

Supplemental Material - Machine Learning Models Using Non-Invasive Tests & B-Mode Ultrasound to Predict Liver-Related Outcomes in Non-Alcoholic Fatty Liver Disease

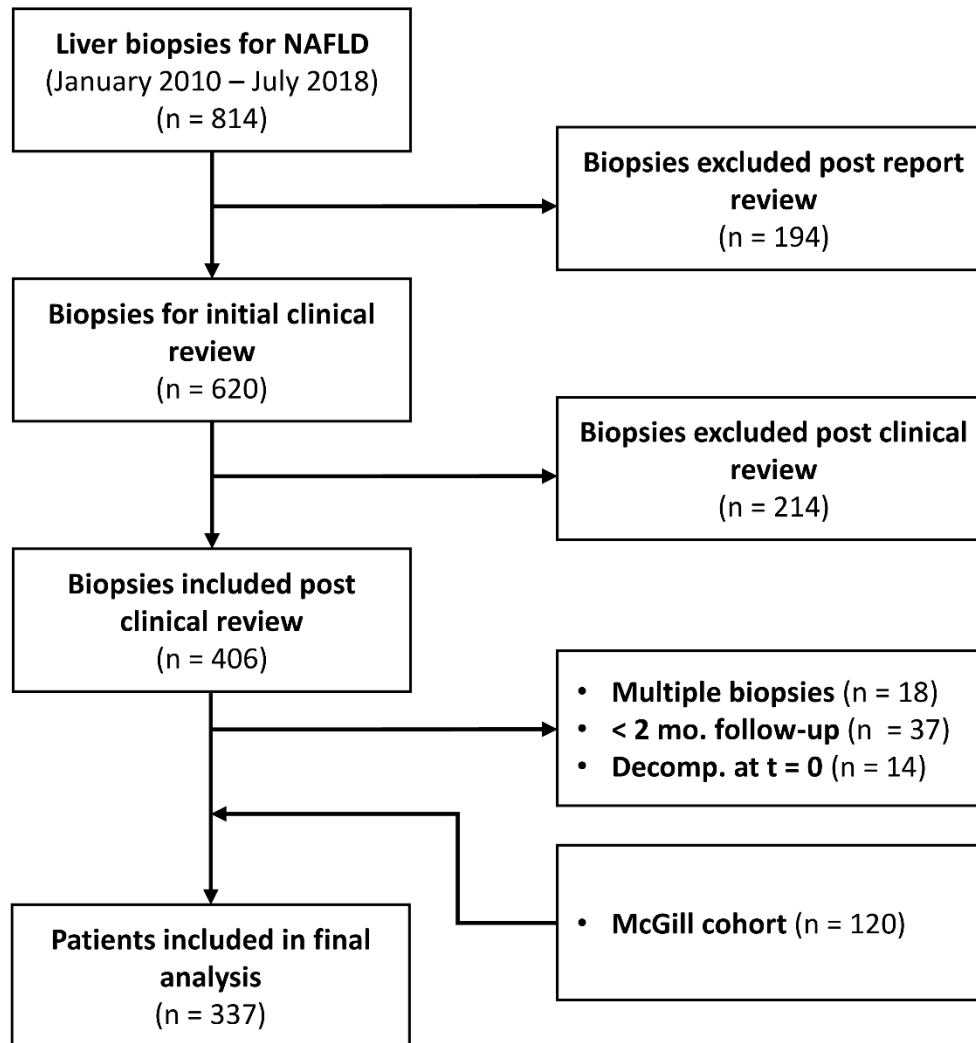
Authors: Heather Mary-Kathleen Kosick, Chris McIntosh, Chinmay Bera, Mina Fakhriyehasl, Mohamed Shengir, Oyedele Adeyi, Giada

Sebastiani, **Kartik Jhaveri**, **Keyur Patel** *Authors names in bold represent shared lead authors.

Table of Contents

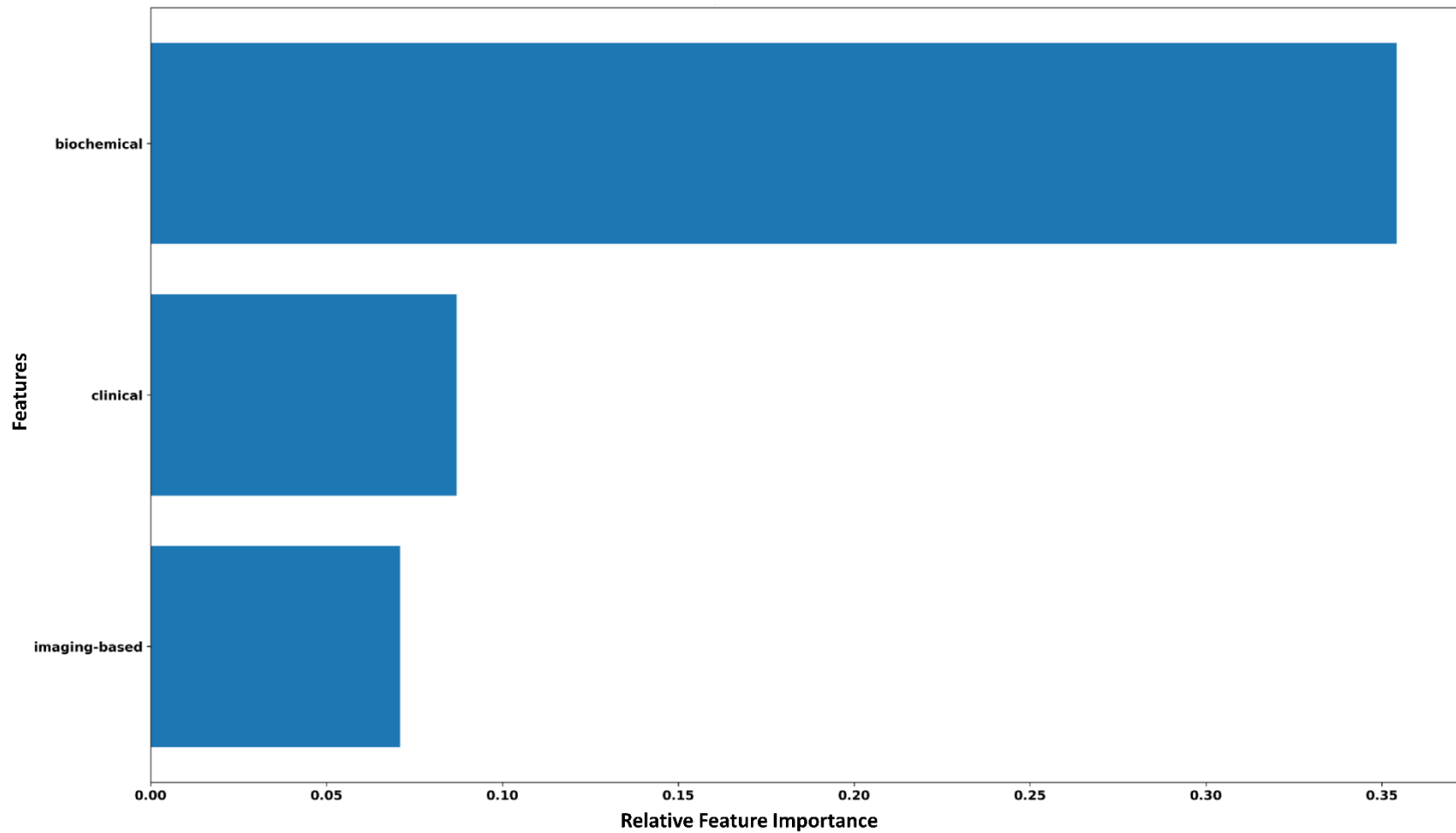
1. Supplementary Figures – 1, 2a-c, 3a-c, 4a-c, 5a-c, 6, 7
2. Supplementary Tables – 1 – 9

