HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 3 (2024)

Personalized Survival Prediction for Hepatocellular Carcinoma Using Artificial Intelligence

Steven Yu Lin[1], Justin Pilon[#], Chen-Wen Yen[#] and Chih-Wen Lin[#]

[1]Kaohsiung American School, Kaohsiung, Taiwan
[#]Advisor

ABSTRACT

Hepatocellular carcinoma (HCC) presents a significant global health challenge. Researchers have explored machine artificial intelligence-based systems for treatment recommendation. However, these models only consider a limited range of treatment options and primarily replicate historical treatment decisions, potentially leading to suboptimal outcomes. The XGBoost-based Accelerated Failure Time (XGBAFT) model, a novel machine learning-based system, predicts survival times for nine HCC treatment options. Developed using data from 8,511 patients at two medical centers in Taiwan, the model integrates the XGBoost algorithm with the accelerated failure time framework. Its key strength lies in comparing predicted survival times and curves across treatment options, facilitating personalized decision-making. The model can also predict survival beyond a specified target survival time, with accuracy and reliability established using binary classification measures. The model achieved a concordance index of 0.831, outperforming the Cox proportional hazards model. Notably, treatment was identified as the most influential covariate, underscoring the importance of selecting appropriate and effective treatments tailored to patient characteristics for improving survival outcomes. In conclusion, the XGBAFT model demonstrates significant potential in predicting HCC outcomes and personalized survival predictions. Future research should focus on integrating advanced machine learning algorithms and improving model interpretability to support clinical decision-making and personalized care.

Introduction

Hepatocellular carcinoma (HCC) presents a significant global health challenge. As reported by the World Health Organization, liver cancer, predominantly HCC, accounted for approximately 830,000 deaths worldwide in 2020 (World Health Organization, 2022). The selection of treatments for HCC is generally guided by the Barcelona Clinic Liver Cancer (BCLC) staging system (Omata et al., 2017; Galle et al., 2018; Heimbach et al., 2018). However, there are discrepancies between the treatments used in clinical settings and those recommended by the BCLC system (Leoni et al., 2014; Mittal et al., 2016), influenced by a variety of factors including cancer stage, liver function, tumor location and distribution, performance status, and comorbidities. These factors collectively impact the clinical outcomes for HCC patients, and the complexity of treatment decisions has hindered large-scale clinical research. Traditional statistical methods struggle to account effectively for these numerous influencing factors, highlighting the need for a new system to assess treatment options for HCC.

In response to these challenges, researchers have explored artificial intelligence-based systems for treatment recommendation (Choi et al., 2020; Lee et al., 2024). However, these models only consider a limited range of treatment options and primarily replicate historical treatment decisions, potentially leading to suboptimal outcomes. Moreover, these systems typically recommend only two treatment alternatives, which is insufficient for addressing the diverse needs of HCC patients.

To address these limitations, we introduce a novel decision tree-based artificial intelligence system designed to predict the survival times associated with nine distinct treatment options for HCC. This system

allows clinicians and patients to compare predicted survival times and corresponding survival curves across various treatment choices, providing a personalized approach based on individual patient characteristics. By considering a wider range of treatment options and offering survival time predictions, our system aims to support more informed and patient-specific decision-making in HCC treatment planning.

We validated our system using a dataset of 8,511 patients from two medical centers. The results demonstrate the potential of our approach to enhance clinical decision-making and improve HCC prognosis by enabling clinicians and patients to make more informed treatment choices based on personalized survival predictions.

Methods

Patients

This study retrospectively analyzed data from patients diagnosed with HCC at two Taiwanese medical centers: E-DA Hospital (E-DAH, 2007-2020) and Kaohsiung Medical University Chung-Ho Memorial Hospital (KMUH, 1993-2019), adhering to the diagnostic guidelines of the American Association for the Study of the Liver Diseases (AASLD). This study has received ethical approval from the Institutional Review Boards of E-DAH and KMUH.

The primary outcome of interest was overall survival (OS), defined as the time from diagnosis until death or the last follow-up date. Survival and mortality status were ascertained from patient medical records. The study included two sets of covariates: pre-treatment features and treatment options.

