. Among all the models evaluated, the late-fusion model incorporating a random forest classifier for tabular clinical features and a support vector machine classifier for text-derived features achieved the best predictive performance. Key predictors of response included lower AFP levels, liver function tests within normal ranges (AST, ALT, ALP, albumin, and bilirubin), higher total protein levels, and better ECOG performance status. Our model was especially valuable in supporting clinical decisionmaking for patients who are at the borderline of eligibility for immunotherapy. By identifying individuals most likely to benefit, these models may help minimize unnecessary exposure to potentially toxic treatments while optimizing therapeutic outcomes.

Alpha-fetoprotein (AFP) is a widely used serum biomarker for assessing treatment response in hepatocellular carcinoma (HCC). Our model confirmed the predictive value of AFP in predicting ICIs response, this is consistent with a meta-analysis of 131 studies showing patients with an AFP response had significantly higher objective radiological response rates [32]. A Chinese study further demonstrated that an early decline in AFP from baseline to the first follow-up (i.e., 4–6 weeks after treatment initiation) was associated with improved overall survival (OS) and progression-free survival (PFS) [33]. Similarly, liver enzymes, which proved to be indicators of hepatic functional reserve, were significant predictors of clinical outcomes. Tian et al. showed that impaired liver function, reflected by higher Child–Pugh or ALBI scores, was associated with poorer prognosis in patients receiving ICIs [34]. In contrast, the predictive value of PD-L1 expression remains unclear. In our model, PD-L1 levels were not associated with response to ICIs, consistent with findings from the KEYNOTE-224 and CheckMate-040 studies [35], [36].

By incorporating routinely available laboratory tests and demographic data as well as clinical text from clinician notes and investigation reports, this study demonstrates that immunotherapy response can be predicted using simple clinical inputs without the need for sophisticated biomolecular testing. This is particularly important given that these laboratory tests are inexpensive, widely available, and easily implemented in routine clinical practice, including in low-resource settings.

Our findings further support the growing body of evidence demonstrating that machine learning approaches leveraging routinely collected clinical data can aid in predicting response to immunotherapy and in identifying patients at increased risk of disease progression. Artificial intelligence (AI) and machine learning (ML) are increasingly transforming patient care by enabling precision medicine and personalised oncology treatment strategies [37]. These methods can efficiently analyse and integrate large volumes of multimodal clinical data, providing timely and clinically actionable predictions [38, 39]. In addition, AI and ML models are capable of identifying subtle or latent patterns within complex datasets that may be predictive of patient response to immune checkpoint inhibitors (ICIs) [40]. Importantly, AI and ML models can iteratively improve during the training process by adjusting model parameters in response to prediction errors. Through repeated exposure to diverse patient profiles and clinical outcomes, these models refine their representations of complex, non-linear relationships among clinical, molecular, and treatment variables, thereby enhancing model generalisability and predictive accuracy [41].

The main strength of our study lies in the systematic evaluation of five machine learning models to identify the optimal predictive approach. In addition, we specifically identified the most relevant clinical features and multimodal late-fusion pipeline to extract meaningful clinical data from unstructured text, thereby enhancing data input quality. To our knowledge, this study incorporates the largest patient numbers and clinical data points than prior work, contributing to improved predictive accuracy.

We acknowledged one of the limitations of this study is being a single-center retrospective study with inherent biases in patient selection. However, our study is the largest to specifically assess immunotherapy response to aHCC using patients' clinical characteristics. Additionally, the number of patients who received the Atezo/Bev and Durva/Treme combinations was insufficient to detect clinically meaningful differences between treatment groups or to support robust subgroup analyses for generalizable clinical inference. Larger, more diverse cohorts are required to validate these findings across immune checkpoint–based combinations, patient demographics, geographic regions, treatment paradigms, and healthcare systems.

This study lays a solid foundation for identifying patients who are likely to benefit from immunotherapy. The next step toward clinical implementation is validation in larger, more diverse populations with broader ethnic backgrounds and heterogeneous disease characteristics through multicenter retrospective studies. Subsequently, prospective realworld studies will be required to evaluate model performance and determine the most clinically effective predictive algorithms.

Conclusion

We have developed a well-validated machine learning–based predictive model capable of distinguishing responders from non-responders to immunotherapy in advanced HCC using clinical features in the form of structured tables and unstructured text. This model may support clinical decision-making in a setting where treatment response is notoriously difficult to predict. With the inclusion of additional clinical

features and external validation in future studies, this approach has the potential to evolve into a clinically useful tool for guiding immunotherapy selection in routine practice.

