


Research Article

Machine Learning Models and HCC Risk Prediction in Patients with HCV cACLD after SVR to DAA

Andrea Compagner^{1,4}, Valeria Piazzolla², Antonio Napolitano², Francesco Giuliani³, Federico Cabitza^{1,5}, Alessandra Mangia^{2*}

Abstract

Background and Aims: For HCC detection after SVR to DAA treatment in patients with chronic HCV infection and cACLD, a single baseline marker is not sufficient to define a risk profile. A combination of tools and evaluation of dynamic changes are needed to optimize accuracy prediction. We aimed to investigate whether developed ML models based on LSM, laboratory parameters and co-morbidities may help to predict HCC occurrence.

Methods: A retrospective prospective cohort of patients who achieved SVR from 2014 to 2024 at a single tertiary referral hospital. Several classes of ML models, namely: LR, DT, XGB, RF, SVM, and NB were tested and compared.

Results: The cohort included a total of 1082 patients, each of whom underwent at least one and up to three visits (T0, T1 and T2). Patients who had the first LSM (T1) later than 1 year and patients who had HCC diagnosis before T1, were excluded. The XGB model demonstrated that LSM results at T1 correlate with an increased risk of developing HCC. Higher baseline LSM values (T0), increased glucose (T1) and AST (T1) levels were also moderately associated with an increased risk, as was being of older age. Glucose was found to be a good predictor of increased risk, independently of baseline levels. Also, for LR, LSM results (both at T1 and T0) were one of the most predictive features for increased risk. Main finding is that XGB and LR substantially outperformed traditional HCC risk scores in predictive performance.

Conclusion: LR and XGB were significantly more accurate than traditional HCC risk scores. This work marks an important step toward precision hepatology, demonstrating how dynamic, data-driven models can reshape surveillance.

Lay summary: Not all the HCV infected patients with chronic compensated advanced liver disease carry the same risk of HCC development after SVR to DAA treatment. Several models are under evaluation to stratify HCC risk for these patients. We propose ML-based risk stratification models that can support more personalised and cost-effective surveillance regimens, reducing unnecessary monitoring while ensuring timely detection in high-risk patients.

Affiliation:

¹Department of Informatics, Systems and Communication, Università Milano-Bicocca, Milano, Italy

²Hepatology Unit, Fondazione Casa Sollievo della Sofferenza, IRCCS, San Giovanni Rotondo, Italy

³Innovation and Research Unit Casa Sollievo della Sofferenza, IRCCS, San Giovanni Rotondo, Italy

⁴IRCCS Ospedale Galeazzi Sant' Ambrogio, Milano, Italy

⁵Digital Health & Wellbeing Center, Fondazione Bruno Kessler (FBK), Trento, Italy

*Corresponding author:

Alessandra Mangia, Hepatology Unit, Fondazione Casa Sollievo della Sofferenza, IRCCS71013 San Giovanni Rotondo

Citation: Andrea Compagner, Valeria Piazzolla, Antonio Napolitano, Francesco Giuliani, Federico Cabitza, and Alessandra Mangia. Machine learning models and HCC risk prediction in patients with HCV cACLD after SVR to DAA. *Journal of Bioinformatics and Systems Biology*. 9 (2026): 73-82.

Received: March 09, 2026

Accepted: March 18, 2026

Published: April 29, 2026

Keywords: HCC; Machine Learning; LSM

Introduction

The widespread use of direct-acting antivirals (DAA) has revolutionized

the treatment of patients with chronic hepatitis C (HCV) infection, achieving outstanding sustained virological response (SVR) rates, regardless of disease severity and safety profiles. Such a favorable option enabled the treatment of large numbers of old patients with compensated advanced chronic liver diseases (cold). However, despite the impressive rates of viral clearance, several studies demonstrated that cold patients continue to be at risk of de novo hepatocellular carcinoma (HCC) [1,2]. Not surprisingly and in line with this evidence, HCC cases are expected to increase by 30% in 2050 [3]. As early detection of HCC is essential for a curative approach, indefinite HCC surveillance is currently recommended after DAA with a difference in target population as Asia-Pacific Associations for the Study of the Liver (APASL), Japanese society and European Association for the Study of the Liver (EASL) recommend surveillance in patients with advanced fibrosis, the American Association for the Study of the Liver (AASLD) only in patients with cirrhosis [4-7]. Overall, the difference between advanced fibrosis and cirrhosis is subtle, and with a pragmatic approach we can include these patients under the umbrella of cold and possibly identify accurate tools for individual risk stratification. Risk stratification is necessary for establishing cost-effective surveillance and increasing patients' compliance with surveillance programs. The analysis of multicenter cohorts has shown that the incidence of HCC after SVR falls below the threshold of 1.32/year, considered cost-effective for surveillance outside the setting of HCV-cured patients [8]. Although prospective real-world data is not available, according to a Markov-based microsimulation, the risk in HCV patients successfully treated has been recently estimated to be 0.7 per 100 person-years (PY) [9]. This evidence further highlights that risk strata should be linked with different surveillance strategies.