Eighteen pre-treatment features relevant to HCC were selected. These features were consistently available across both medical centers with minimal missing data. Initially, data for 6,260 patients from E-DAH and 4,032 patients from KMUH were collected. However, 1,666 patients from E-DAH and 115 from KMUH were excluded due to missing values in one or more pre-treatment features. The remaining 8,511 patients had complete data for all pre-treatment covariates, which are summarized in Table 1.

The second set of covariates comprised nine treatment options, reflecting a broad spectrum of therapeutic approaches: liver transplantation (LT), surgical resection (SR), radiofrequency ablation (RFA) including percutaneous ethanol injection therapy (PEIT), transarterial chemoembolization (TACE), targeted therapies such as sorafenib and lenvatinib, immunotherapies including nivolumab with or without ipilimumab and pembrolizumab with or without lenvatinib, atezolizumab combined with bevacizumab, hepatic artery infusion chemotherapy (HAIC), radiotherapy (RT), and best supportive care (BSC).

Model Development

This study aimed to develop a personalized prediction system for HCC treatment outcomes, designed to aid clinicians and patients in selecting the most appropriate treatment based on pre-treatment characteristics.

To identify the most effective model for predicting survival within our dataset, we evaluated a range of established and advanced methods with available open-source codes. These included the Cox proportional hazards model (Cox, 1972), machine learning approaches like Random Survival Forests (RSF) (Ishwaran et al., 2008), XGBoost (Barnwal, 2022), and advanced deep learning models such as Deep Cox Mixture and Deep Survival Machines (Nagpal et al., 2022).

We ultimately chose a nonlinear accelerated failure time (AFT) model, proposed by Barnwal et al. (2022), which incorporates the XGBoost gradient boosting algorithm developed by Chen and Guestrin (2016). This model, hereinafter referred to as the XGBAFT model, integrates the AFT framework into its loss function and substitutes the linear component with an ensemble of decision trees. The XGBAFT model is particularly

effective at capturing complex, nonlinear relationships between covariates and survival times, and it efficiently handles various types of censored survival data.

Our choice of the XGBAFT model was motivated by three factors: XGBoost's robust ability to quickly process complex data patterns, the model's capacity for direct survival time prediction which facilitates easy comparison of prognostic outcomes across different treatments, and its provision of several metrics for assessing feature importance, thus enhancing interpretability.

To train and validate the XGBAFT model, we employed a two-step approach. First, Bayesian optimization (Snoek et al., 2012) identified the optimal hyperparameters to maximize model performance on our dataset. The model was then trained using a five-fold cross-validation method to ensure the robustness and generalizability of the survival time predictions. This approach allowed each patient's survival time to be predicted precisely once during the validation phase, providing a survival time estimate for every individual.

Following the cross-validation process, the predicted survival times for each patient in the five validation subsets were rounded to the nearest integer month. This integer value represents the predicted survival duration in months. For each patient, we then identified a subgroup of patients from the validation subsets whose predicted survival times were rounded to the same integer value. This subgroup essentially represents a cohort of patients with similar predicted prognoses.

We then leveraged these subgroups to generate personalized Kaplan-Meier (KM) survival curves. A KM curve depicts the estimated probability of survival over time for a specific patient group. In our case, the KM curve for each patient reflects the survival probability over time for patients within the validation subsets whose predicted survival closely aligns with the patient's own prediction. This approach provides a more patient-centric view of the predicted survival trajectory.

Results

The XGBAFT model offers a robust method for predicting and visualizing survival outcomes in HCC patients, integrating detailed individual predictions with cohort-based survival curves. To demonstrate the utility of our approach, we used a stage 0 patient as an example. Initially treated with liver transplantation, this patient had a censored survival time of 68 months. By contrast, the XGBAFT model predicted 118 months. After altering the treatment method to surgical resection, TACE, and best supportive care, the predicted survival times changed to 54, 43, and 33 months, respectively, with corresponding survival curves illustrated in Figure 1.