Supplementary Figures

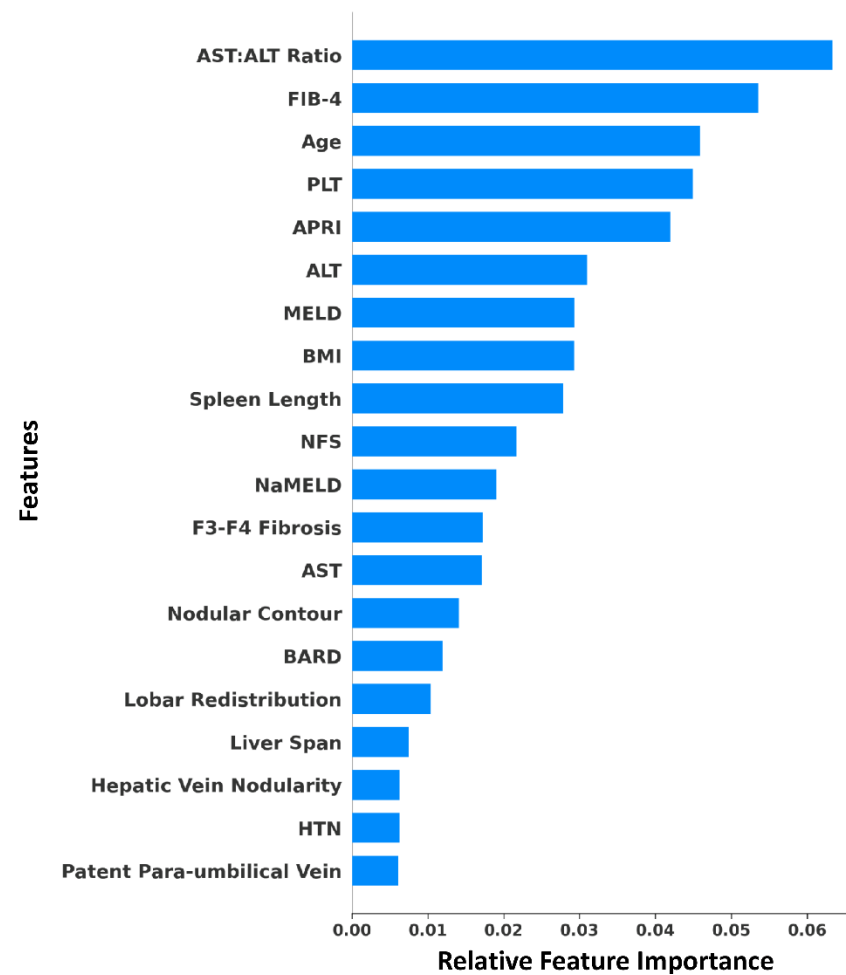


Supplementary Figure 1 – Study Flow Diagram.

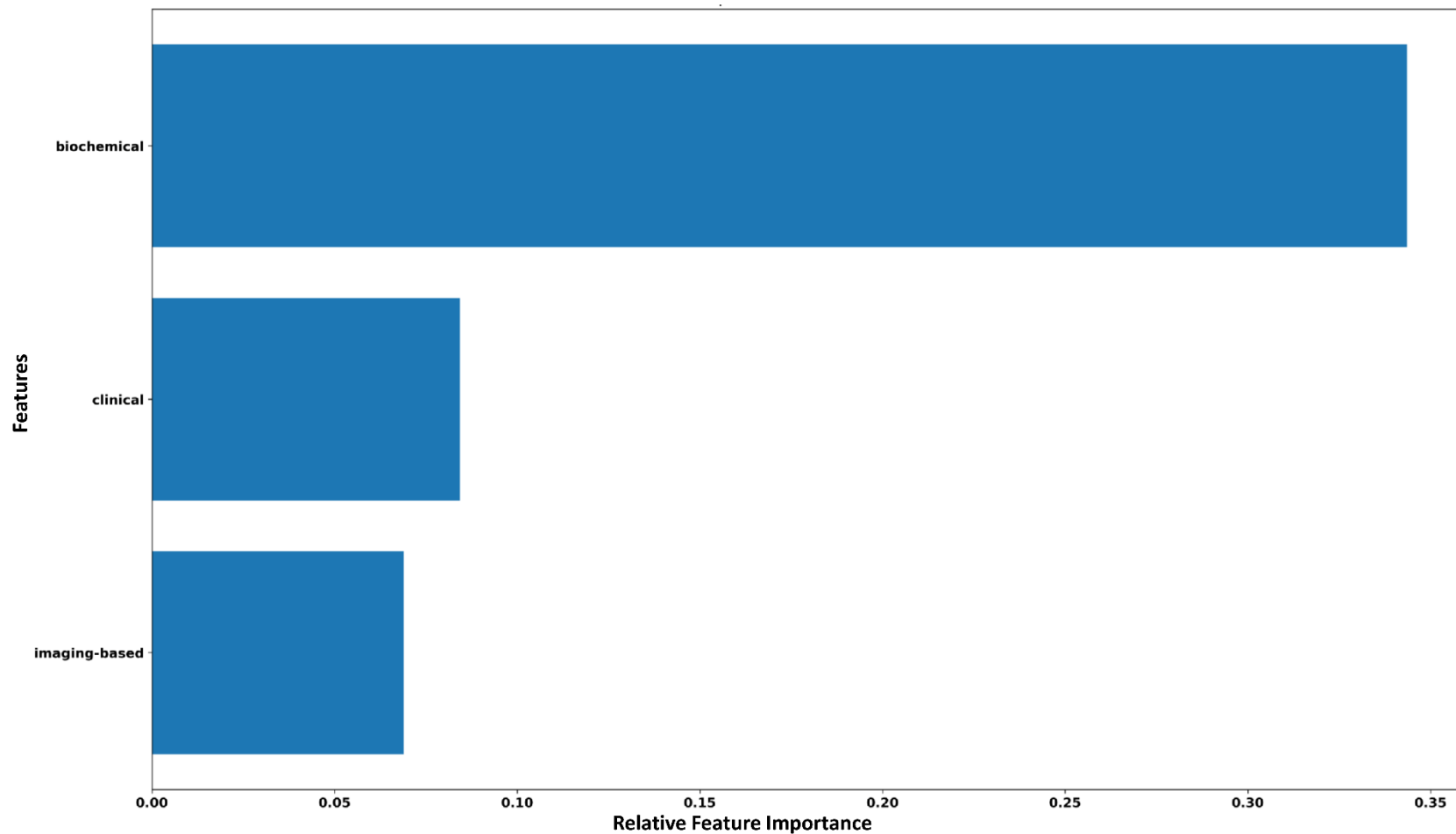
Supplementary Figure 1 represents the study flow diagram, from initial review of all patients undergoing a liver biopsy for NAFLD between 2010 – 2021, to clinical review, based on study inclusion/exclusion criteria.



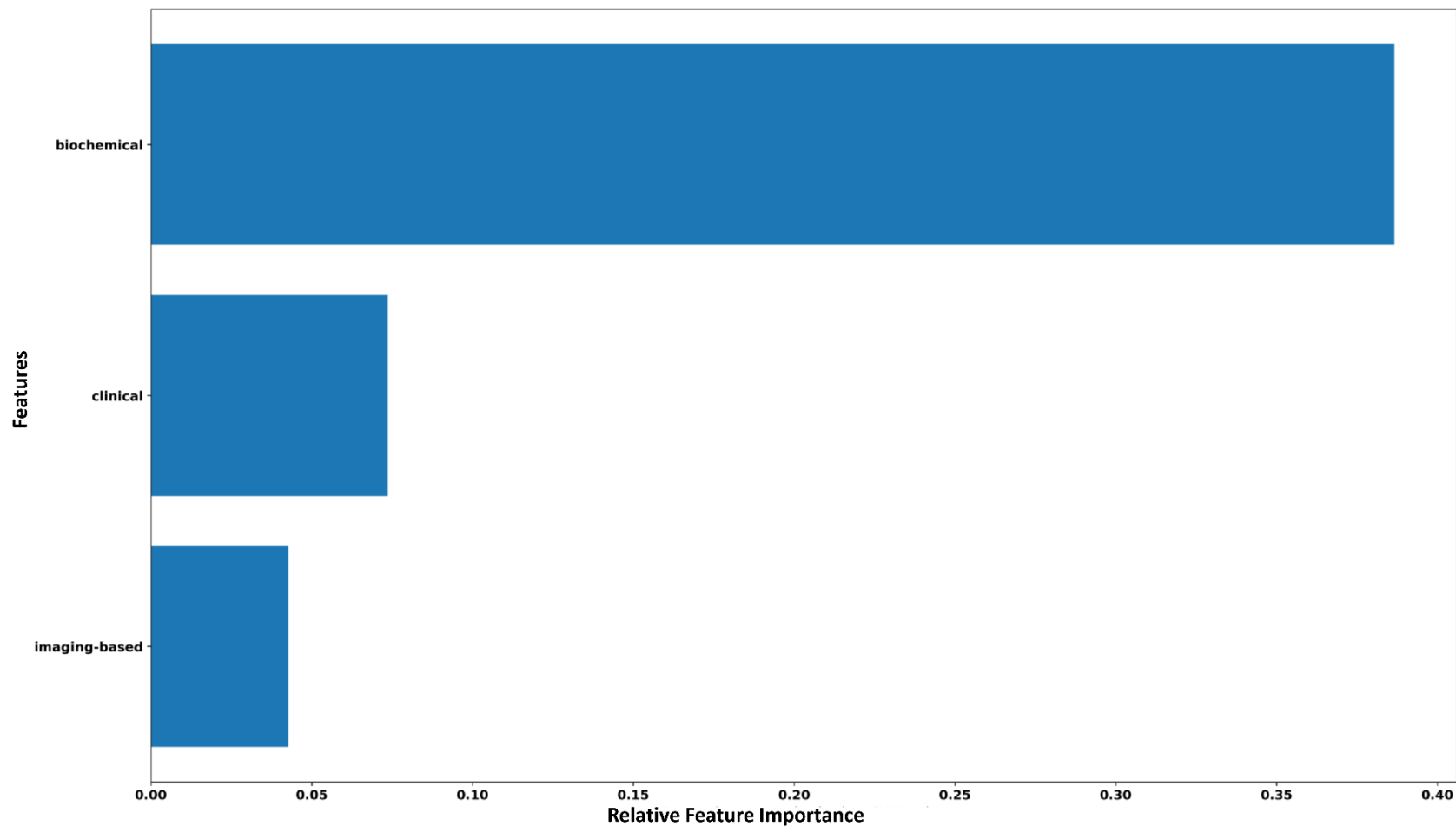
Supplementary Figure 2a – grouped relative feature importance based on allowable predictive variable classification (biochemical, clinical, and imaging-based variables) included in the machine learning model for prediction of outcome ‘liver-related outcomes’.



