Supplementary Table

Supplementary Table 1 Performance comparison of 10 algorithms on the validation cohort.

Algorithm	95% CI		AUROC
PLS	0.698	0.887	0.793
AdaBoost	0.700	0.883	0.792
GBM	0.699	0.882	0.790
KNN	0.679	0.871	0.775
RF	0.676	0.867	0.771
SVM	0.666	0.866	0.766
XGBoost	0.659	0.856	0.757
NNET	0.655	0.857	0.756
GLMNET	0.646	0.853	0.750
RPART	0.629	0.813	0.721

PLS = Partial Least Squares. AdaBoost = Adaptive Boosting. GBM = Gradient Boosting Machine. KNN = K-Nearest Neighbors. RF = Random Forest. SVM = Support Vector Machine. XGBoost = Extreme Gradient Boosting. NNET = Neural Network. GLMNET = Generalized Linear Model with Elastic Net regularization. RPART = Recursive Partitioning and Regression Trees. AUROC = Area Under the Receiver Operating Characteristic Curve.

Supplementary Table 2 Assessment of the MASHRisk score.

	Training cohort	Validation cohort	Test cohort 1	Test cohort 2
Sensitivity	0.729	0.689	0.605	0.779
Specificity	0.738	0.833	0.906	0.689
Positive Predictiv Value	0.686	0.795	0.839	0.104
Negative Predictive Value	0.776	0.741	0.738	0.985
Precision	0.686	0.795	0.839	0.104
Recall	0.729	0.689	0.605	0.779
F1	0.707	0.738	0.703	0.184
Prevalence	0.44	0.484	0.448	0.044
Detection Rate	0.321	0.333	0.271	0.035
Detection Prevalence	0.468	0.419	0.323	0.332

Supplementary Table 3 ICD-10 Codes for Study Outcomes

	ICD-10 code
Liver-related death	B15-B19, C22, K70.0, K70.1, K70.2, K70.3, K70.4, K70.9, K71.7, K71.9, K72.1, K72.9, K73.0, K73.1, K73.2, K73.8, K73.9, K74.0, K74.1, K74.2, K74.5, K74.6, K75.1, K75.2, K75.8, K75.9, K76.0, K76.1, K76.2, K76.3, K76.4, K76.5, K76.6, K76.7, K76.8, K76.9, K77.0, K77.8
Cirrhosis	K70.3, K71.7, K72.1, K74.4, K74.5, K74.6, K76.6, I85.0, I85.9, I98.2, I98.3, I86.4
Liver cancer	C22.0, C22.7, C22.8
MASH	K75.8
ASCVD	I21.0, I21.1, I21.2, I21.3, I21.4, I21.9, I22.0, I22.1,