To illustrate the performance of the proposed XGBAFT model, Figure 2 displays the survival curve generated by the XGBAFT model for the entire population of 8,511 patients alongside the KM curve derived from the actual survival data of these patients. This comparison allows for a direct assessment of the model's ability to accurately predict survival probabilities over time. The results indicate that the model's effectiveness declines as the prediction period lengthens. This trend is expected, given that our pre-treatment covariates primarily reflect the patient's initial condition. Over time, the relevance of these initial conditions to the patient's ongoing health status may lessen, and post-treatment variables, which our model does not account for, can significantly impact predictive accuracy. Additionally, the number of long-term survivors decreases over time, resulting in less data and consequently less robust statistical confidence in long-term predictions. This scarcity of data complicates the training of models for accurate long-term forecasts.

To gauge how well our model compares survival outcomes for different treatment options, we employed the concordance index (C-index) (Harrell, 1982). The C-index essentially measures how accurate our model is in predicting which treatment will lead to better survival. The C-index considers all possible pairs of patients in the data. For each pair, it checks if the patient our model predicts to have a shorter lifespan actually dies before the other patient. A higher C-index means our model is more accurate at predicting which patient will live longer.

To assess the XGBAFT model's performance, we compared it to the well-established Cox proportional hazards model, a common benchmark in survival analysis. This comparison allows us to directly evaluate how the XGBAFT model performs against a validated method. In our study, the XGBAFT model achieved a C-index of 0.831, indicating slightly better predictive accuracy compared to the Cox model (C-index of 0.821). This suggests that the XGBAFT model might offer more precise estimations of survival probabilities for patients with HCC.

A key advantage of the XGBAFT model over the Cox model is its ability to predict whether a patient will survive beyond a specified target survival time (TST). We analyzed this capability by treating it as a series of binary classification tasks, where patients with survival times shorter than the TST were classified as the positive class, and those with survival times equal to or longer than the TST as the negative class. This approach was evaluated across TSTs ranging from 1 to 120 months.

To assess the overall effectiveness of this method in predicting whether a patient can survive beyond a specified survival time threshold (TST), we utilized the Matthews Correlation Coefficient (MCC) as described by Chicco & Jurman (2020). This performance metric, displayed in Figure 3, is plotted as a function of TST. The MCC provides a balanced evaluation by considering all outcomes: true positives, true negatives, false positives, and false negatives. This measure is particularly valuable in scenarios with class imbalances, ranging from -1 (indicating total disagreement between prediction and actual outcomes) to +1 (indicating perfect prediction), with 0 suggesting that the prediction capability is no better than random chance. Figure 3 also illustrates the class imbalance ratio (CIR), defined as the ratio of the number of positive class data points to negative class data points. Extreme values of CIR complicate binary classification, as models may show bias towards the majority class, leading to diminished performance, particularly in identifying less frequent but critical events.

The accuracy of predictions, in terms of correctly identifying non-survivors and survivors, was further examined using sensitivity and specificity, which are plotted in Figure 4 as a function of TST. High CIR values generally resulted in a tendency to predict the positive class more frequently, enhancing sensitivity but reducing specificity. Conversely, a low CIR often led to higher specificity but reduced sensitivity. These trends are consistent with the results depicted in Figure 4.

Figure 5 demonstrates the reliability of our predictions across different TSTs by showing the positive predictive value (PPV) and negative predictive value (NPV). With a high CIR, the PPV is generally higher due to a greater likelihood of correctly predicting the positive class, albeit at the expense of more false positives, thereby reducing the NPV. In contrast, a lower CIR tends to decrease the PPV while increasing the NPV, reflecting fewer false positive predictions. These observations align with the findings presented in Figure 5.
Finally, Figure 6 illustrates the impact of various covariates on survival time, employing the GAIN feature importance measure from the XGBoost algorithm. This metric quantifies the average change in the loss function resulting from splits on specific features within the XGBoost ensemble's decision trees. The treatment method was identified as the most influential covariate, followed by maximum tumor size, number of tumors, and performance status. The prominence of treatment as a predictor underscores its critical role in influencing survival outcomes. This suggests that selecting the most appropriate and effective treatment, tailored to individual patient characteristics and disease specifics, is essential for improving survival outcomes.