Declarations

Acknowledgements

The authors would like to acknowledge UPMC for the compilation of and provision of access to the patient datasets used in this research.

Ethics Approval and Consent to Participate

This study was reviewed and approved by the University of Pittsburgh Institutional Review Board (IRB) (protocol number: M0D23070075-002). This study used retrospectively collected, de-identified patient data, and was conducted in accordance with relevant guidelines and regulations. Informed consent to participate was granted by all participants in the study.

Data Availability Statement

The datasets generated and/or analyzed during the current study are not publicly available due to data use agreements and patient confidentiality but may be available from the UPMC subject to their approval and applicable ethical and legal requirements.

Funding

This study was funded by Curenetics Ltd.

Conflicts of Interest

The authors declare no potential conflicts of interest.

Statement of Significance

This study develops a multimodal AI model predicting HCC immunotherapy response (AUC 0.78) using routine clinical labs and text data, outperforming PD-L1 biomarkers. It identifies actionable predictors enabling precision selection for Atezo/Bev and Durva/Treme, addressing the critical 70% non-response rate.

References

1. Global Burden of Disease Liver Cancer Collaboration. 'The Burden of Primary Liver Cancer and Underlying Etiologies From 1990 to 2015 at the Global, Regional, and National Level: Results From the Global Burden of Disease Study 2015', *JAMA Oncol*, vol. 3, no. 12, pp. 1683–1691, Dec. 2017, 10.1001/jamaoncol.2017.3055
2. Rumgay H, et al. Global burden of primary liver cancer in 2020 and predictions to 2040. J Hepatol. Dec. 2022;77(6):1598–606. 10.1016/j.jhep.2022.08.021.
3. Sung H, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. Cancer J Clin. 2021;71(3):209–49. 10.3322/caac.21660.
4. McGlynn KA, Petrick JL, Groopman JD. Liver Cancer: Progress and Priorities. Cancer Epidemiol Biomarkers Prev. Oct. 2024;33(10):1261–72. 10.1158/1055-9965.EPI-24-0686.
5. Falette Puisieux M et al. Jan., 'Therapeutic Management of Advanced Hepatocellular Carcinoma: An Updated Review', *Cancers*, vol. 14, no. 10, p. 2357, 2022, 10.3390/cancers14102357
6. Shiani A, Narayanan S, Pena L, Friedman M. The Role of Diagnosis and Treatment of Underlying Liver Disease for the Prognosis of Primary Liver Cancer. Cancer Control. 2017;24(3). 10.1177/1073274817729240.
7. Erratum. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2020;70(4):313. 10.3322/caac.21609.
8. Garg P, Pareek S, Kulkarni P, Horne D, Salgia R, Singhal SS. Next-Generation Immunotherapy: Advancing Clinical Applications in Cancer Treatment. J Clin Med. Jan. 2024;13(21):6537. 10.3390/jcm13216537.
9. Finn RS, et al. Atezolizumab plus Bevacizumab in Unresectable Hepatocellular Carcinoma. N Engl J Med. May 2020;382(20):1894–905. 10.1056/NEJMoa1915745.
10. Abou-Alfa GK et al. Jun., Tremelimumab plus Durvalumab in Unresectable Hepatocellular Carcinoma, *NEJM Evidence*, vol. 1, no. 8, 2022, https://doi.org/10.1056/evidoa2100070
11. Qin R, Jin T, Xu F. Biomarkers predicting the efficacy of immune checkpoint inhibitors in hepatocellular carcinoma. Front Immunol. 2023;14:1326097. 10.3389/fimmu.2023.1326097. PMID: 38187399; PMCID: PMC10770866.
12. Ventura I, Sanchiz L, Legidos-García ME, Murillo-Llorente MT, Pérez-Bermejo M. 'Atezolizumab and Bevacizumab Combination Therapy in the Treatment of Advanced Hepatocellular Cancer', *Cancers*, vol. 16, no. 1, p. 197, Jan. 2024, 10.3390/cancers16010197
13. Doroshow DB, et al. PD-L1 as a biomarker of response to immunecheckpoint inhibitors. Nat Rev Clin Oncol. Jun. 2021;18(6):345–62. 10.1038/s41571-021-00473-5.
14. Yang Y, et al. The predictive value of PD-L1 expression in patients with advanced hepatocellular carcinoma treated with PD-1/PD-L1 inhibitors: A systematic review and meta-analysis. Cancer Med. 2023;12(8):9282–92. 10.1002/cam4.5676.
15. Adlung L, Cohen Y, Mor U, Elinav E. 'Machine learning in clinical decision making', *Med*, vol. 2, no. 6, pp. 642–665, Jun. 2021, 10.1016/j.medj.2021.04.006