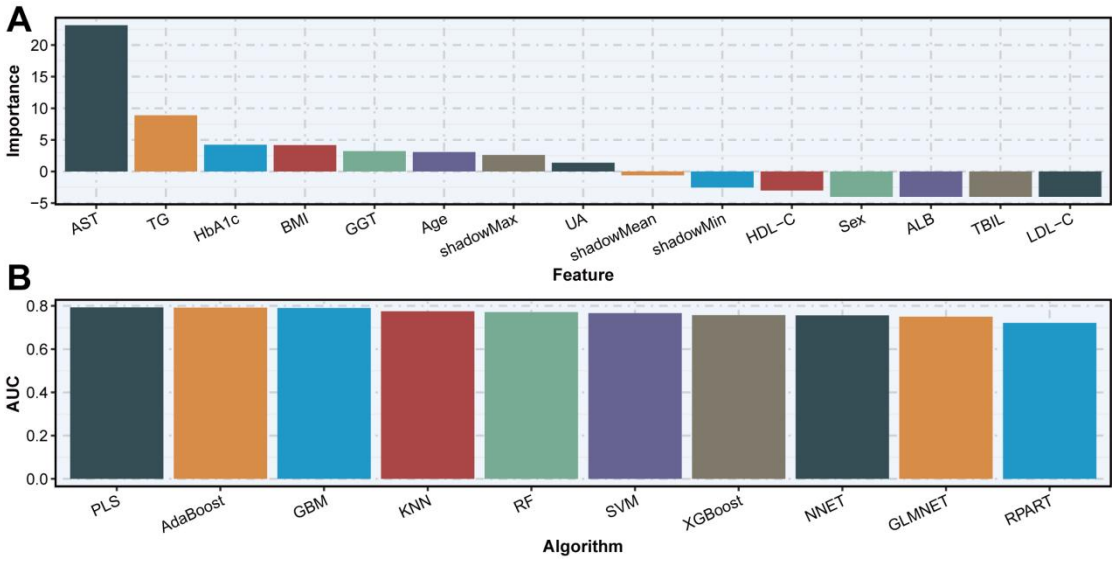
	I22.8, I22.9, I23.0, I23.1, I23.2, I23.3, I23.4, I23.5, I23.6, I23.8, I24.0, I24.1, I25.5, I25.6, I25.8, I25.9, I20.0, I20.1, I20.8, I20.9, I48.0, I48.1, I48.2, I48.9, I74.0, I74.1, I74.2, I74.3, I74.4, I74.5, I74.8, I74.9, I80.0, I80.1, I80.2, I80.3, I80.8, I80.9, I82.2, I82.3, I82.8, I82.9, I50.0, I50.1, I50.9, I60.0, I60.1, I60.2, I60.3, I60.4, I60.5, I60.6, I60.7, I60.8, I60.9, I61.0, I61.1, I61.2, I61.3, I61.4, I61.5, I61.6, I61.8, I61.9, I62.0, I62.1, I62.9, I63.0, I63.1, I63.2, I63.3, I63.4, I63.5, I63.6, I63.8, I63.9, I64, I69.0, I69.1, I69.2, I69.3, I69.4, I69.8

Abbreviations: ICD=International Classification of Diseases. MASH=metabolic dysfunction–associated steatohepatitis. ASCVD=atherosclerotic cardiovascular disease.

Supplementary Figure

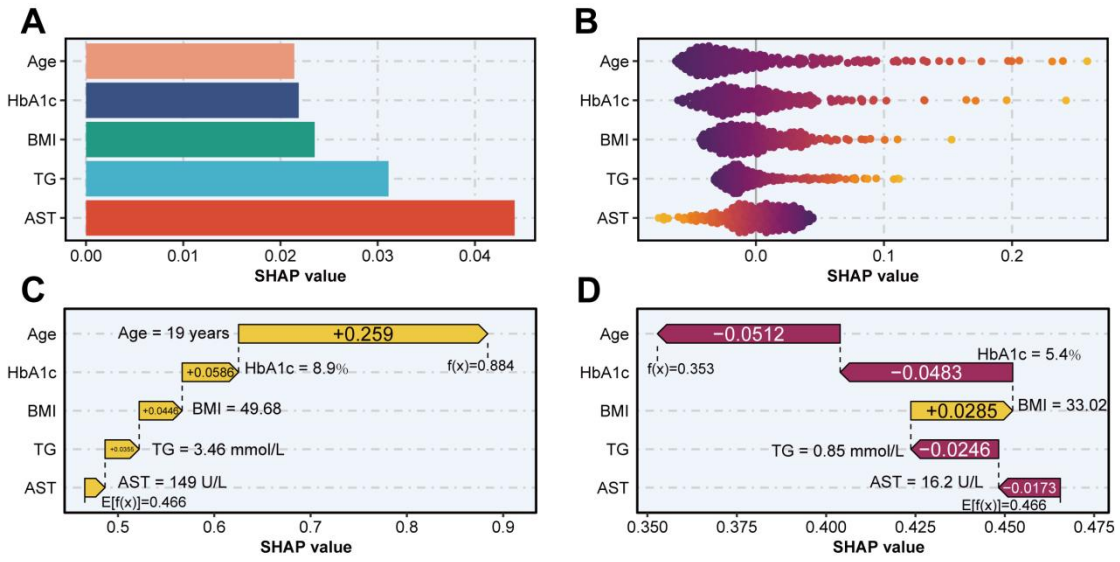
Supplementary Fig. 1 The process of constructing the MASHRisk score.

a. Ranking the importance of features included in the training cohort. **b.** Area under the receiver operating characteristic curve (AUC) ranking of different algorithms.