Discussion

In this study, we introduce a novel survival analysis system, the XGBAFT model, tailored for patients with HCC. The model’s primary output predicts survival times, offering critical insights that could significantly enhance clinical decision-making and patient outcomes.

Despite its efficacy, the XGBAFT model has several limitations that require consideration. Initially, the development and validation of the model were confined to a cohort of Taiwanese patients. This limitation

might restrict its generalizability due to variations in genetic, environmental, and lifestyle factors that differ across ethnicities and regions. These factors influence disease progression and treatment responses. This underscores the need for further validation of the model in more diverse populations to confirm its effectiveness globally.

Additionally, the current iteration of the XGBAFT model does not provide mechanisms for suggesting or evaluating the suitability of various treatment options tailored to individual patient needs. The appropriateness of treatments-ranging from surgical resection and transplantation to ablation therapies and systemic treatments-depends on numerous patient-specific factors, including liver function, tumor stage, cirrhosis presence, overall health, and comorbidities. Thus, the crucial role of clinicians remains, as they must assess the viability of each treatment option based on the comprehensive clinical profile of each patient.

Despite these challenges, the XGBAFT model employs a robust machine learning framework to generate personalized survival estimates, invaluable for prognostic assessments. Future enhancements should aim to expand the training datasets to include a broader spectrum of patient demographics and to incorporate features that assist in treatment recommendations. Exploring the integration of more advanced machine learning algorithms for survival analysis and improving the interpretability of model predictions are also promising directions. Enhanced interpretability could involve developing methods to quantify the influence of each covariate on predicted survival outcomes. This advancement would not only deepen our understanding of the factors driving the model's predictions but also aid in delineating the specific roles that each covariate plays in determining the survival outcomes of patients with HCC.

Conclusion

While the XGBAFT model shows significant potential in predicting survival outcomes for HCC patients and personalized survival predictions, its current limitations highlight the necessity for further refinement and broader validation. Addressing these limitations is essential for evolving this model into a more comprehensive and universally applicable survival prediction tool. Such advancements could support clinical decision-making and improve patient outcomes across diverse global populations.

Acknowledgment

I would like to appreciate my advisor Justin Pilon and my mentor Professor Chen-Wen Yen (National Sun Yat-Sen University, Kaohsiung, Taiwan) and Chih-Wen Lin (E-Da Hospital, I-Shou University) for all of the support and guidance you have given me along the way.

References

1. Barnwal, A., Cho, H., & Hocking, T. (2022). Survival regression with accelerated failure time model in XGBoost. Journal of Computational and Graphical Statistics, 31(4), 1292-1302.
2. Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
3. Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC genomics, 21, 1-13.
4. Choi, G. H., Yun, J., Choi, J., Lee, D., Shim, J. H., Lee, H. C., ... & Kim, K. M. (2020). Development of machine learning-based clinical decision support system for hepatocellular carcinoma. Scientific reports, 10(1), 14855.