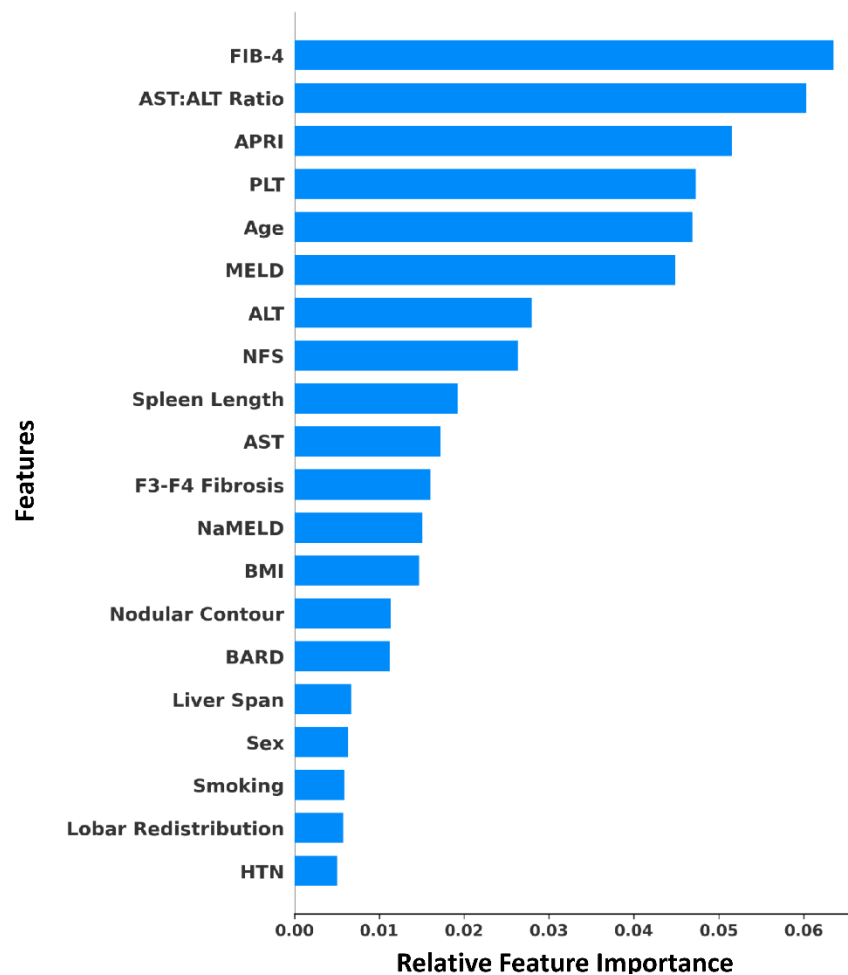
Supplementary Figure 2b – relative feature importance of allowable predictive variables included in the machine learning model for prediction of outcome ‘liver-related outcomes’, including biopsy F3-4 as an allowable predictive variable.



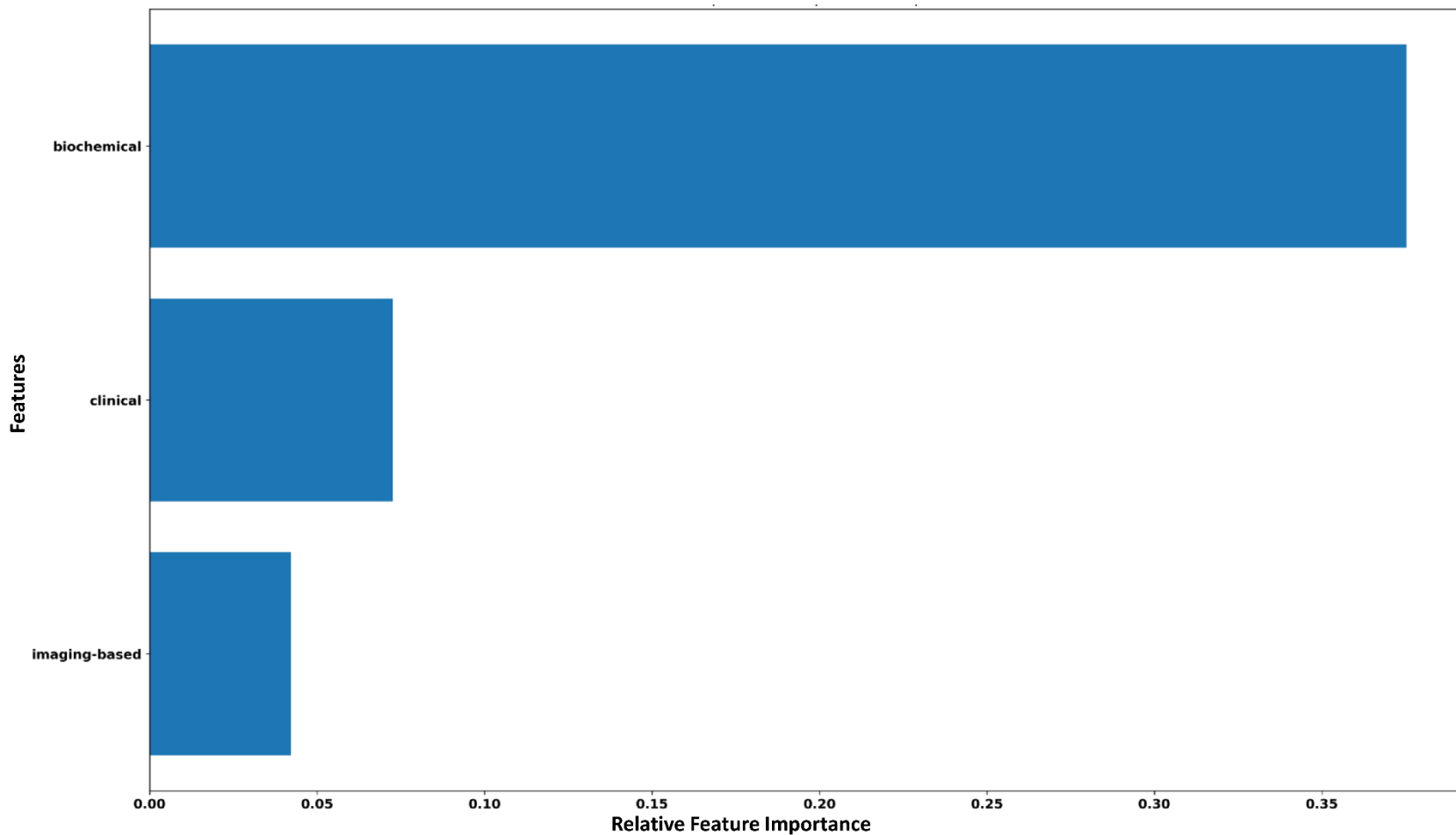
Supplementary Figure 2c – grouped relative feature importance based on allowable predictive variable classification (biochemical, clinical, and imaging-based variables) included in the machine learning model for prediction of outcome ‘liver-related outcomes’, including biopsy F3-4 as an allowable predictive variable.