For early HCC detection after SVR, to define a risk profile, a single marker is not sufficient and a combination of tools is needed to optimize sensitivity for patients at risk of HCC, regardless of DAA treatment [10]. Usually, the Cox proportional hazards model is used for risk prediction with predictors identified in a multivariate analysis. Risk scores determined by the total of scores assigned to the coefficient of each predictor are available [11-13]. Among several retrospective studies, conducted with the aim of identifying what increases the risk of HCC -in HCV patients after SVR- large heterogeneity emerged, limiting the widespread adoption of the proposed algorithms. Some cohorts do not include only HCV infected patients [11-14] or only DAA treated patients [15]; others do not focus on patients with severe disease carrying the highest risk, while they include all the HCV stages of fibrosis. Only a few studies evaluate non-invasive testing (NIT) for fibrosis as liver stiffness measurements (LSM) by transient elastography (TE). Among algorithms including LSM, Pons et al and Semmler et al [13-16] studied a population of cold; while Pons et al included DAA treated patients, Semmler et al also included

patients who had achieved SVR after Interferon. Moreover, the predictive performance of LSM by TE may differ at baseline and after DAA therapy. To increase heterogeneity, liver-related biochemical parameters, to be combined with LSM to increase the accuracy of prediction, also exhibit dynamic changes [17]. Finally, the identification of the optimal time point when to use this variable is critical. By utilizing selected biomarkers, we scotomize all the co-factors coexisting in the same individual and we flatten their role in a cancer arising after SVR, if a linear relationship between risk factors and outcomes exists. AI algorithms fit well with the needs of identifying at-risk patients using NITs. The ability of machine learning (ML) algorithms to predict outcomes can be exploited starting from high-quality representative datasets to eliminate the potential for bias [18]. Ioannou et al found that a ML algorithm outperforms conventional regression models in predicting the development of HCC [19]. The aim of the study was to investigate whether developed ML models based on NITs, laboratory parameters and co-morbidities may help to predict HCC occurrence in cold patients successfully treated with DAA for HCV infection and longitudinally followed.

Methods

Study design and patient population

This study focused on a retrospective prospective cohort of patients with HCV infection and baseline evidence of cALD diagnosed and successfully treated with DAA at "Casa Sollievo della Sofferenza", IRCCS, Liver Unit. Eligibility criteria were: age ≥ 18 years, antiviral treatment based on DAA started from December 2014 to January 2024, regular surveillance for HCC, after SVR defined as undetectable HCV RNA 24 weeks after the end of treatment. Remarkably, all the patients had started treatment within 12 weeks from the baseline LSM. HCC surveillance was performed in accordance with the current guidelines [4]. LSM by TE was assessed by two certified operators V.P. and A.M. Co-morbidities were defined as follow: alcohol abuse as >30 g/day for men and > 20 g/day for women; diabetes as at least one discharge diagnosis for T2D or use of anti-diabetic drugs; cardiovascular disease as recent or past myocardial infarction and heart failure. The cohort included a total of 1082 patients, each of whom underwent at least two and up to three visits (T0 at baseline, T1 and T2 after the end of treatment). As patients had a large variability in the time between T0 and T1 visits see (Figure 1), we selected only patients for which this event occurred within 365 ± 30 days from the T0 visit. Additionally, all patients diagnosed with HCC before the T1 visit, as well as patients with incomplete data, were excluded. Overall, 536 patients were included, while the remaining 546 patients were excluded from further analyses. The list of continuous variables, together with their distribution, is reported in Table 1, while the list of categorical variables is reported in (Table-2). Before analysis, categorical variables were one-hot

encoded, using a separate feature for the “Missing” category for the genotype feature. The distributions of times between the T1 visit and HCC diagnosis and liver-related death, are reported in (Figures 1) and (Figure 2), respectively.

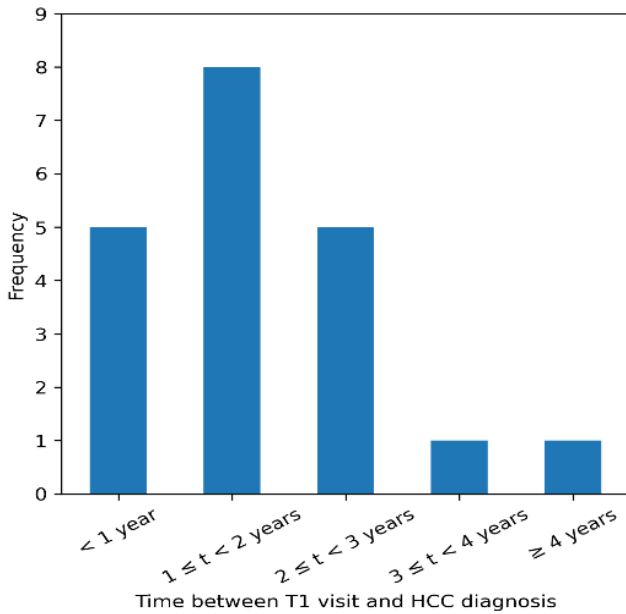


Figure 1: Histogram of the time (in years) between the T1 visit and HCC diagnosis.