5. Cox, D. R. (1972). Regression analysis of length of life. Journal of the Royal Statistical Society: Series B (Methodological), 34(2), 187-220.
6. Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests.
7. Lee, K. H., Choi, G. H., Yun, J., Choi, J., Goh, M. J., Sinn, D. H., ... & Kim, K. M. (2024). Machine learning-based clinical decision support system for treatment recommendation and overall survival prediction of hepatocellular carcinoma: a multi-center study. NPJ Digital Medicine, 7(1), 2.
8. Galle, P. R., Forner, A., Llovet, J. M., Mazzaferro, V., Piscaglia, F., Raoul, J. L., ... & Vilgrain, V. (2018). EASL clinical practice guidelines: management of hepatocellular carcinoma. Journal of hepatology, 69(1), 182-236.
9. Harrell Jr, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. JAMA, 247(18), 2543-2546.
10. Heimbach, J. K., Kulik, L. M., Finn, R. S., Sirlin, C. B., Abecassis, M. M., Roberts, L. R., ... & Marrero, J. A. (2018). AASLD guidelines for the treatment of hepatocellular carcinoma. Hepatology, 67(1), 358-380.
11. Leoni, S., Piscaglia, F., Serio, I., Terzi, E., Pettinari, I., Croci, L., ... & Bolondi, L. (2014). Adherence to AASLD guidelines for the treatment of hepatocellular carcinoma in clinical practice: experience of the Bologna Liver Oncology Group. Digestive and Liver Disease, 46(6), 549-555.
12. Mittal, S., Kanwal, F., Ying, J., Chung, R., Sada, Y. H., Temple, S., ... & El-Serag, H. B. (2016). Effectiveness of surveillance for hepatocellular carcinoma in clinical practice: a United States cohort. Journal of hepatology, 65(6), 1148-1154.
13. Nagpal, C., Potosnak, W., & Dubrawski, A. (2022, December). Auton-survival: An open-source package for regression, counterfactual estimation, evaluation and phenotyping with censored time-to-event data. In Machine Learning for Healthcare Conference (pp. 585-608). PMLR.
14. Omata, M., Cheng, A. L., Kokudo, N., Kudo, M., Lee, J. M., Jia, J., ... & Sarin, S. K. (2017). Asia–Pacific clinical practice guidelines on the management of hepatocellular carcinoma: a 2017 update. Hepatology international, 11, 317-370.
15. Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. Advances in neural information processing systems, 25.
16. World Health Organization. (2021). Cancer. https://www.who.int/news-room/fact-sheets/detail/cancer.

Table 1. Demographic features of 8511 patients with hepatocellular carcinoma

Characteristics		All patients (N=8511)
Age, year		63.0 ± 12.0
Gender	Male	6235 (73.2)
	Female	2276 (26.8)
Body mass index, kg/m^2		24.7 ± 4.1
Performance status	0	5870 (68.9)
	1	1714 (20.1)
	2	552 (6.4)
	3	264 (3.1)
	4	111 (1.5)
Etiology of liver disease	HBV	3827 (44.9)
	HCV	3266 (38.3)
Child-Pugh class	A	4870 (57.2)

	B	2945 (34.6)
	C	696 (8.2)
Maximal tumor size		7.1 ± 4.7
Tumor number	Single	3824 (44.9)
	Multiple	4687 (55.1)
Cirrhosis	Absence	3033 (35.6)
	Presence	5479 (64.4)
Lymph nodules	Absence	7813 (91.8)
	Presence	698 (8.2)
Extrahepatic spread	Absence	7309 (85.8)
	Presence	1202 (14.1)
Metastasis	Absence	7677 (90.2)
	Presence	834 (9.8)
Macrovascular invasion	Absence	6892 (80.1)
	Presence	1619 (19.9)
Laboratory findings	AFP, ng/mL	31792.7 ± 111098.6
	Total bilirubin, mg/dL	2.3 ± 3.87
	INR	1.4 ± 0.76
	Creatinine, mg/dL	1.0 ± 0.96

Data shown as mean ± standard deviation or number (%).

HBV: Hepatitis B virus; HCV: Hepatitis C virus; AFP: Alpha-fetoprotein; INR: International normalization ratio;