16. Huang L, et al. Artificial Intelligence Can Predict Personalized Immunotherapy Outcomes in Cancer. Cancer Immunol Res. Feb. 2025;13(7):964–77. https://doi.org/10.1158/2326-6066.CIR-24-1270.
17. Sharma L, Balaji V, Katumba A, Mohan E, Adeleke S. Abstract A015: Precision medicine approach to melanoma immunotherapy: Predicting response, adverse events, and hospital admissions using machine learning and explainable artificial intelligence. Clin Cancer Res. Jul. 2025;31:A015–015. 10.1158/1557-3265.AIMACHINE-A015.
18. Sharma L, et al. 2P Using machine learning (ML) and explainable artificial intelligence (AI) to accurately predict immune-checkpoint inhibitor (ICI) response in small cell (SCLC) and non-small cell (NSCLC) lung cancer patients. ESMO Open. Mar. 2025;10:104157–104157. https://doi.org/10.1016/j.esmoop.2025.104157.
19. Sharma L, et al. 3P Predicting immunotherapy-related adverse events in melanoma patients using machine learning algorithms and explainable artificial intelligence. ESMO Open. Mar. 2025;10:104158. https://doi.org/10.1016/j.esmoop.2025.104158.
20. Sharma L et al. Immunotherapy toxicity prediction in patients with melanoma using machine learning algorithms. J Clin Oncol, 42, no. 16_suppl, pp. e21567–e21567, Jun. 2024, https://doi.org/10.1200/jco.2024.42.16_suppl.e21567
21. Xu J, et al. Predicting Immunotherapy Response in Unresectable Hepatocellular Carcinoma: A Comparative Study of Large Language Models and Human Experts. J Med Syst. May 2025;49(1):64. 10.1007/s10916-025-02192-1.
22. Eisenhauer EA, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. Jan. 2009;45(2):228–47. 10.1016/j.ejca.2008.10.026.
23. van Buuren S, Flexible Imputation of Missing Data., Second Edition. Second edition. | Boca Raton, Florida: CRC Press, [2019] |: Chapman and Hall/CRC, 2018. https://doi.org/10.1201/9780429492259
24. Kuhn M, Johnson K. Applied predictive modeling. New York: Springer; 2019.
25. Guyon I, Elisseeff A. 'An introduction to variable and feature selection', *J. Mach. Learn. Res.*, vol. 3, no. null, pp. 1157–1182, Mar. 2003.
26. Breiman L, Forests' 'Random. *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, 10.1023/A:1010933404324
27. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. 'Support vector machines', *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18–28, Jul. 1998, 10.1109/5254.708428
28. Chen T, Guestrin C. 'XGBoost: A Scalable Tree Boosting System', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794. 10.1145/2939672.2939785
29. Lundberg S, Lee S. 'A Unified Approach to Interpreting Model Predictions', Nov. 25, 2017, *arXiv*: arXiv:1705.07874. 10.48550/arXiv.1705.07874
30. Abou-Alfa GK et al. Jul., 'Tremelimumab plus Durvalumab in Unresectable Hepatocellular Carcinoma', *NEJM Evidence*, vol. 1, no. 8, p. EVIDoa2100070, 2022, 10.1056/EVIDoa2100070