Development of ML models

To develop the model several classes of Machine Learning models, namely: logistic regression (LR), decision tree (DT), extreme gradient boosting (XGB), random forest (RF), support vector machine (SVM), and naive Bayes (NB) were tested. Models were developed as three-step pipelines, where the first step was feature normalization (the normalization method was considered as an hyper-parameter to be optimized), followed

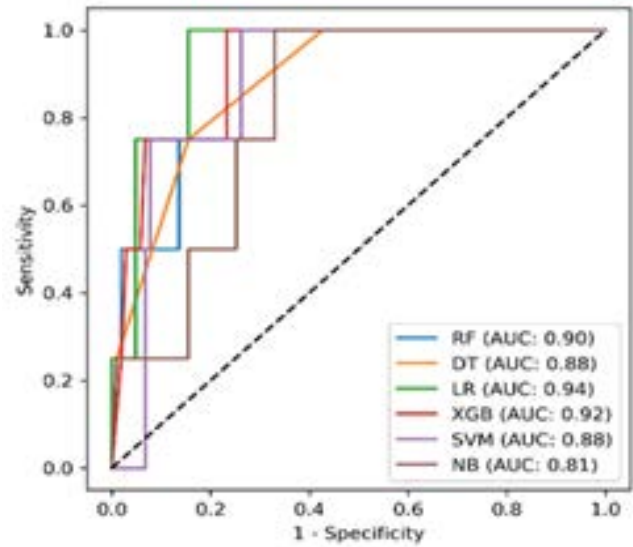


Figure 2: ROC curves for the developed models. The dashed black line represents the chance baseline.

by a feature selection step (using ANOVA F-value between the features and the target as selection score, and setting the number of features to be selected as a hyper-parameter to be optimized), followed by the model itself. The data was split using a hold-out approach, with 80% of the original data used for training (and hyper-parameter optimisation), and the remaining 20% solely for hold-out testing. The size of the test set (108 instances) was estimated to be sufficiently large to detect performance differences between the ML models with a power of 80% (minimum sample size = 81, estimated by Hoeffding inequality) [20]. Training, optimisation and re-training methods are reported as supplementary material. The High-Confidence (HC) counterparts of the confusion matrix metrics (sensitivity, specificity, PPV, NPV) were also Analysed, in order to assess the uncertainty quantification

Table 1: Descriptive statistics for the continuous features. P values have been corrected for multiple hypothesis testing using Benjamini-Hochberg procedure.

Feature name	Mean	St. Dev.	Min	Max	Pvalue
Age (years)	63,54	13,24	25,4	93	0.43
BMI (Kg/m ²)	25,97	3,91	16	47	0.648
Platelet Count x10 ⁹ /L	176,62	84,22	20	652	0.035
Baseline HCV RNA levels (IU/ml)	1716148,81	2453528,36	13	3.00E+07	0.345
ALT (IU/L) (T0)	75,83	68,43	10	552	0.839
AST (IU/L) (T0)	58,23	49,14	12	402	0.345
Albumin (g/dl) (T0)	4,28	0,54	0,3	5,6	0.109
Creatinine (mg/dl) (T0)	0,85	0,45	0,02	9,2	0.072
Glucose (mg/dl) (T0)	107,41	32,04	11	388	0.422
INR (T0)	1,08	0,2	0,3	2,66	0.516
Bilirubin (mg/dl) (T0)	0,82	0,52	0,1	5	0.035
LSM KPa (T0)	16,07	9,72	10,4	75	< 0.001
LSM KPa (T1)	12,71	8,64	3	75	< 0.001

Glucose (mg/dl) (T1)	104,46	29,2	1	399	0.133
INR (T1)	1,1	0,27	0,5	5	0.072
Albumin (g/dl) (T1)	4,42	0,55	1	5,9	0.141
Bilirubin (mg/dl) (T1)	0,81	0,59	0,1	4,7	0.1
ALT (IU/L) (T1)	28,91	24,01	6	281	0.008
AST (IU/L) (T1)	29,01	22,58	7	239	0.005
Platelet Count x10 ⁹ /L (T1)	180,83	81,94	24	683	0.072
ALT Difference	-46,92	68,89	-532	187	0.345
AST Difference	-29,23	50,38	-387	170	0.649
LSM Difference	-3,37	5,1	-58,9	17,8	0.934
Platelets Difference	4,21	49,44	-335	304	0.888
Glucose Difference	-2,95	32,79	-280	197	0.345
INR Difference	0,03	0,25	-1,2	3,96	0.422
Bilirubin Difference	-0,01	0,49	-2,4	3,8	0.345
Albumin Difference	0,14	0,58	-3,21	4,9	0.648

Table 2: Descriptive statistics for the categorical features. N/A means “Not applicable”. P values have been corrected for multiple hypothesis testing using Benjamini-Hochberg procedure.

Feature Name	Categories	P value
Gender	Female: 220, Male: 316	0.345
HCV Genotype	1: 341, 2: 110, 3: 52, 4: 26, Missing: 7	0.195
Combination Treatment	SOF/VEL: 173, SOF/RBV: 159, SOF/LDV: 94, SOF/SMV: 70, Other: 40	0.345
Oesophageal Varices	No: 486, Yes: 50	0.345
MELD score	< 10: 221, >= 10: 36, Not applicable: 279	0.005
CHILD-Pugh score	A: 349, B/C: 21, Not applicable: 166	0.141
Menopause	No: 17, Yes: 203, Not applicable: 316	0.422
Fibrosis stage (F3/F4)	F3: 277, F4: 259	0.005
Obesity	No: 464, Yes: 72	0.667
Cardiovascular damage	No: 466, Yes: 44	0.934
Alcohol abuse	No: 484, Yes: 52	0.345
T2 Diabetes	No: 466, Yes: 70	0.133

capability of the developed models: this approach aimed to exclude cases where the model’s output was effectively a guess or only marginally informative. The high-confidence threshold was set at 75% (for the positive class, corresponding to 25% for the negative class). For all metrics, to account for sampling uncertainty, confidence intervals were computed at a 95% confidence level: for count-based metrics, the intervals were computed as the smallest width resulting from either Hoeffding’s inequality or the binomial approximation method; by contrast, for AUC, Balanced Accuracy, F2 score, Brier score and Net Benefit, the formulas defined by Brodersen et al [21], Cabeza et al [22], and Lam et al [23] were employed.