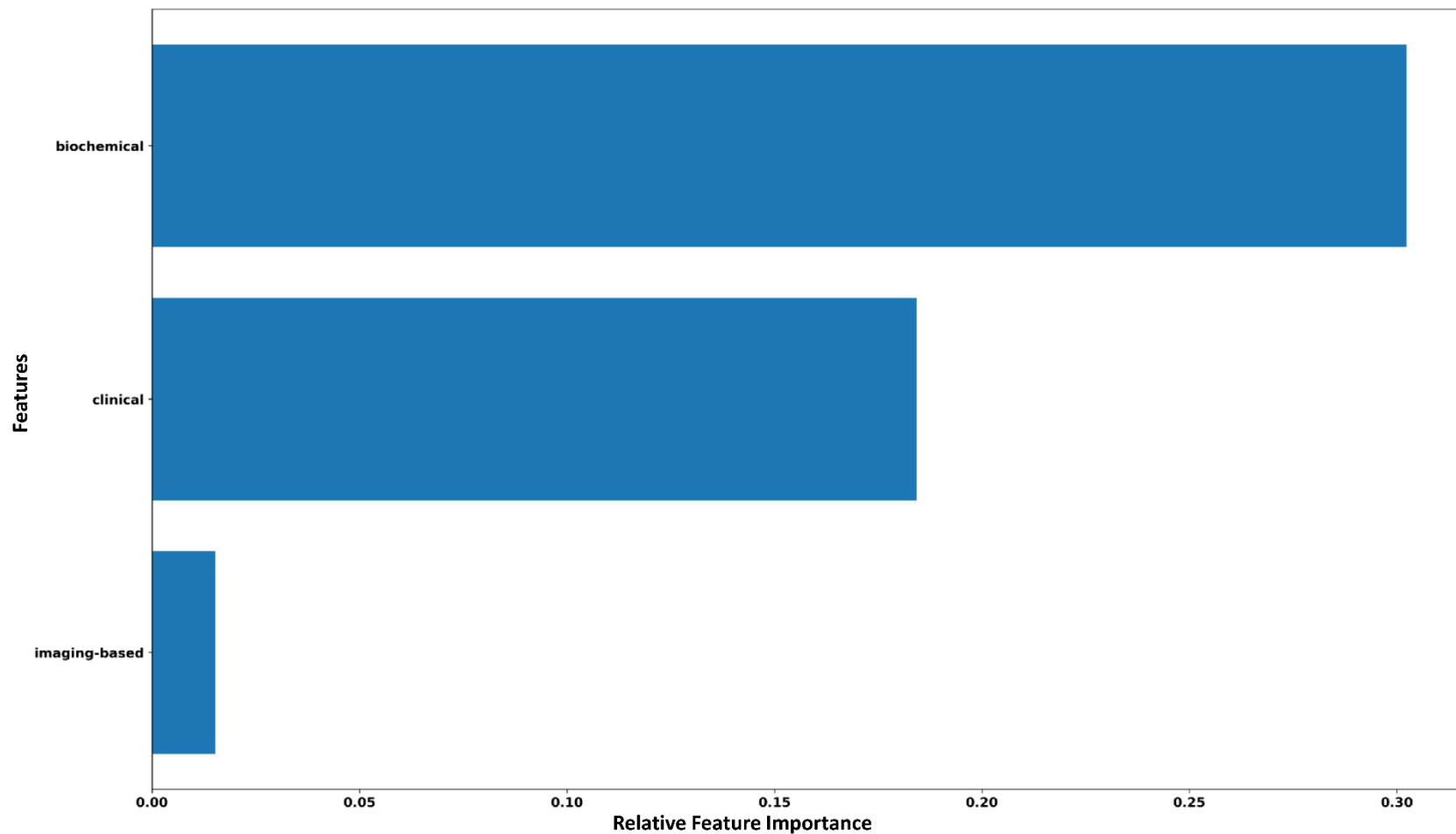
Supplementary Figure 3a – grouped relative feature importance based on allowable predictive variable classification (biochemical, clinical, and imaging-based variables) included in the machine learning model for prediction of outcome ‘hepatic decompensation’.



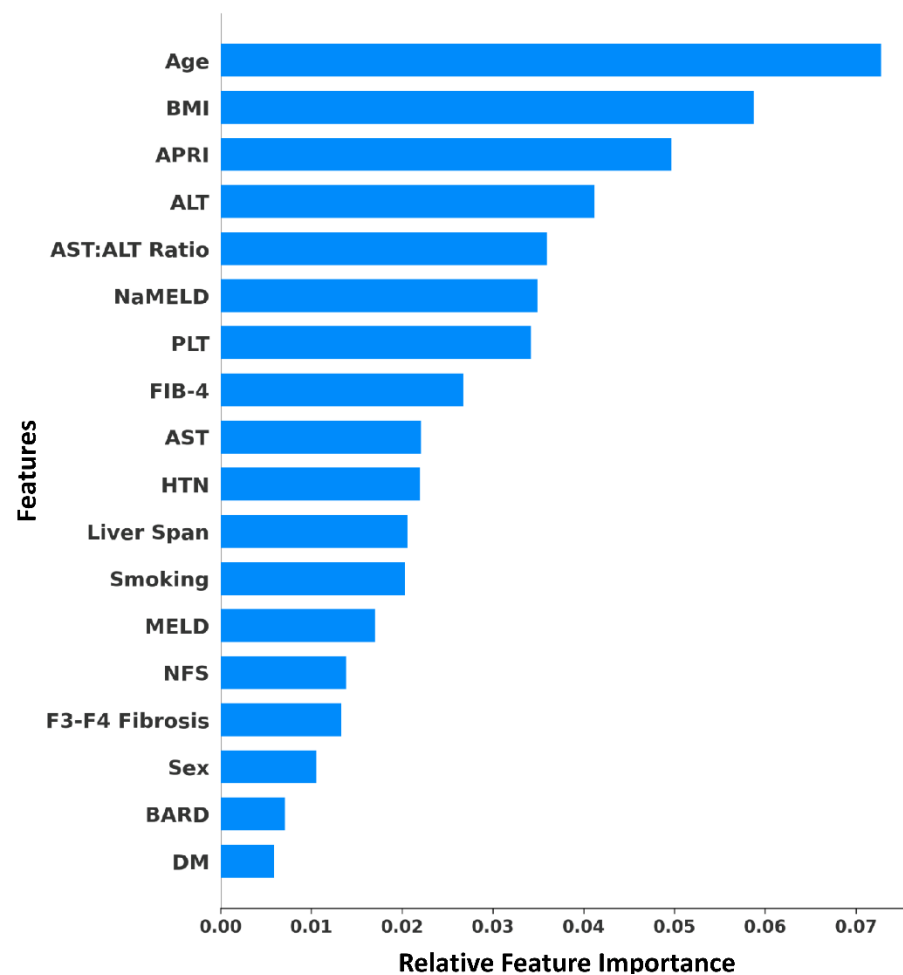
Supplementary Figure 3b – relative feature importance of allowable predictive variables included in the machine learning model for prediction of outcome ‘hepatic decompensation’, including biopsy F3-4 as an allowable predictive variable.



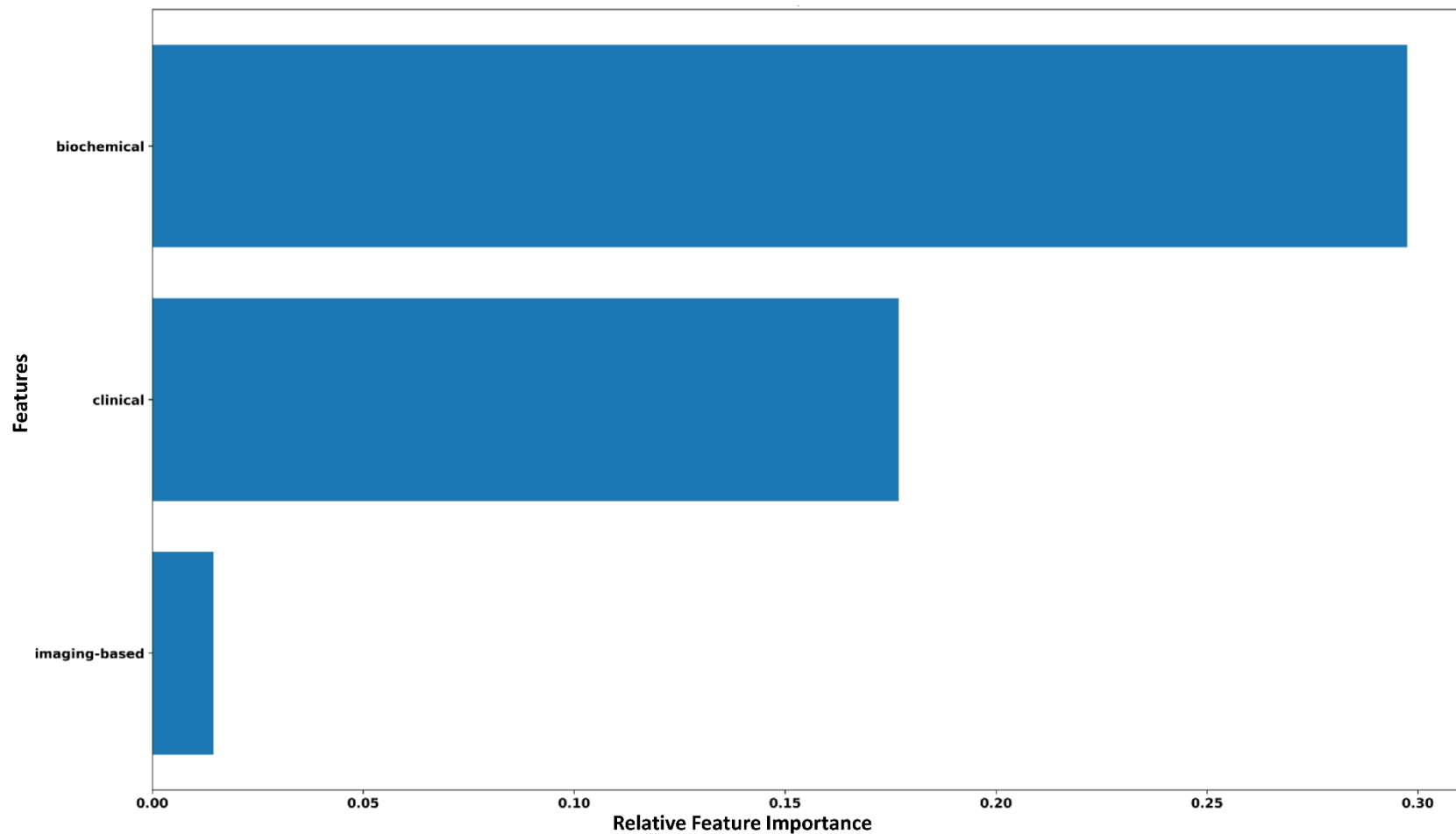
Supplementary Figure 3c – grouped relative feature importance based on allowable predictive variable classification (biochemical, clinical, and imaging-based variables) included in the machine learning model for prediction of outcome ‘hepatic decompensation’, including biopsy F3-4 as an allowable predictive variable.