31. Salton G, Buckley C. ‘Term-weighting approaches in automatic text retrieval’, *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, Jan. 1988, 10.1016/0306-4573(88)90021-0
32. Tian et al. The prognostic and predictive value of AFP in immune checkpoint inhibitor-treated hepatocellular carcinoma: a systematic review and metaanalysisFront. Immunol., 04 November 2025 Sec. Cancer Immunity and Immunotherapy Volume 16–2025 | https://doi.org/10.3389/fimmu.2025.1695861
33. Hou G, Liu B, Fan ZQ, Li C, Zhang JP, Guo YH, Zhang RY, Zheng Y, Zhu H, Wang NY. Association between early response of alpha-fetoprotein and treatment efficacy of systemic therapy for advanced hepatocellular carcinoma: A multicenter cohort study from China. Front Oncol. 2023;12:1094104. 10.3389/fonc.2022.1094104. . PMID: 36686731; PMCID: PMC9846773.
34. Tian BW, Yan LJ, Ding ZN, Liu H, Han CL, Meng GX, Xue JS, Dong ZR, Yan YC, Hong JG, Chen ZQ, Wang DX, Li T. Evaluating liver function and the impact of immune checkpoint inhibitors in the prognosis of hepatocellular carcinoma patients: A systemic review and meta-analysis. Int Immunopharmacol. 2023;114:109519. 10.1016/j.intimp.2022.109519. . Epub 2022 Nov 30. PMID: 36459922.
35. Zhu AX, Finn RS, Edeline J, Cattan S, Ogasawara S, Palmer D, et al. Pembrolizumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib (KEYNOTE-224): a non-randomised, open-label phase 2 trial. Lancet Oncol. 2018;19(7):940–52. 10.1016/S1470-2045(18)30351-6.
36. Yau T, Kang YK, Kim TY, El-Khoueiry AB, Santoro A, Sangro B, Melero I, Kudo M, Hou MM, Matilla A, Tovoli F, Knox JJ, Ruth He A, El-Rayes BF, Acosta-Rivera M, Lim HY, Neely J, Shen Y, Wisniewski T, Anderson J, Hsu C. Efficacy and Safety of Nivolumab Plus Ipilimumab in Patients With Advanced Hepatocellular Carcinoma Previously Treated With Sorafenib: The CheckMate 040 Randomized Clinical Trial. JAMA Oncol. 2020;6(11):e204564. 10.1001/jamaoncol.2020.4564. Epub 2020 Nov 12. Erratum in: JAMA Oncol. 2021 Jan 1;7(1):140. doi: 10.1001/jamaoncol.2020.6961. PMID: 33001135; PMCID: PMC7530824.
37. Hectors SJ, Lewis S, Besa C, King MJ, Said D, Putra J, Ward S, Higashi T, Thung S, Yao S, Laface I, Schwartz M, Gnjatic S, Merad M, Hoshida Y, Taouli B. MRI radiomics features predict immuno-oncological characteristics of hepatocellular carcinoma. Eur Radiol. 2020;30(7):3759–69. 10.1007/s00330-020-06675-2. . Epub 2020 Feb 21. PMID: 32086577; PMCID: PMC7869026.
38. Rakaee M, Tafavvoghi M, Ricciuti B, Alessi JV, Cortellini A, Citarella F, Nibid L, Perrone G, Adib E, Fulgenzi CAM, Hidalgo Filho CM, Di Federico A, Jabar F, Hashemi S, Houda I, Richardsen E, Rasmussen Busund LT, Donnem T, Bahce I, Pinato DJ, Helland Å, Sholl LM, Awad MM, Kwiatkowski DJ. Deep Learning Model for Predicting Immunotherapy Response in Advanced Non-Small Cell Lung Cancer. JAMA Oncol. 2025;11(2):109–18. 10.1001/jamaoncol.2024.5356. . PMID: 39724105; PMCID: PMC11843371.
39. Qin R, Jin T, Xu F. Biomarkers predicting the efficacy of immune checkpoint inhibitors in hepatocellular carcinoma. Front Immunol. 2023;14:1326097. 10.3389/fimmu.2023.1326097. PMID: 38187399; PMCID: PMC10770866.

40. Liu J, Fu R, Su Y, Li Z, Huang X, Wang Q, Shi Z, Wei S. Applications of artificial intelligence in cancer immunotherapy: a frontier review on enhancing treatment efficacy and safety. Front Immunol. 2025;16:1676112. 10.3389/fimmu.2025.1676112. . PMID: 41246300; PMCID: PMC12615365.
41. Gao Q, Yang L, Lu M, Jin R, Ye H, Ma T. The artificial intelligence and machine learning in lung cancer immunotherapy. J Hematol Oncol. 2023;16(1):55. 10.1186/s13045-023-01456-y. . PMID: 37226190; PMCID: PMC10207827.