Results

Characteristics of the study population

Baseline characteristics of the 536 patients included are shown in table 1. Overall 316 (58.9%) were men; mean age was 63.5 ± 13.2 years. During a mean of follow-up of 4.29 years (SD= 5.9), 20 (3.7%) patients developed incidental

HCC. The incidence rate of HCC in the selected population was 0.009 PY, while the incidence rate of liver-related death was 0.004 person-years. The incidence proportion of HCC at 1, 3 and 5 years in the selected population were, respectively, 1.31%, 2.99%, 3.36%.

ML model and relevance of variables

The performance of the models on the hold-out test set is reported in Table 3, while the corresponding ROC curves, decision curves and calibration curves are reported in (Figures 2 and 3). XGB reported the best balanced accuracy, sensitivity and NPV, SVM reported the best specificity, PPV, and Brier score, while LR reported the best AUC. All models were uniformly better than the Treat All baseline at all relevant threshold values (< 0.15, see Figure 3), for almost all such thresholds, LR reported higher Net Benefit than all other models. In terms of calibration (Figure 3), all models were largely uncalibrated, providing over-confident probability estimates across the entire probability range. Focusing on the

high confidence predictions lead to a uniform increment in performance for all models except RF and NB: in particular, LR was the best model across all HC metrics while, at the same time, having high coverage. As there was not uniformly best model, our analysis focused on XGB and LR, as the best black-box and interpretable models, respectively.

Predictive ability of the model and comparison with the conventional score models

DT, XGB, LR and SVM (or superior to) state-of-the-art models and risk scores for HCC prediction (Table-4). Compared with all other models, DT, LR and XGB reported higher AUC. Furthermore, XGB reported higher sensitivity than all state-of-the-art models, with specificity lower than only that of the model proposed by Chang et al [24]. By contrast, DT, LR and SVM outperformed all state-of-the-art models in terms of specificity, while reporting sensitivity lower than only a MAP [12].

We also conducted a post-hoc explainability analysis, using both local and global explanation methods (supplementary). As regards the XGB model, and in terms of local explainability, we applied Shapley value analysis to assess the importance of features: the results are reported in (Figure 3). The most important features were LSM results at T1, with high values of this variable correlating with an increased risk of developing HCC. Higher values of baseline LSM (T0), increased glucose (T1), and AST (T1) levels were also moderately associated with an increased risk of developing of HCC, as was being of older age. As expected, lower Platelets baseline values (T0) were associated with high HCC risk. Interestingly, an increase in Glucose levels was found to be a good predictor of increased HCC risk, independently of baseline levels. Regarding global explainability, a surrogate Decision Tree (DT) model, aimed at describing the behavior of the optimal XGB model was developed. The surrogate DT was trained on the whole training set and reported a balanced fidelity of 92.22%: the DT is visualized in (Figure 4). The importance of

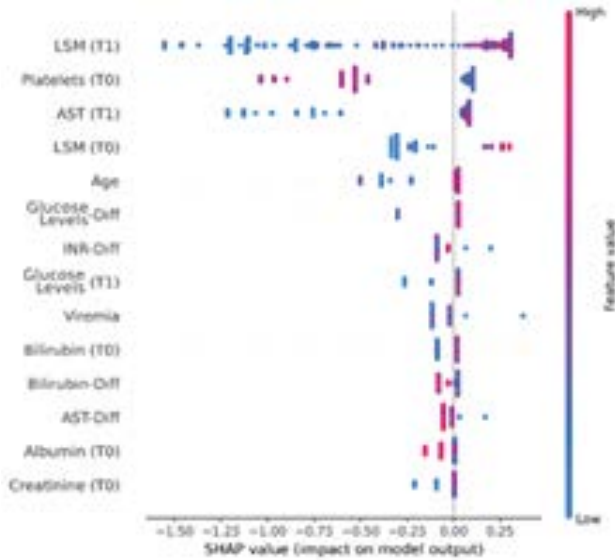


Figure 3: Shapley value explainability analysis for the XGB model. For each feature, dots represent in-instances in the dataset. Red dots correspond to high values for the corresponding feature, while blue dots correspond to low values. Values on the right of the solid gray line are associated with an increased risk of HCC, while values on the left are associated with decreased risk of HCC.

Table 3: Results of the ML models on the hold-out validation set.