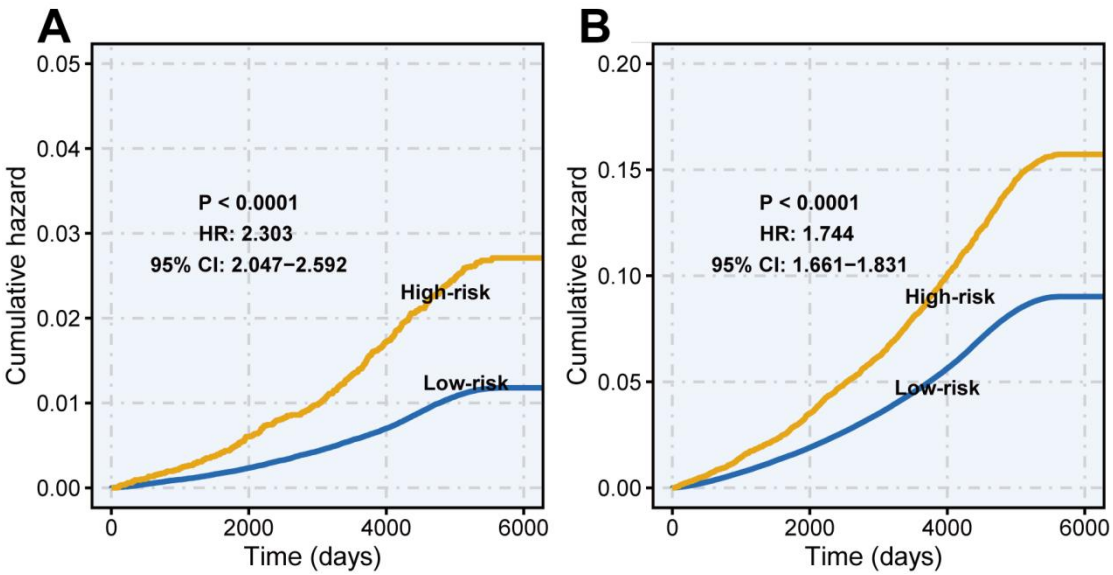
Supplementary Fig. 2 The interpretability of the MASHRisk score.

a. Importance ranking of various variables based on the mean (|SHapley Additive exPlanations (SHAP) value|) derived from the MASHRisk score. **b.** Stability and interpretability of different risk factors assessed with the MASHRisk score. **c.** SHAP value interpretation in a classical sample with metabolic dysfunction-associated steatohepatitis (MASH). **d.** SHAP value interpretation in a classical sample without MASH.



Supplementary Fig. 3. Prognostic value of MASHRisk score in the prognostic cohort.

a. Cumulative hazard plot comparing high-risk and low-risk patient groups over time (in days), with atherosclerotic cardiovascular disease (ASCVD) as the outcome. **b.** Cumulative hazard plot comparing high-risk and low-risk patient groups over time (in days), with all-cause mortality as the outcome.


A
Cumulative hazard
0.05
0.04
0.03
0.02
0.01
0.00
P < 0.0001
HR: 2.303
95% CI: 2.047−2.592
High-risk
Low-risk
0
2000
4000
6000
Time (days)
B
Cumulative hazard
0.20
0.15
0.10
0.05
0.00
P < 0.0001
HR: 1.744
95% CI: 1.661−1.831
High-risk
Low-risk
0
2000
4000
6000
Time (days)