Figures

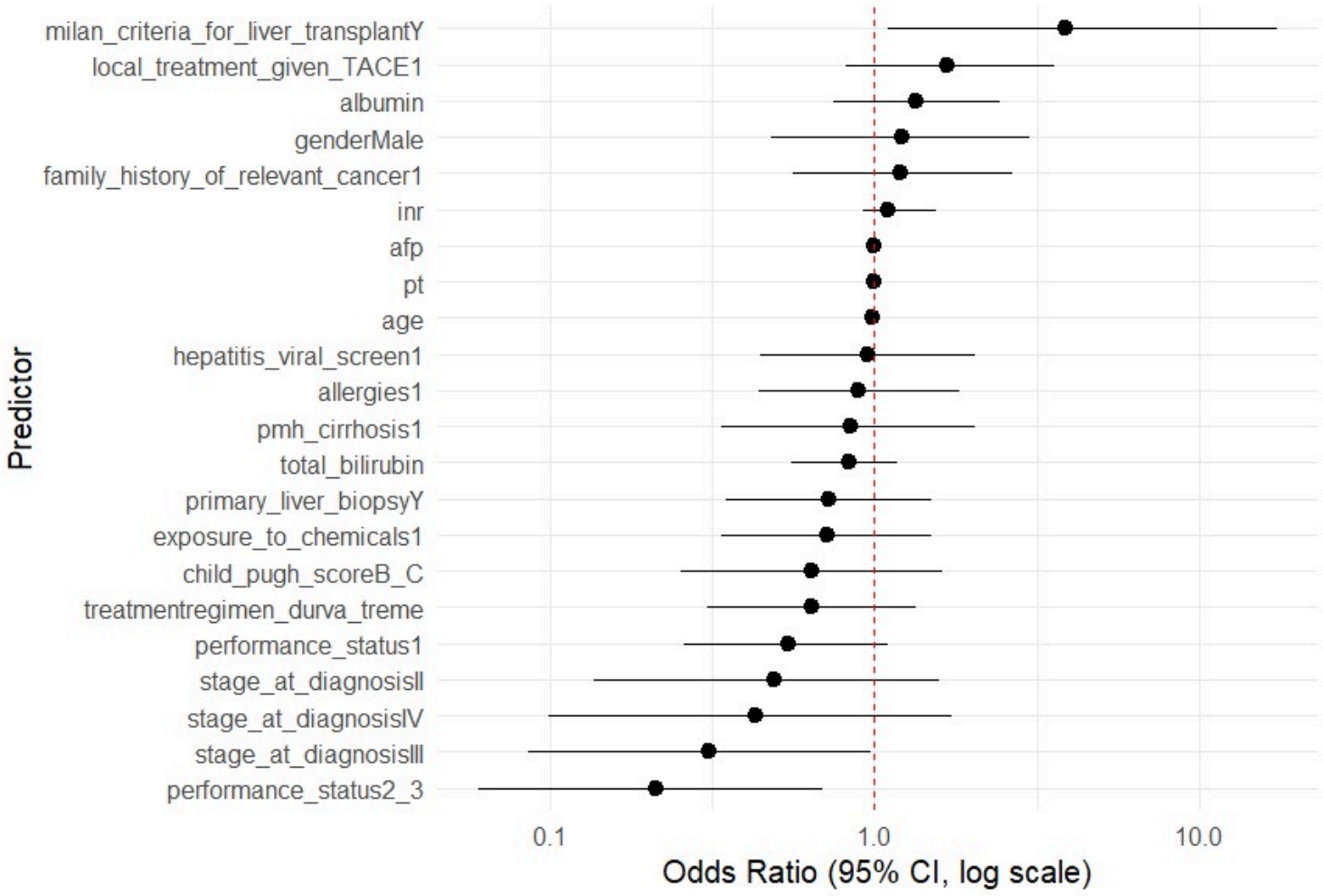


Figure 1

Odds Ratios of Logistic Regression Model

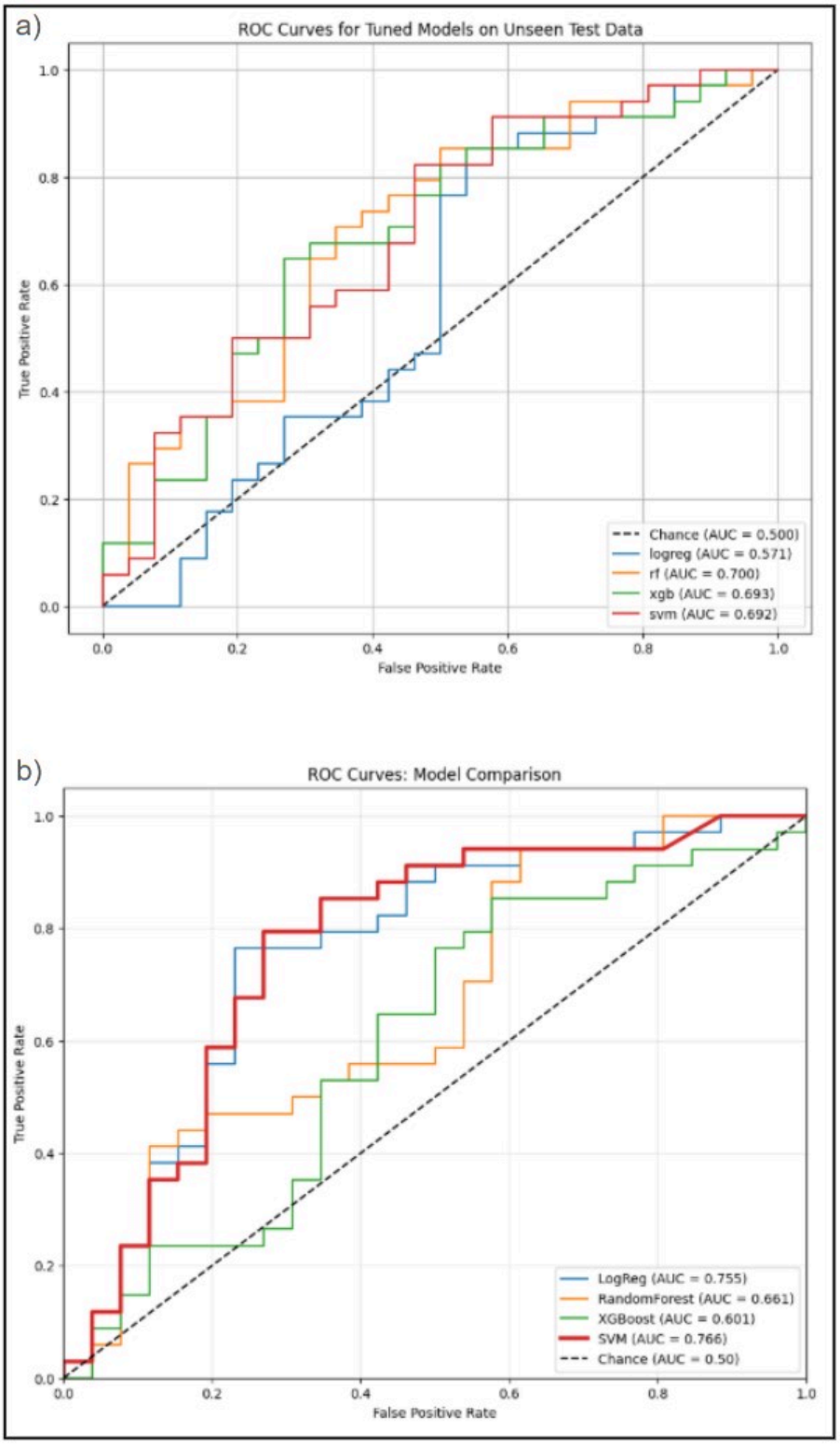