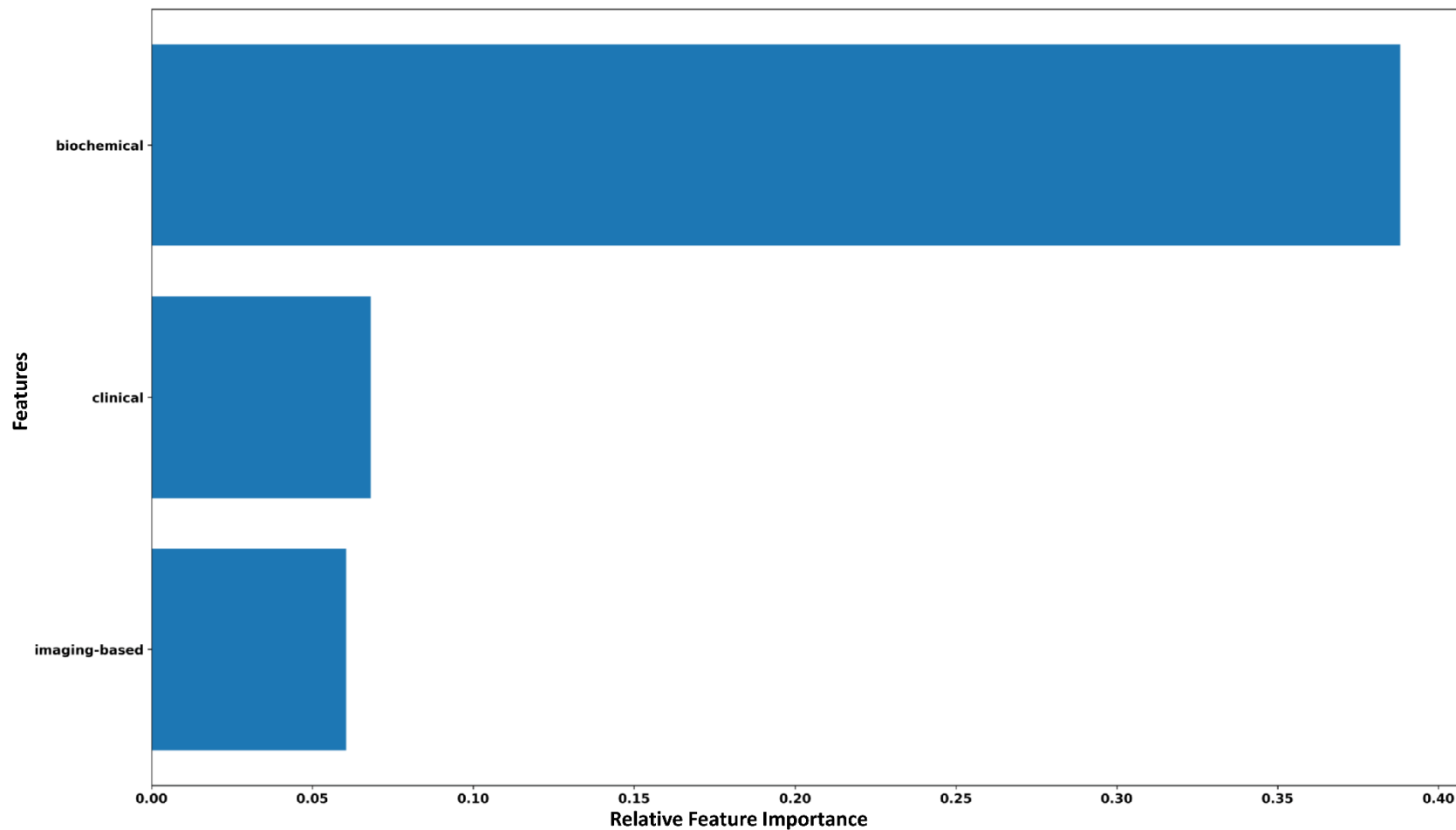
Supplementary Figure 4a – grouped relative feature importance based on allowable predictive variable classification (biochemical, clinical, and imaging-based variables) included in the machine learning model for prediction of outcome ‘hepatocellular carcinoma’.



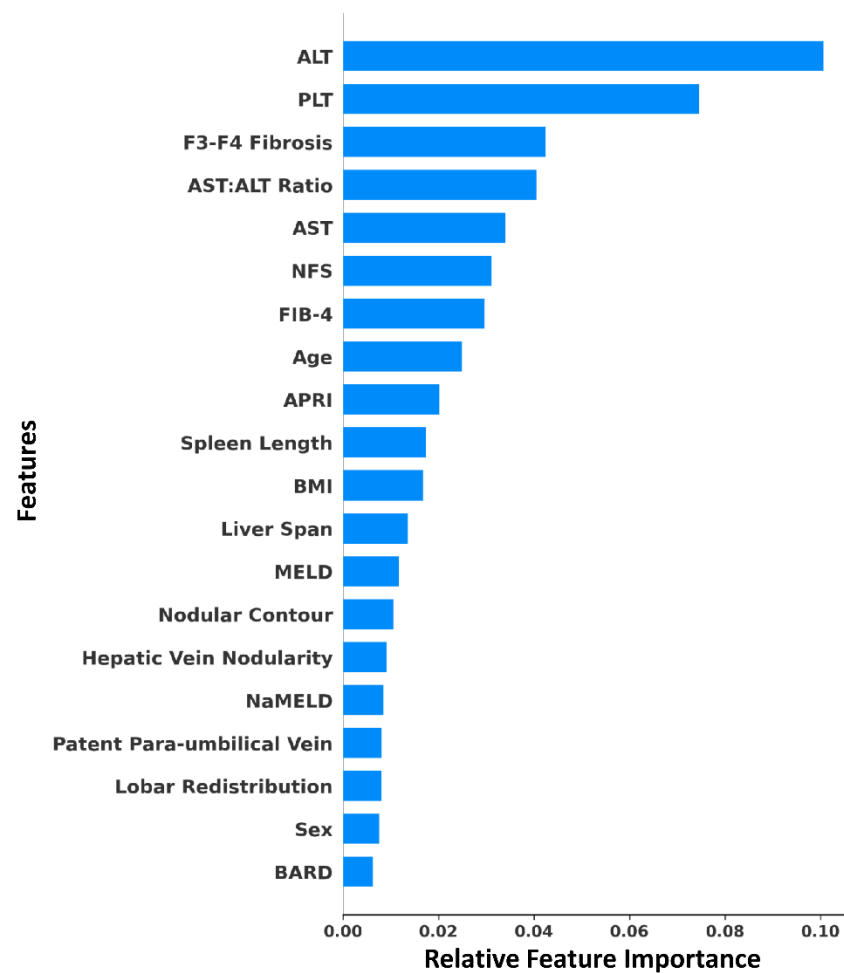
Supplementary Figure 4b – relative feature importance of allowable predictive variables included in the machine learning model for prediction of outcome ‘hepatocellular carcinoma’, including biopsy F3-4 as an allowable predictive variable.