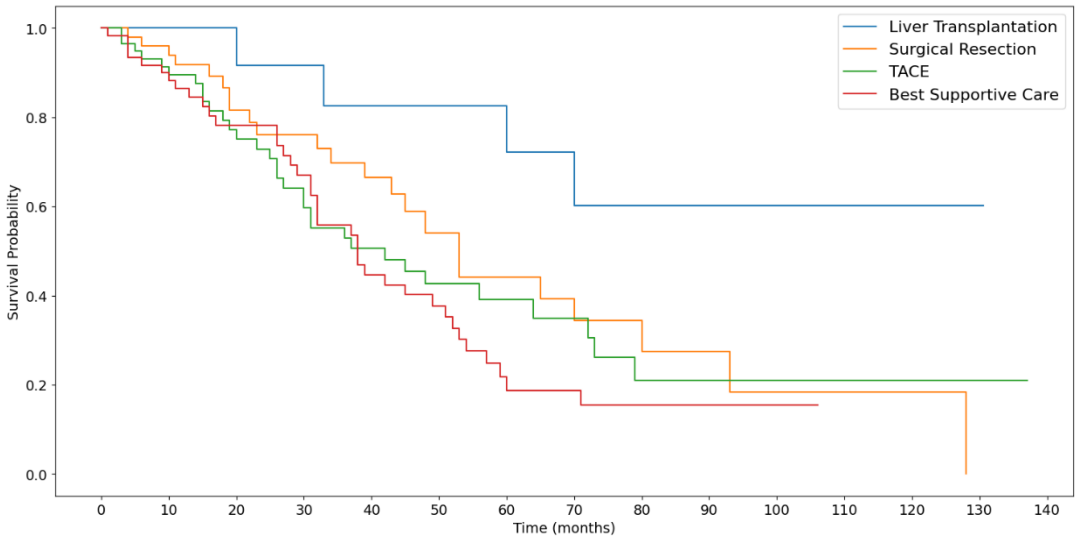

Figure 1. Survival curves comparing outcomes for four different treatment options for hepatocellular carcinoma Barcelona Clinic Liver Cancer stage 0 patient.