Figure 2

Receiver Operating Characteristics for Machine Learning Models

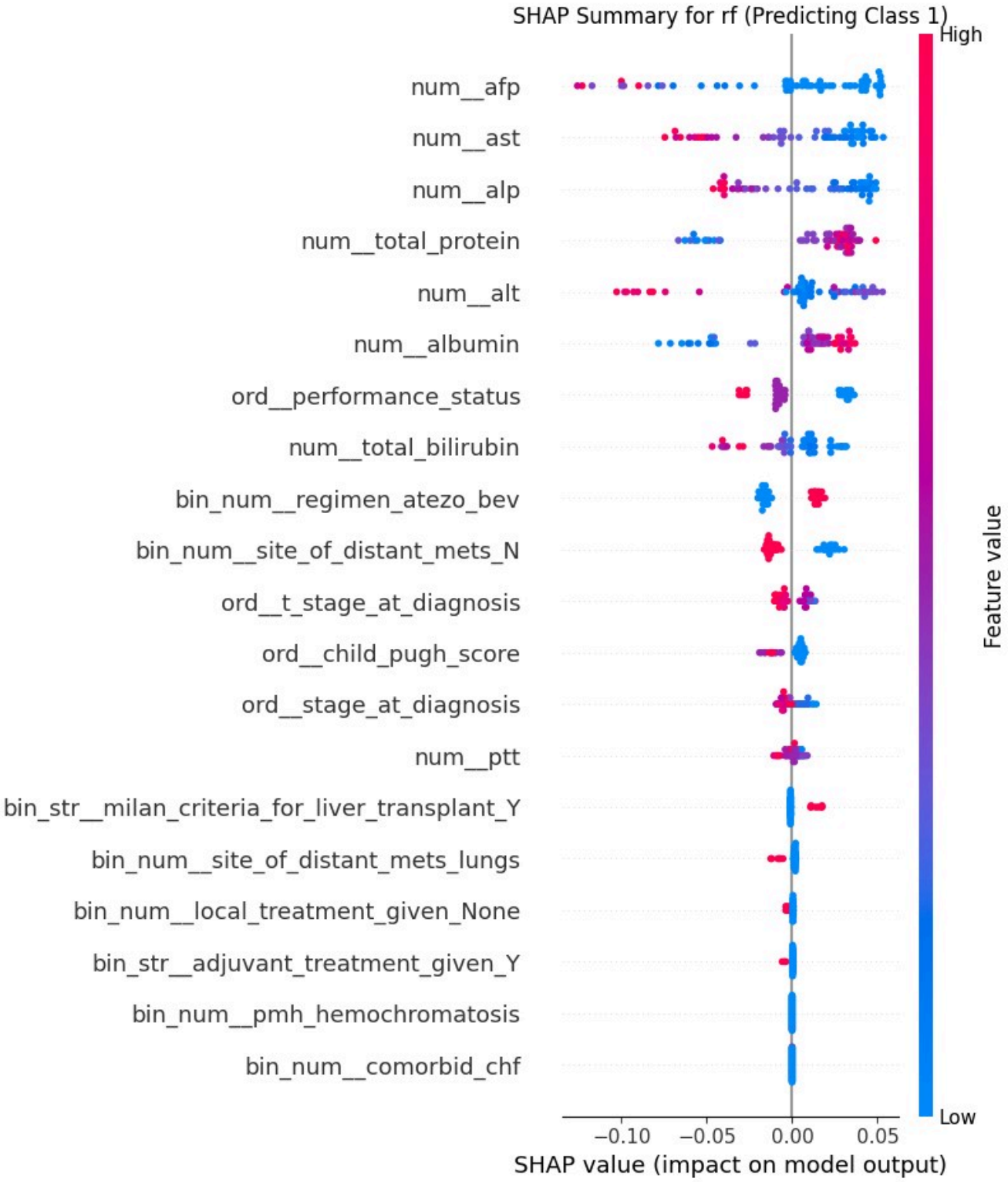


Figure 3