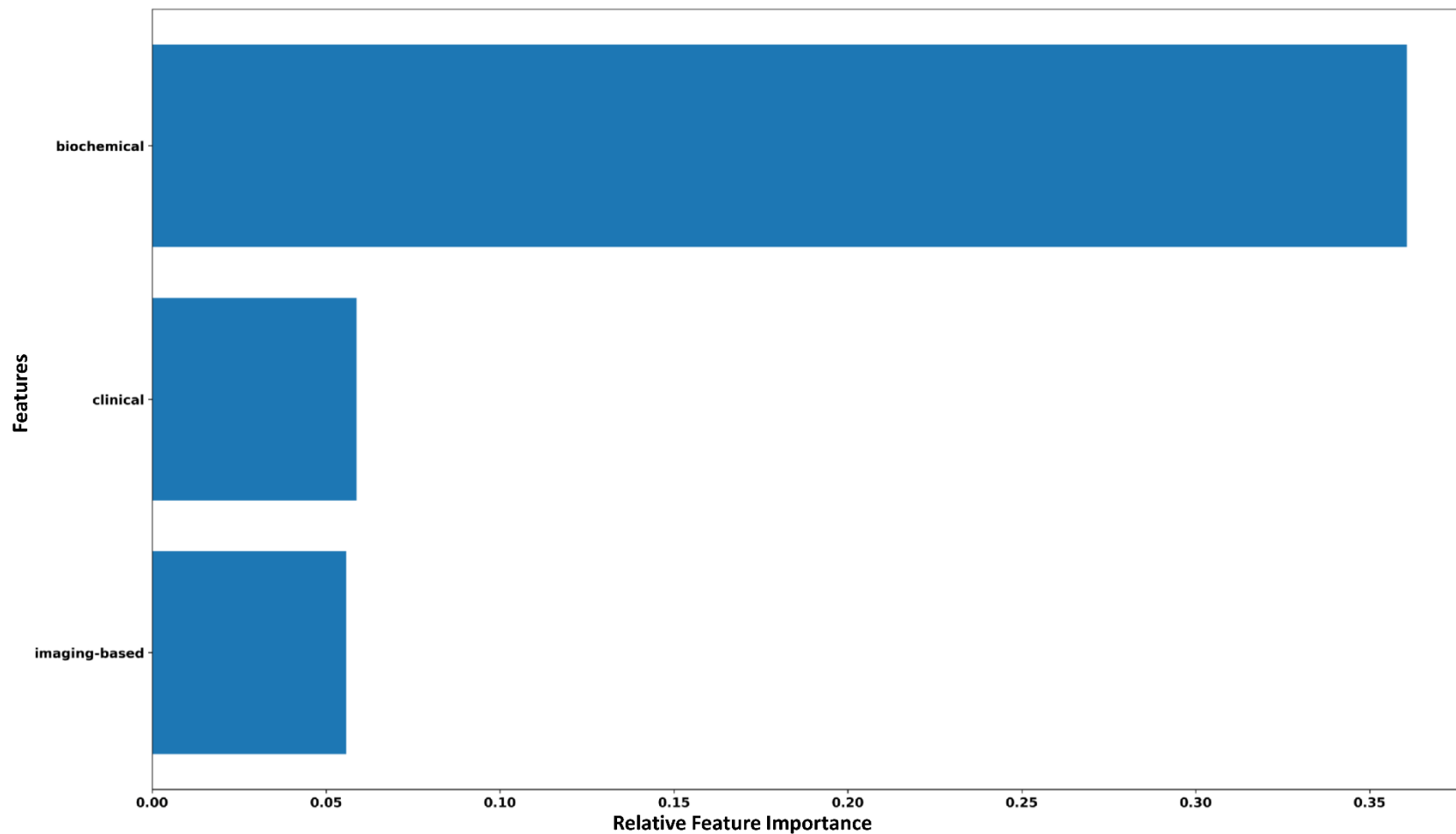
Supplementary Figure 4c – relative feature importance based on allowable predictive variable classification (biochemical, clinical, and imaging-based variables) included in the machine learning model for prediction of outcome ‘hepatocellular carcinoma’, including biopsy F3-4 as an allowable predictive variable.



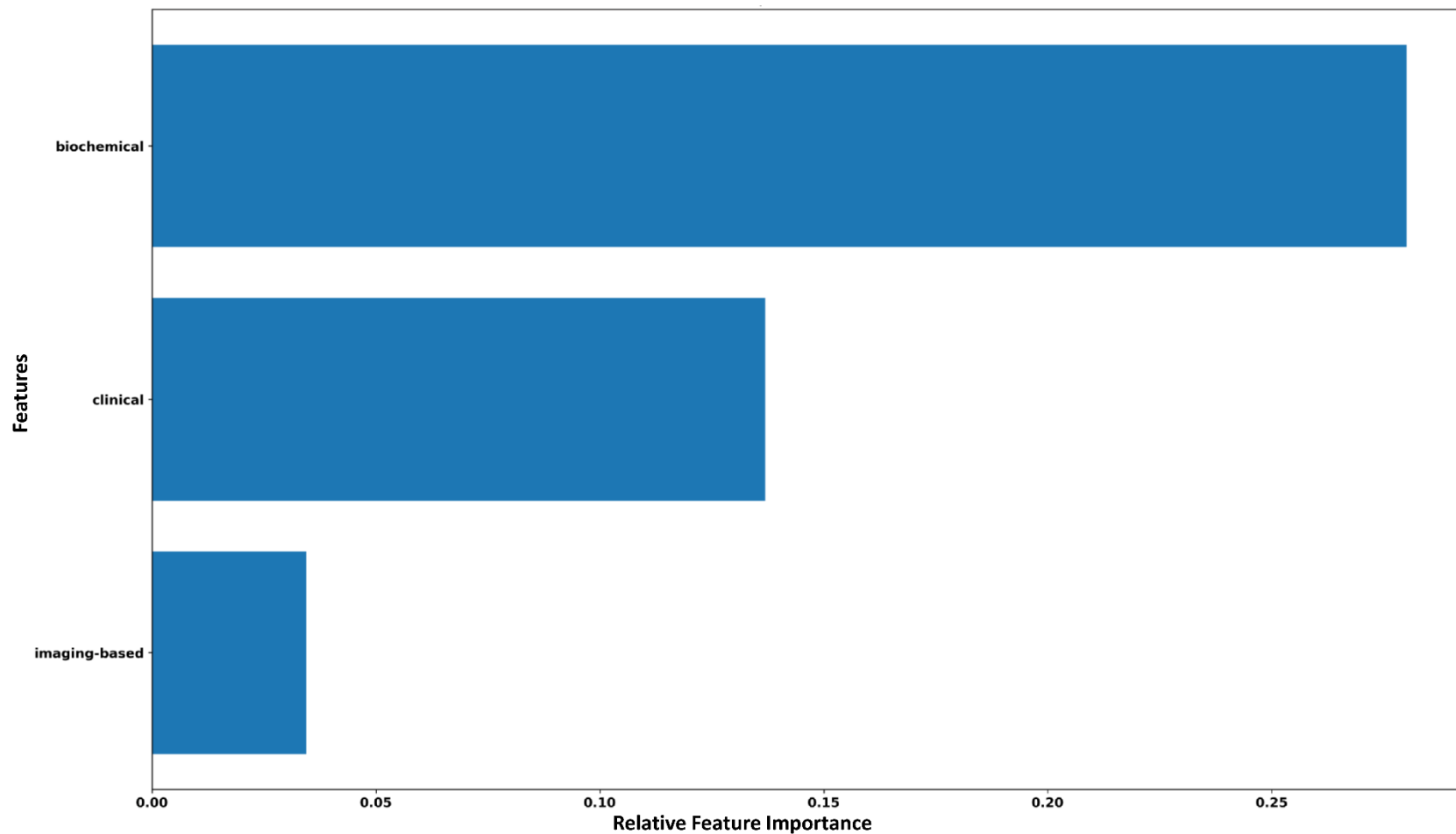
Supplementary Figure 5a – grouped relative feature importance based on allowable predictive variable classification (biochemical, clinical, and imaging-based variables) included in the machine learning model for prediction of outcome ‘liver-related mortality’.