	RF	DT	LR	XGB	SVM	NB
Balanced Accuracy	0.69±0.02	0.80±0.02	0.81±0.02	0.84±0.03	0.82±0.02	0.79±0.03
F2	0.49±0.07	0.72±0.07	0.73±0.07	0.92±0.05	0.73±0.07	0.90±0.05
Sensitivity	0.50±0.09	0.75±0.09	0.75±0.09	1.00±0.00	0.75±0.09	1.00±0.00
Specificity	0.87±0.06	0.84±0.07	0.87±0.06	0.68±0.09	0.89±0.06	0.57±0.09
PPV	0.13±0.20	0.16±0.20	0.19±0.20	0.11±0.20	0.21±0.20	0.08±0.20
NPV	0.98±0.03	0.99±0.02	0.99±0.02	1.00±0.00	0.99±0.02	1.00±0.00
AUC	0.90±0.02	0.88±0.02	0.94±0.01	0.92±0.01	0.88±0.02	0.81±0.02
Brier	0.10±0.02	0.16±0.04	0.11±0.05	0.23±0.05	0.04±0.03	0.40±0.09
HC-Sens	0.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.75±0.09	1.00±0.00
HC-Spec	1.00±0.00	0.74±0.08	0.89±0.06	0.61±0.09	0.89±0.06	0.58±0.10
HC-PPV	0.00±0.00	0.16±0.20	0.23±0.20	0.13±0.20	0.21±0.20	0.09±0.20
HC-NPV	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.99±0.02	1.00±0.00
Coverage	0.55±0.09	0.61±0.09	0.91±0.06	0.50±0.09	1.00±0.00	0.96±0.04
sNB	-2.35±0.36	-2.93±0.43	-0.06±0.35	-3.46±0.83	0.00±0.31	-1.16±1.09
sNB	-0.36±0.36	-0.17±0.43	0.54±0.35	-0.36±0.83	0.41±0.31	0.34±1.09
sNB (t: 0.005)	0.87±0.36	0.89±0.43	0.94±0.35	0.87±0.83	0.87±0.31	0.91±1.09

Table 4: Results of state-of-the-art ML models and risk scores for HCC prediction.

Metric	Fan et al	Semmler et al.	Lopez et al.	Chen et al.	Wong et al.	Minami et al.	Sharma et al.
Sensitivity	0.92	-	-	0.73	0.9	-	
Specificity	0.42	-	-	0.83	0.6	-	
PPV	0.03	-	-	0.14	0.16	-	
NPV	0.99	-	-	0.98	0.99	-	
AUC	0.82	0.88	0.81	0.82	0.84	0.84	0.76

the predictive role of LSM results was also confirmed by the global explainability, where LSM results (T1) were selected as the root (most discriminative) feature. Similarly, also Platelets lower than 166.5 mm³ (T0) and increased AST (T1) were reported among the most discriminative features. Since LR is a transparent model, we directly analysed the model's coefficients as a form of global explainability. In particular, the model's coefficients were adjusted by the respective variables' standard deviations, to standardize the variables' contribution to the model predictions. The standardized coefficients are reported in (Figure 4). Similarly to the case of XGB, also for LR, LSM appears (both at T1 and T0) as one of the most predictive features for increased risk of developing HCC. By contrast, the most predictive features were baseline (T0) AST levels, an increase in AST or Glucose levels between T0 and T1. These features were nonetheless found to be among the most important features also by XGB. Similarly to XGB, also for LR, lower baseline (T0) Platelets count was identified as predictive for increased risk of HCC. Additionally, LR identified lower (T0 and T1) ALT as predictors for increased risk of HCC, highlighting the role of the AST/ALT ratio as a good prognostic, independent predictor of HCC occurrence. Finally, we focused on this latter model because, while remaining an inherently explainable and auditable model, it exhibits well-balanced predictive performance and clinically

Table 5: Performance values regarding the LR model, observed at t=10% and, for each metric, the comparison with t=50% via the corresponding difference Δ.

Metric	t = 10%	Δ
Threshold	0.1	-0.4
Prevalence (Base rate)	0.037	0
Specificity (TNR)	0.738	-0.136
Sensitivity (Recall, TPR)	1	0.25
Accuracy	0.748	-0.121
PPV (Precision)	0.129	-0.059
NPV	1	0.011
F1 Score	0.229	-0.071
Balanced Accuracy	0.869	0.057
Symmetric Balanced Accuracy	0.717	0.017
Matthews correlation coefficient	0.309	-0.023
Angular Agreement score	0.6	-0.008
Net Benefit (NB)	0.009	0.102
Standardized Net Benefit (SNB)	0.25	2.75
Treat-all SNB	-1.861	22.889
Treat-none SNB	0	0
Treat-all NB	-0.07	0.855
Treat-none NB	0	0
Brier Score	0.1081	0
Uncertainty (Baseline variance)	0.036	0
Reliability (Calibration error)	0.0802	0
Resolution (Stratification power)	0.0086	0
Scaled Brier	-2.0051	0
ECE (Expected Calibration Error)	0.1411	0
Global ECI	0.873	0
Local ECI	0.9109	0.2537
Total AUC (AUROC)	0.9369	0
Local AUC	1	0.2

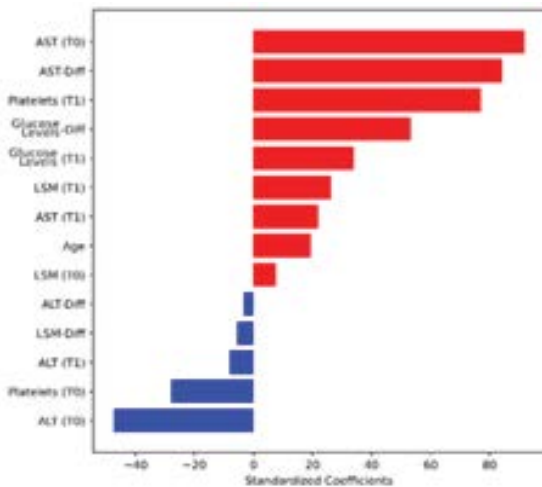


Figure 4: Standardized coefficients for the logistic regression (LR) model. Features depicted in red denote increased risk of HCC, as predicted by the LR model, while features depicted in blue denote decreased risk.