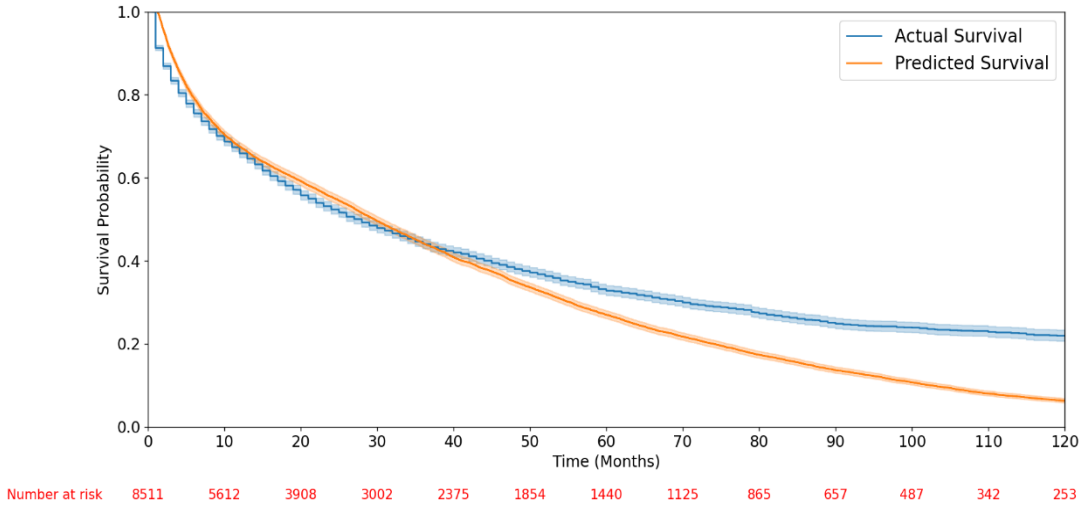

Figure 2. Comparison of actual vs. predicted survival curves

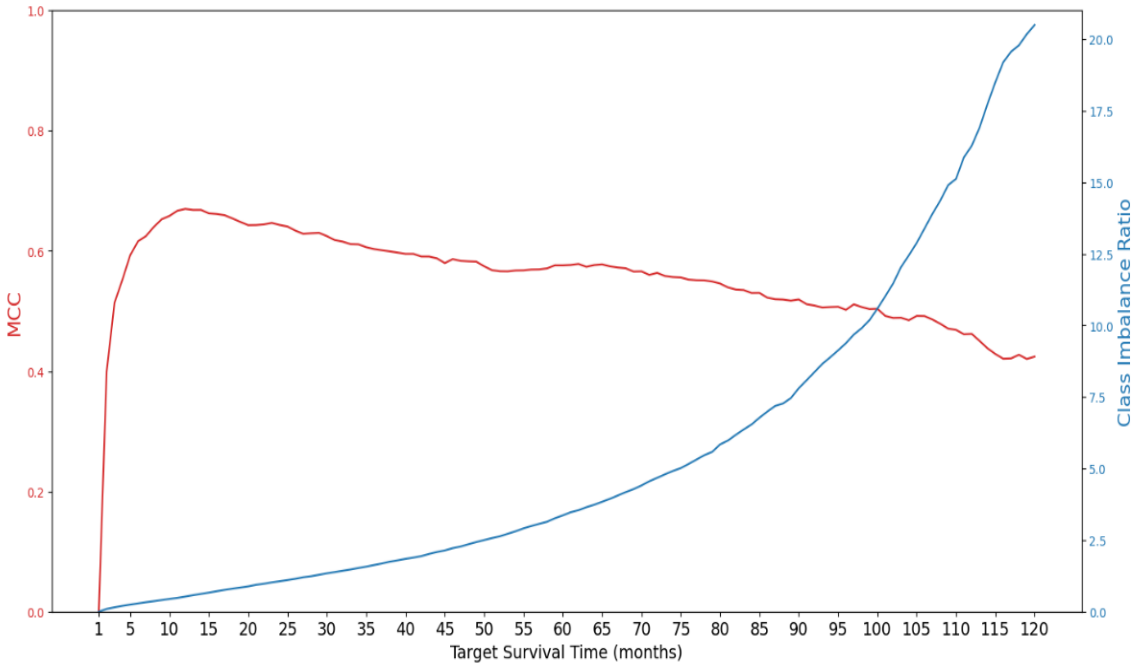

Figure 3. Trends in Matthews Correlation Coefficient (MCC) and Class Imbalance Ratio by Target Survival Time for hepatocellular carcinoma predictions

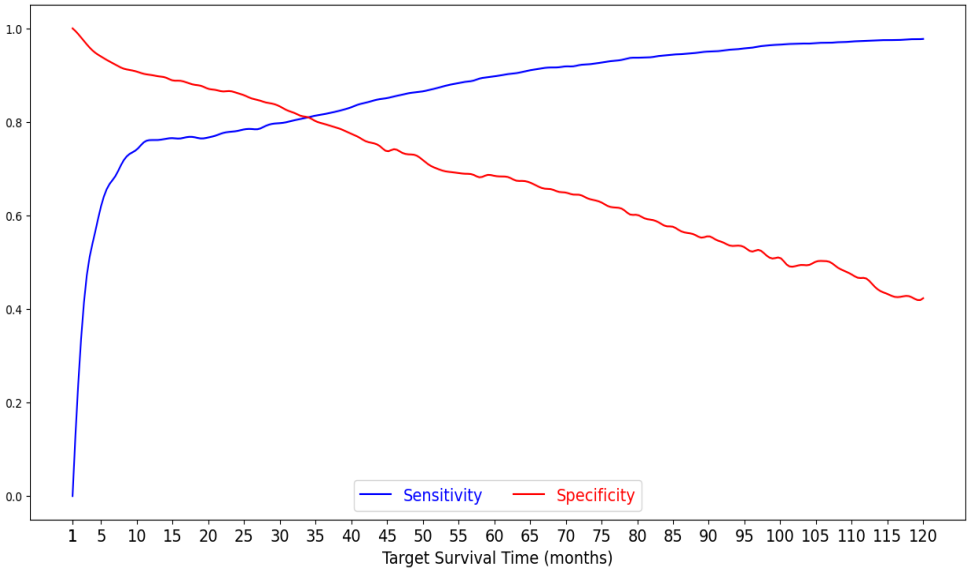

Figure 4. Sensitivity and specificity across Target Survival Times for hepatocellular carcinoma predictions