SHAP Beeswarm Summary Plot of Top Tabular Features

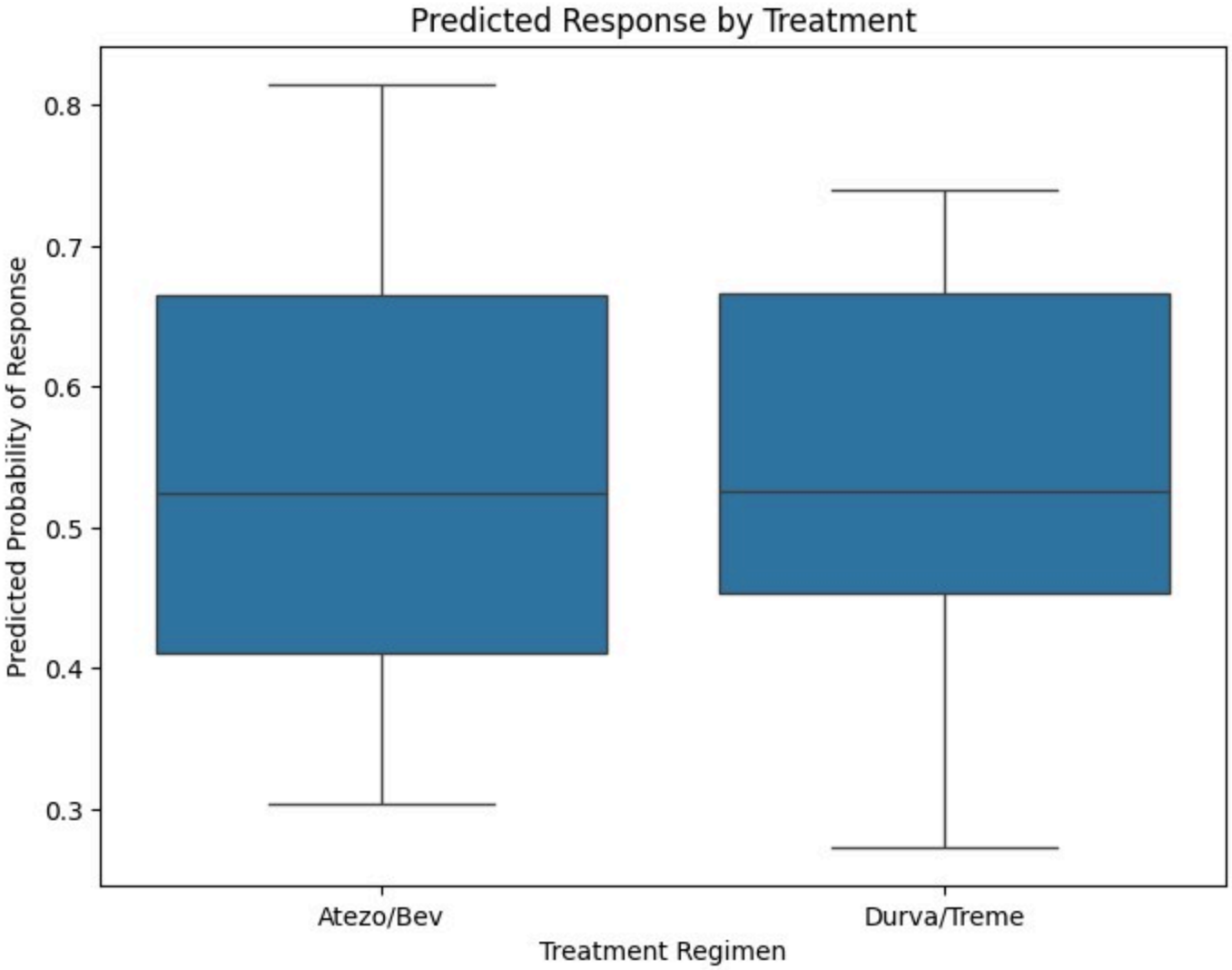


Figure 4

Boxplots of Predicted Responses