acceptable calibration. At the low decision threshold preferred by the involved clinicians ($t = 10\%$), LR achieves maximal sensitivity and negative predictive value, together with high balanced accuracy and a positive net benefit. This performance profile supports its use in low-threshold decision regimes aimed at early identification or safe rule-out, without incurring substantial calibration degradation. In Table 5, the full set of performance, calibration, and decision-analytic metrics computed at the clinically preferred threshold ($t = 10\%$), alongside the corresponding values at a conventional (default) threshold ($t = 50\%$) and their absolute differences are reported. This makes explicit how threshold choice reshapes model behaviour and clarifies the trade-offs that are most relevant for clinical decision-making.

Discussion

This study explored whether ML models developed for the early prediction of HCC in patients with HCV-related cACLD after SVR following DAA treatment may help in HCC risk stratification. The overarching goal was to assess whether ML models could outperform existing HCC risk scores and improve patient stratification for post-SVR surveillance. In line with recent literature highlighting the limitations of single-marker surveillance strategies [25], our research proposed a composite, data-driven approach based on laboratory and clinical parameters assessed at baseline (T0) and approximately one year later (T1). Prior studies have underscored the residual risk of HCC following HCV clearance, particularly among patients with advanced fibrosis or cirrhosis. Conventional regression-based risk scores have attempted to incorporate clinical, virological, and fibrotic parameters, but their predictive accuracy has been limited by cohort heterogeneity and reliance on static or baseline features. The use of LSM as NIT has demonstrated improved prognostic potential in recent models [16].

However, many previous approaches do not fully leverage the temporal evolution of clinical variables nor the complex, potentially non-linear interactions among them. Our study expands on this by integrating ML methodologies, including LR, RF, XGB, and others—on a well-characterized longitudinal cohort, enabling the capture of dynamic and composite risk patterns. Our main finding is that ML models, particularly XGB and LR, substantially outperformed traditional HCC risk scores in terms of predictive performance. XGB achieved the highest balanced accuracy (0.84), sensitivity (1.00), and negative predictive value (NPV), while LR yielded the best area under the ROC curve ($AUC = 0.94$) and exhibited strong calibration, particularly at clinically relevant thresholds (e.g., 10%). Importantly, LR also demonstrated the most robust performance under high-confidence prediction constraints, suggesting its potential suitability for clinical decision support where interpretability and reliability are critical. These findings are significant

for several reasons. First, they confirm that integrating longitudinal clinical features, particularly LSM at follow-up (T1), alongside dynamic laboratory parameters (e.g., AST, glucose, platelet counts), enhances HCC risk stratification. LSM results at T1 emerged as the strongest predictor in both global and local explainability analyses, reinforcing its value in post-SVR surveillance strategies. A meta-analysis of 8 studies investigating the usefulness of LSM in HCC risk stratification after SVR demonstrated that LSM is a reliable predictor of HCC risk [26]. However, reduction in LSM results relative to baseline may reflect not only fibrosis but also inflammation activity improvement. The longer the follow-up the higher the probability that the observed decrease in LSM results truly represents true fibrosis reduction. In the previously reported meta-analysis the main factor influencing the role of LSM by TE in predicting HCC was the study design [26] with length of follow-up being one of the key aspects. Although it appears, also in our study, that the incidence of HCC after SVR gradually decreases with time (Figure 2), a more precise identification of HCC risk in individual patients over time will have implications for cost-effectiveness and resource use and may increase the reported poor adherence to HCC surveillance [27]. Several studies with shorter follow-up have identified a 20% or 30% reduction from baseline, at variance our models provided an absolute value of <10.9 KPa 1 year after T0 visit, in association with baseline PLT count $>166.5 \times 10^9 \text{ mm}^3$ as the threshold for defying a low risk stratum and reducing surveillance costs.

Interestingly, increases in glucose and AST levels between T0 and T1 (an increase in glucose levels of 5 mg/dl led to a 5-fold increase in HCC risk) were also consistently associated with heightened HCC risk, suggesting that metabolic and inflammatory shifts post-treatment may contribute to carcinogenesis—even after viral eradication [28]. Rather than an overt diabetic or pre-diabetic condition, what requires an intensified surveillance are changes in glucose levels.