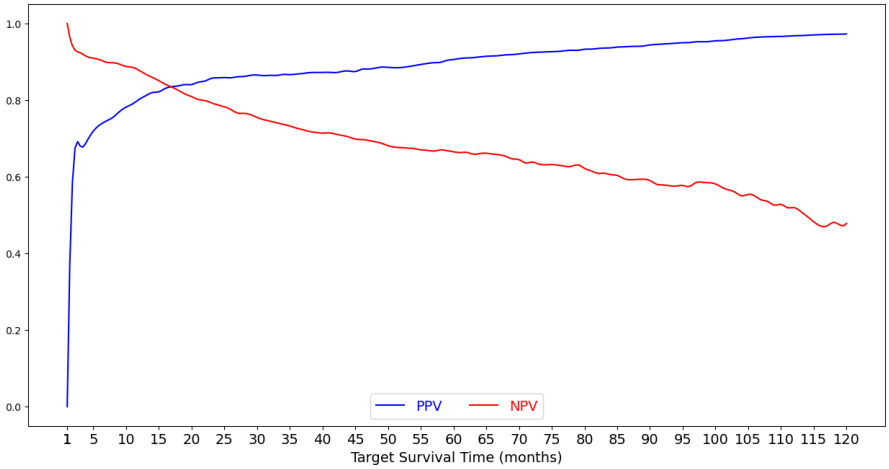

Figure 5. Trends in positive and negative predictive values over Target Survival Times for hepatocellular carcinoma predictions

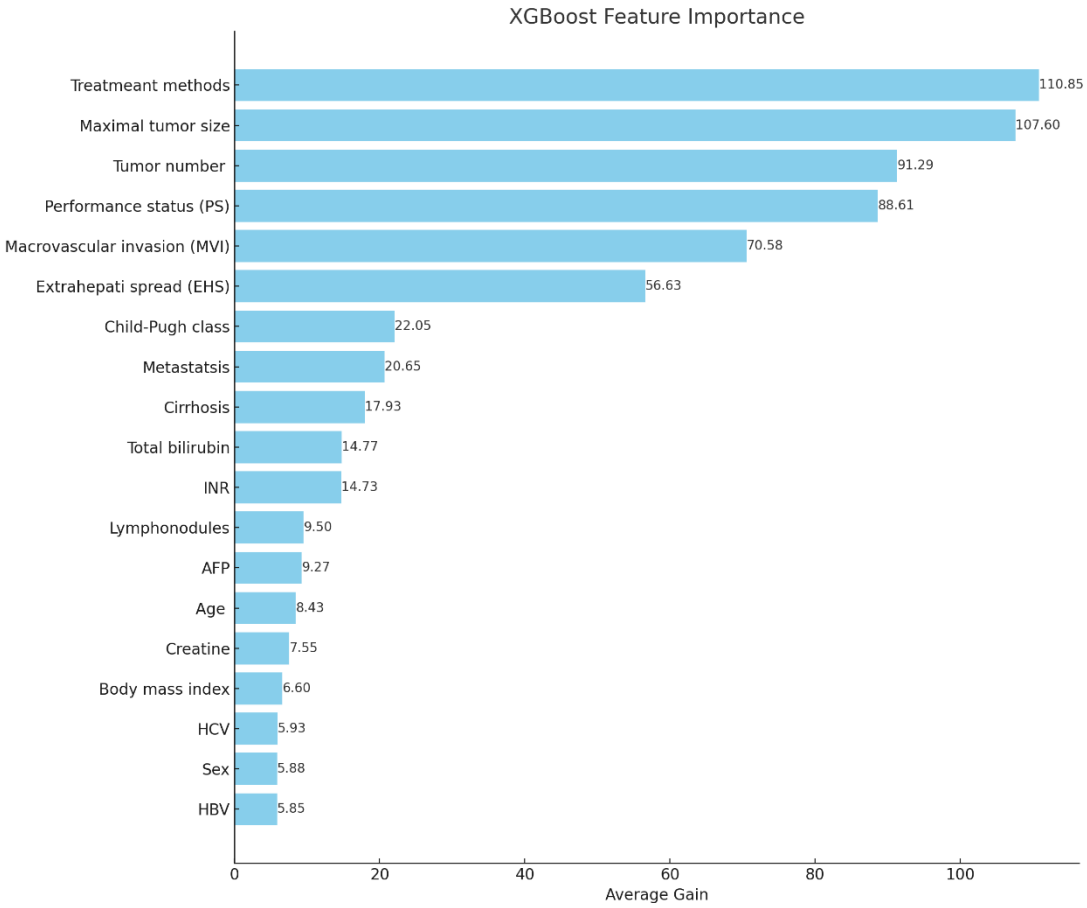

Figure 6. Feature importance in predicting hepatocellular carcinoma outcomes using XGBoost