by Treatment Subgroup

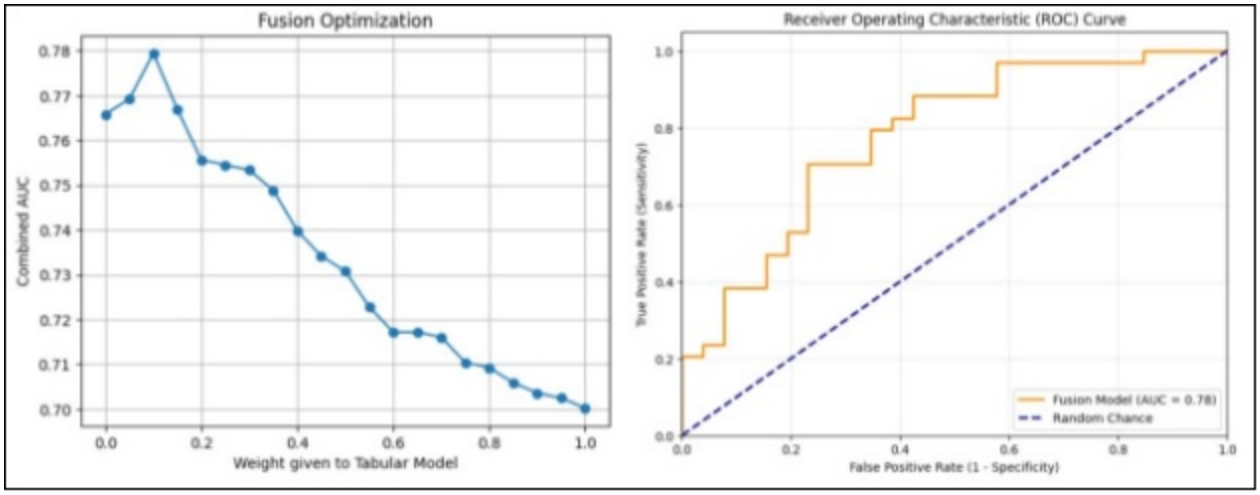


Figure 5

Results of Late-Fusion a) Optimisation and b) Performance