Second, while black-box models like XGB offered the highest sensitivity, their limited calibration and explainability may hinder clinical translation. Conversely, the interpretable LR model provided a compelling balance of predictive strength, calibration, and transparency. For example, standardized coefficients highlighted not only the importance of LSM and AST, but also the role of AST/ALT ratios, a marker previously associated with fibrosis progression and liver dysfunction, as an independent HCC predictor. Moreover, LR's high NPV at low-risk thresholds makes it particularly suitable for "rule-out" applications—potentially alleviating the burden of unnecessary surveillance in low-risk patients as non-diabetic subjects with no T1 increase in LSM results. Not unexpectedly, variables such as ALT levels, showed inverse associations with HCC risk and this confirms the limited sensitivity of ALT in patients with advanced disease as a consequence of limited inflammation [29]. Lower ALT

levels at both T0 and T1 were associated with increased risk in the LR model. The AST/ALT ratio's results may suggest the presence of co-morbidities such as fatty liver and/or overt or hidden alcoholic consumption and might be of superior prognostic utility over raw transaminase values. Compared with previous ML-based HCC prediction models, including those by Minami et al. and Wong et al. [30-32] our approach offers a higher AUC (0.94 vs. 0.84) and improved sensitivity, particularly for XGB. Our findings thus substantiate the advantage of integrating both baseline and follow-up data and optimized specifically for DAA successfully treated cold patients, rather than more heterogeneously treated HCV cohorts. In clinical terms, the model may help clinicians in reducing the burden of follow-up visits for patients with reduction of LSM results below the threshold of 9.95 KPa one year after the start of treatment regardless of the treatment duration and in the absence of risk factors such as an increase in glucose levels. A strength of this study was that starting the follow-up at the time of treatment start, but excluding from the analysis all the potential HCC cases already present (although not detectable) by the elimination of the HCC diagnosed within one year from the treatment start, we may avoid the risk of underestimation of incidence as reported when starting follow-up after SVR. Although the involvement of a single centre cohort may be considered a limitation, it offers the advantage of a homogeneous follow-up procedures, the adoption of similar criteria during surveillance and, not irrelevant, the performance of TE by only two dedicated and certified operators. The study offers several implications for future research. First, external validation of the LR and XGB models is essential to confirm their Generalisability across broader geographic and clinical settings. Second, incorporating additional dynamic features such as imaging biomarkers or real-time clinical decision inputs could further enhance model robustness. Third, developing user-facing tools based on the interpretable LR model could facilitate clinical adoption, especially if integrated into electronic health record systems.

Conclusion

This study, demonstrated that the persistent clinical challenge of stratifying HCC risk in patients with cACLD, after SVR following DAA therapy for HCV, may be addressed with the help of ML models—particularly LR and XGB. These models significantly outperform traditional HCC risk scores in predictive accuracy. The LR model offers a practical balance of interpretability, robustness, and high negative predictive value at low-risk thresholds, making it especially suited for rule-out strategies in post-SVR surveillance. These findings suggest that ML-based risk stratification can support more personalized and cost-effective surveillance regimens, reduce unnecessary monitoring while ensure timely detection in high-risk patients. For example, a more expensive RMI based surveillance could be advised in patients with baseline

PLT count $<166.5 \times 10^9 \text{ mm}^3$ in case of LSM $>10.9 \text{ KPa}$, AST and glucose increase 1 year after baseline evaluation. Notably, LSM at follow-up and dynamic shifts in AST and glucose emerged as critical predictors, highlighting the prognostic value of longitudinally collected data. Future research should focus on validating these models across diverse populations, incorporating imaging biomarkers, and translating interpretable models such as LR into clinical decision support tools. This work demonstrates how dynamic, data-driven models can reshape HCC surveillance and improve outcomes for patients with HCV-related liver disease.

Credit Authorship Contribution Statement

AM: conception, data generation and collection, study design, final drafting and approval;

FC: conceptualisation, ML supervision, final drafting, critical revision and approval of the article;

VP: data generation, data management, first draft of the article,

AN: data acquisition, article preparation;

FG: study design;

AC: ML analysis, first draft of the article

Financial support: This study had no specific external financial support.

Conflict-of-interest statement: Alessandra Mangia has served in an advisory role and received research grants from Gilead Sciences, Intercept, Madrigal, Vertex, Akero, Angelini, Madrigal. The other Authors have no conflict of interests to declare.

Etics statement: This research was conducted in accordance with the principles of the Declaration of Helsinki and ethical standards that promote and ensure respect for all participants and protect their health and rights. Every precaution was taken to protect the privacy of research participants and the confidentiality of their personal information.

References

1. Waziri R, Hajari Zadeh B, Greely J, et al. Hepatocellular carcinoma risk following direct-acting antiviral HCV therapy: A systematic review, meta-analyses, and meta-regression. *J Hepatol* 67 (2017): 1204-1212.
2. Cheung MCM, Walker AJ, Hudson BE, et al. Outcomes after successful direct-acting antiviral therapy for patients with chronic hepatitis C and decompensated cirrhosis *J Hepatol* 65 (2016): 741-747.
3. Chan SL, Sun HC, Xu Y, et al. The Lancet Commission on addressing the global hepatocellular carcinoma burden: comprehensive strategies from prevention to treatment. *Lancet* 406 (2025): 731-778.