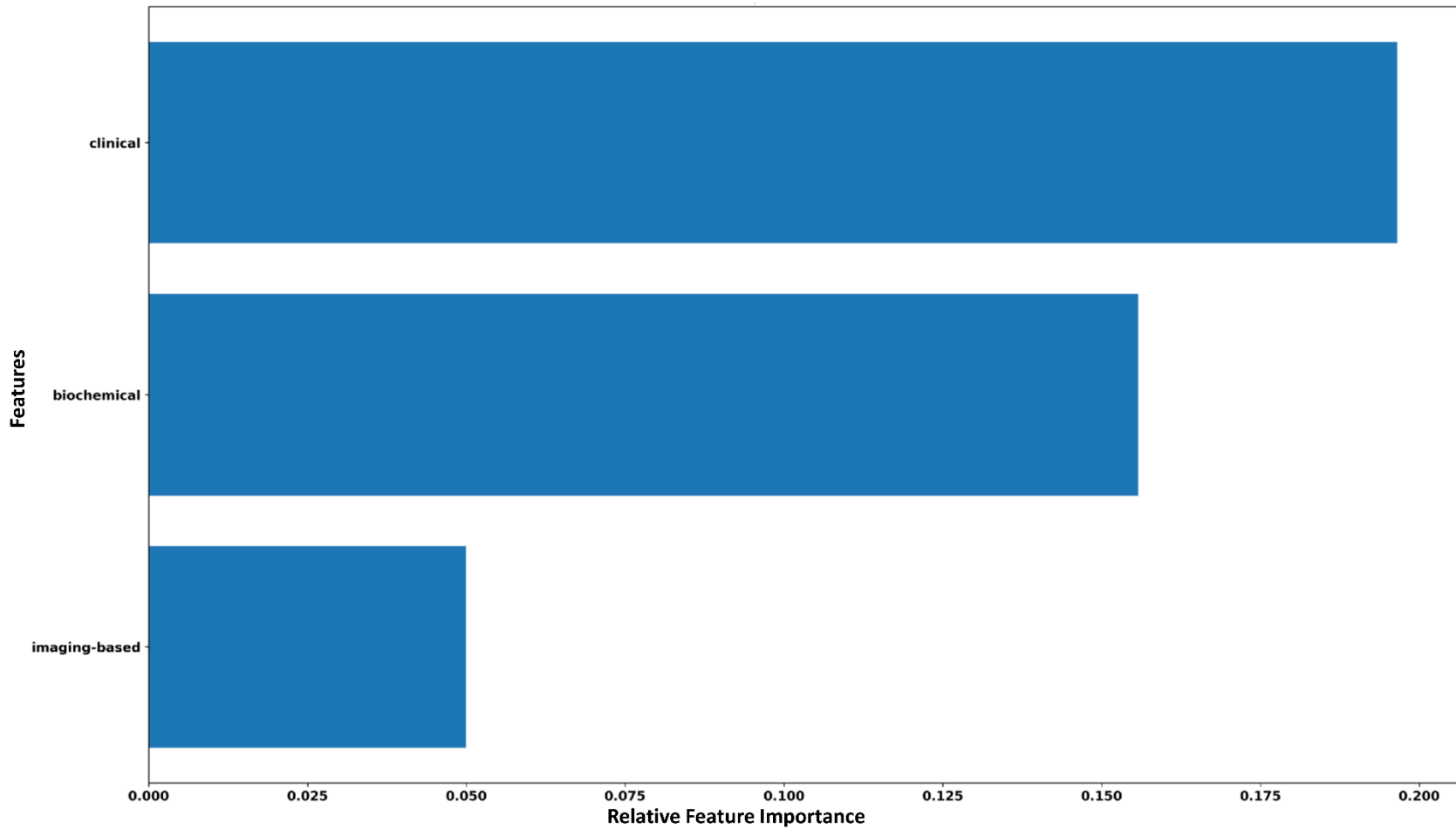
Supplementary Figure 5b – relative feature importance of allowable predictive variables included in the machine learning model for prediction of outcome ‘liver-related mortality’, including biopsy F3-4 as an allowable predictive variable.



Supplementary Figure 5c – grouped relative feature importance based on allowable predictive variable classification (biochemical, clinical, and imaging-based variables) included in the machine learning model for prediction of outcome ‘liver-related mortality’, including biopsy F3-4 as an allowable predictive variable.



Supplementary Figure 6 – Grouped relative feature importance based on allowable predictive variable classification (biochemical, clinical, and imaging-based variables) included in the machine learning model for prediction of outcome ‘F3-4 fibrosis’.



Supplementary Figure 7 – Grouped relative feature importance based on allowable predictive variable classification (biochemical, clinical, and imaging-based variables) included in the machine learning model for prediction of outcome ‘fibrotic NASH’. NASH – non-alcoholic steatohepatitis; NAS – NAFLD activity score.

Supplementary Tables

Supplementary Table 1 - Baseline Clinical Characteristics, Toronto and McGill Cohorts

Patient Characteristic	Toronto (n = 337)	McGill (n = 120)	<i>p</i>
Fibrosis stage	F0-2 – 54.6% (184) F3-4 – 45.4% (153)	F0-2 – 56.7% F3-4 – 43.3%	-
% Males (n)	52.8% (178)	56.7% (68)	0.4623
Age (mean ± SD) [years]	47.9 ± 13.2	52.8 ± 11.4	*0.0003
% Hypertension (n)	32.3% (109)	33.6% (38/113)	0.7989
% Diabetes (n)	28.5% (96)	41.1% (46/112)	*0.0131
% Smoking (reported) (n = 314)	32.1% (97/302)	7.1% (8/112)	*<0.0001
BMI (mean ± SD) [kg/m²]	32.6 ± 7.1 (171)	33.2 ± 6.7 (100)	0.4938
AST (mean ± SD) [U/L]	62.3 ± 55.9 (297)	54 ± 52 (98)	0.1956
ALT (mean ± SD) [U/L]	91.6 ± 78.2 (300)	71 ± 58 (104)	*0.0143

NAS (median, IQR)	4 (3 – 5) (335)	4 (4 – 6) (118)	*0.0025
MELD (median, IQR)	7 (6 – 8) (281)	7 (6 – 8) (90)	0.1170
NaMELD (median, IQR)	8 (7 – 10) (190)	9 (7 – 10) (85)	*0.0384
NFS (n = 163)	-1.67 ± 2.27 (162)	-0.82 ± 1.63 (69)	0.9470
FIB-4 (n = 303)	1.71 ± 1.77 (295)	2.10 ± 1.65 (95)	0.0584
BARD (median, IQR) (n = 182)	2 (1 – 3) (181)	2 (1 – 3) (79)	*0.0213
APRI (n = 303)	0.77 ± 0.79 (295)	0.77 ± 0.65 (97)	1.0000
AST/ALT ratio (n = 305)	0.79 ± 0.44 (297)	0.83 ± 0.35 (96)	0.4176
Liver Span (mean ± SD) [cm] (n = 148)	16.1 ± 2.6 (146)	20.0 ± 3.0 (3)	*0.0113
% Fatty Liver (n = 161)	84.2% (133/158)	88.8% (88/99)	0.3024
% Hepatic Nodularity Contour (n = 161)	19.6% (31/158)	25.4% (15/59)	0.3532
% Hepatic Vein Nodularity (n = 161)	8.2% (13/158)	-	-
% Lobar Redistribution (n = 161)	8.2% (13/158)	-	-

% Patent Para-Umbilical Vein (n = 161)	1.3% (2/158)	-	-
Spleen Length (mean ± SD) [cm] (n = 153)	11.6 ± 2.8 (149)	12.8 ± 2.4 (71)	*0.0021

n – number of patients; * – $p < 0.05$; † – p calculated using Mann-Whitney test; BMI – body mass index; AST – aspartate aminotransferase; ALT – alanine aminotransferase; NAS – NAFLD Activity Score; MELD – Model for End-Stage Liver Disease; NaMELD – MELD sodium score; NFS – NAFLD Fibrosis Score; APRI – AST to Platelet Ratio Index; US – ultrasound; SD – standard deviation; IQR – interquartile range; US – ultrasound.

Supplementary Table 2 – Study Outcomes

Outcome		Total (n = 457)	F0-2 (n = 252)	F3-4 (n = 205)	p
Hepatic Decompensation (n)	Total	4.8% (22)	1.6% (4)	8.8% (18)	*0.0004
	Ascites	3.3% (15)	0.8% (2)	6.3% (13)	*0.0010
	HE	0.8% (4)	0.4% (1)	1.5% (3)	0.2144
	Variceal bleed	0.7% (3)	0.4% (1)	1.0% (2)	0.4344
	Jaundice	0.2% (1)	0% (0)	0.5% (1)	0.2616
	Other	0.2% (1)	0.4% (1)	0% (0)	0.3652
Liver-Related Outcomes° (n)	Total	6.3% (29)	2.4% (6)	11.2% (23)	*0.0001
	Hepatic Decompensation	4.8% (22)	1.6% (4)	8.8% (18)	*0.0004
	HCC	2.0% (9)	1.2% (3)	2.9% (6)	0.1930
	Liver Transplant	0.2% (1)	0% (0)	0.5% (1)	0.2616
	Liver-Related Mortality	1.1% (5)	0% (0)	2.4% (5)	*0.0135
HCC (n)		2.0% (9)	1.2% (3)	2.9% (6)	0.1930
Liver-Related Mortality (n)		1.1% (5)	0% (0)	2.4% (5)	*0.0135

n – number of patients; *– $p < 0.05$; HCC – hepatocellular carcinoma. °Liver-related outcomes – combined clinical outcome of hepatic decompensation (ascites, jaundice, encephalopathy, variceal bleeding) + HCC + liver transplant + liver-related mortality).