4. European Association for the Study of the Liver. EASL Clinical Practice Guidelines on the management of hepatocellular carcinoma J Hepatol. 82(2025): 315-374.
5. 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 68 (2018): 723-750.
6. Singal AG, Llovet JM, Yarchoan M, et al. AASLD Practice Guidance on prevention, diagnosis, and treatment of hepatocellular carcinoma. Hepatology 78 (2023): 1922-1965.
7. Hasegawa K, Takemura N, Yamashita T, et al. Clinical Practice Guidelines for Hepatocellular Carcinoma: The Japan Society of Hepatology 2021 version (5th JSH-HCC Guidelines). Hepatol Res 53 (2023): 383-390.
8. Kanwal F, Kramer JR, Asch SM, et al. Long-Term Risk of Hepatocellular Carcinoma in HCV Patients Treated With Direct Acting Antiviral Agents. Hepatology 71 (2020): 44-55
9. Zangneh FH, Wong WWL, Sander B, et al. Cost Effectiveness of Hepatocellular Carcinoma Surveillance After a Sustained Virologic Response to Therapy in Patients With Hepatitis C Virus Infection and Advanced Fibrosis. Clin Gastroenterol Hepatol 17 (2019): 1840-1849.
10. Chhatwal J, Hajjar A, Mueller PP, et al. Hepatocellular Carcinoma Incidence Threshold for Surveillance in Virologically Cured Hepatitis C Individuals. Clin Gastroenterol Hepatol 22 (2024): 91-101.
11. Sharma SA, Kowgier M, Hansen BE, et al. Toronto HCC risk index: A validated scoring system to predict 10-year risk of HCC in patients with cirrhosis. J Hepatol (2017).
12. Fan R, Papatheodoridis G, Sun J, et al. aMAP risk score predicts hepatocellular carcinoma development in patients with chronic hepatitis. J Hepatol 73 (2020): 1368-1378.
13. Semmler G, Meyer EL, Kozbial K, et al. HCC risk stratification after cure of hepatitis C in patients with compensated advanced chronic liver disease. J Hepatol 76 (2022): 812-821.
14. Fujiwara N, Lopez C, Marsh TL, et al. Phase 3 Validation of PAaM for Hepatocellular Carcinoma Risk Stratification in Cirrhosis. Gastroenterology 168 (2025): 556-567.
15. Ioannou GN, Green PK, Berry K. HCV eradication induced by direct-acting antiviral agents reduces the risk of hepatocellular carcinoma. J Hepatol. Published online September 5, 2017.
16. Pons M, Rodríguez-Tajes S, Esteban JI, et al. Non-invasive prediction of liver-related events in patients with HCV-associated compensated advanced chronic liver disease after oral antivirals. J Hepatol 72 (2020): 472-480.
17. López SA, Manzano ML, Gea F, et al. A Model Based on Noninvasive Markers Predicts Very Low Hepatocellular Carcinoma Risk After Viral Response in Hepatitis C Virus-Advanced Fibrosis. Hepatology 72 (2020): 1924-1934.
18. Calderaro J, Seraphin TP, Luedde T, et al. Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma. J Hepatol 76 (2022): 1348-1361.
19. Ioannou GN, Tang W, Beste LA, et al. Assessment of a Deep Learning Model to Predict Hepatocellular Carcinoma in Patients with Hepatitis C Cirrhosis. JAMA Netw Open 3(2020): e2015626.
20. Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301), 13-30.
21. Brodersen KH, Ong CS, Stephan KE, et al. "The Balanced Accuracy and Its Posterior Distribution," 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey 22 (2020): 3121-3124
22. Cabitza F, Campagner A, Soares F, et al. The importance of being external. methodological insights for the external validation of machine learning models in medicine. *Comput Methods Programs. Biomed* 208 (2021): 106288.
23. Lam KFY, Gopal V, Qian J. "Confidence Intervals for the F1 Score: A Comparison of Four Methods." arXiv preprint arXiv:2309.8 (2026): 1751763.
24. Chang, Y. P., Chen, Y. C., & Liu, C. H. (2026). Risk Scores for Stratifying Hepatocellular Carcinoma and Optimizing Surveillance Strategies. *Cancers* 18 (2026): 158.
25. Reiberger T, Lens S, Cabibbo G, et al. EASL position paper on clinical follow-up after HCV cure. J Hepatol 81 (2024): 326-344.
26. You MW, Kim KW, Shim JJ, et al. Impact of liver-stiffness measurement on hepatocellular carcinoma development in chronic hepatitis C patients treated with direct-acting antivirals: A systematic review and time-to-event meta-analysis. J Gastroenterol Hepatol 36 (2021): 601-608.
27. Zhao C, Jin M, Hieu Le R, et al. Poor adherence to Hepatocellular Carcinoma Surveillance Liver Intern 38 (2018): 503-513.
28. Lockart I, Yeo MGH, Hajarizadeh B, et al. HCC incidence after hepatitis C cure among patients with advanced fibrosis or cirrhosis: A meta-analysis. Hepatology 76 (2022): 139-154.

29. Sullivan MK, Daher HB, Rockey DC. Normal or near normal aminotransferase levels in patients with alcoholic cirrhosis. *Am J Med Sci* 63 (2022): 484-489.
30. Wong GL, Hui VW, Tan Q, et al. Novel machine learning models outperform risk scores in predicting hepatocellular carcinoma in patients with chronic viral hepatitis. *JHEP Rep* 4 (2022): 100441.
31. Minami T, Sato M, Toyoda H, et al. Machine learning for individualized prediction of hepatocellular carcinoma development after the eradication of hepatitis C virus with antivirals. *J Hepatol* 79 (2023): 1006-1014.
32. Cheng B, Zhou P, Chen Y. Machine-learning algorithms based on personalized pathways for a novel predictive model for the diagnosis of hepatocellular carcinoma. *BMC Bioinformatics* 23 (2022): 248.



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC-BY\) license 4.0](https://creativecommons.org/licenses/by/4.0/)