Supplementary Table 3 – Allowable Predictors for ML Algorithms

Variable	Classification	Outcomes					
		Liver-Related Outcomes	Hepatic Decompensati on	HCC	Liver-Related Mortality	F3-4 Fibrosis	Fibrotic NASH
Age	Clinical	1	1	1	1	1	1
Sex		1	1	1	1	1	1
HTN		1	1	1	1	1	1
DM		1	1	1	1	1	1
Smoking		1	1	1	1	1	1
BMI (kg/m ²)		1	1	1	1	1	1
AST (IU/L)	Biochemical	1	1	1	1	1	1
ALT (IU/L)		1	1	1	1	1	1
PLT (x10 ⁹)		1	1	1	1	1	1
MELD		1	1	1	1	1	1
MELDNa		1	1	1	1	1	1
NFS		1	1	1	1	1	0

FIB-4		1	1	1	1	1	0
BARD		1	1	1	1	1	0
APRI		1	1	1	1	1	0
AST:ALT		1	1	1	1	1	0
Liver Span (cm)	Radiographic	1	1	1	1	1	1
Fatty Liver		0	0	0	0	0	1
Nodular Contour		1	1	0	1	1	1
Hepatic Vein Nodularity		1	1	0	1	1	1
Hepatic Lobar Redistribution		1	1	0	1	1	1
Patent Para- Umbilical Vein		1	1	0	1	1	1
Spleen Length (cm)		1	1	0	1	1	1

HCC – hepatocellular carcinoma; NASH – non-alcoholic steatohepatitis; HTN – hypertension; DM – diabetes; DLD – dyslipidemia; BMI – body mass index; AST – aspartate aminotransferase; ALT – alanine aminotransferase; PLT – platelet count; MELD – Model for End-Stage Liver Disease score; NaMELD – sodium MELD score; NFS – NAFLD Fibrosis Score; APRI – AST to PLT ratio index. All variables are ‘yes/no’ unless otherwise specified. A ‘1’ under a given outcome indicates that this variable was used as an allowable predictor, while a ‘0’ indicates that it was not. *Grade of fatty liver - 1 = mild, 2 = mild-moderate, 3 = moderate, 4 = moderate-severe, 5 = severe.

Supplementary Table 4 – Performance of NIT for Prediction of Liver-Related Outcomes

	NIT	Sensitivity	Specificity	PPV	NPV	+LR	-LR	AUROC	Youden
Combined Cohort	FIB-4	88.0%	71.8%	17.6	98.9	3.12	0.17	0.84 (0.80 – 0.88)	>1.82
	NFS	93.7%	78.1%	24.2	99.4	4.29	0.08	0.88 (0.83 – 0.92)	>-0.22
	APRI	72.0%	80.9%	20.5	97.7	3.77	0.35	0.75 (0.71 – 0.79)	>1.02
	AST:ALT	84.0%	76.1%	19.3	98.6	3.51	0.21	0.84 (0.79 – 0.87)	>0.92
	BARD	88.9%	69.4%	17.8	98.8	2.91	0.16	0.78 (0.73 – 0.83)	>2
	ML Algorithm	77.4%	77.8%	19.0	98.1	3.49	0.29	0.87 (0.76 – 0.96)	-

NIT – non-invasive tests; PPV – positive predictive value; NPV – negative predictive value; +LR – positive likelihood ratio; -LR – negative likelihood ratio; AUROC – area under the receiver operating curve; NFS – NAFLD fibrosis score; APRI – AST to platelet ratio index.

Supplementary Table 5 – Performance of NIT for Prediction of Hepatic Decompensation

	NIT	Sensitivity	Specificity	PPV	NPV	+LR	-LR	AUROC	Youden
Combined Cohort	FIB-4 (398)	95.0%	71.4%	15.2	99.6	3.32	0.07	0.88 (0.85 – 0.91)	>1.82
	NFS (232)	100.0%	77.5%	21.0	100.0	4.45	0.00	0.93 (0.88 – 0.96)	>-0.22
	APRI (400)	75.0%	80.4%	17.0	98.4	3.82	0.31	0.80 (0.76 – 0.84)	>1.02
	AST:ALT (401)	90.0%	70.8%	14.2	99.2	3.08	0.14	0.85 (0.81 – 0.88)	>0.86
	BARD (261)	86.7%	68.6%	14.4	98.8	2.76	0.19	0.77 (0.71 – 0.82)	>2
	ML Algorithm	88.2%	77.8%	16.7	99.2	3.97	0.15	0.90 (0.79 – 0.98)	-

NIT – non-invasive tests; PPV – positive predictive value; NPV – negative predictive value; +LR – positive likelihood ratio; -LR – negative likelihood ratio; AUROC – area under the receiver operating curve; NFS – NAFLD fibrosis score; APRI – AST to platelet ratio index.

Supplementary Table 6 – Performance of NIT for Prediction of HCC

	NIT	Sensitivity	Specificity	PPV	NPV	+LR	-LR	AUROC	Youden
Combined Cohort	FIB-4	66.7%	79.4%	4.8	99.3	3.24	0.42	0.69 (0.64 – 0.74)	>2.35
	NFS	75.0%	81.5%	6.7	99.5	4.05	0.31	0.70 (0.64 – 0.76)	>0.32
	APRI	66.7%	82.9%	5.7	99.4	3.90	0.40	0.61 (0.56 – 0.65)	>1.18
	AST:ALT	83.3%	74.4%	4.8	99.7	3.26	0.22	0.75 (0.70 – 0.79)	>0.93
	BARD	100.00%	66.4%	4.4	100.0	2.98	0.00	0.80 (0.75 – 0.85)	>2
	ML Algorithm	68.8%	70.7%	4.6	99.1	2.35	0.44	0.72 (0.29 – 0.96)	-

NIT – non-invasive tests; PPV – positive predictive value; NPV – negative predictive value; +LR – positive likelihood ratio; -LR – negative likelihood ratio; AUROC – area under the receiver operating curve; NFS – NAFLD fibrosis score; APRI – AST to platelet ratio index.

Supplementary Table 7 – Performance of NIT for Prediction of Liver-Related Mortality

	NIT	Sensitivity	Specificity	PPV	NPV	+LR	-LR	AUROC	Youden
Combined Cohort	FIB-4	100.0%	71.0%	3.4	100.0	3.45	0.00	0.88 (0.84 – 0.91)	>1.92
	NFS	100.0%	92.1%	14.3	100.0	12.67	0.00	0.97 (0.94 – 0.99)	>1.68
	APRI	100.0%	44.1%	1.8	100.0	1.79	0.00	0.66 (0.61 – 0.71)	>0.46
	AST:ALT	100.0%	76.1%	4.1	100.0	4.18	0.00	0.91 (0.88 – 0.94)	>0.95
	BARD	100.0%	66.1%	3.3	100.0	2.95	0.00	0.82 (0.77 – 0.86)	>2
	ML Algorithm	74.7%	79.8%	4.0	99.6	3.70	0.32	0.79 (0.31 – 0.98)	-

NIT – non-invasive tests; PPV – positive predictive value; NPV – negative predictive value; +LR – positive likelihood ratio; -LR – negative likelihood ratio; AUROC – area under the receiver operating curve; NFS – NAFLD fibrosis score; APRI – AST to platelet ratio index.

Supplementary Table 8 – Performance of NIT for Prediction of F3-4 Fibrosis

	NIT	Sensitivity	Specificity	PPV	NPV	+LR	-LR	AUROC	Youden
Combined Cohort	FIB-4 (398)	73.5%	75.5%	69.8	78.7	3.00	0.35	0.78 (0.74 – 0.821)	>1.32
	NFS (232)	50.5%	92.7%	85.7	68.5	6.82	0.54	0.76 (0.70 – 0.81)	> -0.26
	APRI (400)	77.6%	59.9%	59.7	77.8	1.94	0.37	0.74 (0.70 – 0.79)	>0.46
	AST:ALT (401)	73.4%	65.0%	62.3	75.7	2.10	0.41	0.71 (0.66 – 0.75)	>0.69
	BARD (261)	50.0%	79.4%	68.9	63.5	2.43	0.63	0.71 (0.65 – 0.76)	>2
	ML Algorithm	79.0%	69.7%	68.0	80.3	2.61	0.30	0.83 (0.76 – 0.87)	-

NIT – non-invasive tests; PPV – positive predictive value; NPV – negative predictive value; +LR – positive likelihood ratio; -LR – negative

likelihood ratio; AUROC – area under the receiver operating curve; NFS – NAFLD fibrosis score; APRI – AST to platelet ratio index.

Supplementary Table 9 – Performance of NIT for Prediction of Fibrotic NASH

	NIT	Sensitivity	Specificity	PPV	NPV	+LR	-LR	AUROC	Youden
Combined Cohort	FIB-4 (386)	63.4%	69.8%	68.0	65.4	2.10	0.52	0.68 (0.63 – 0.73)	> 1.3
	NFS (228)	79.0%	44.0%	60.6	65.8	1.41	0.48	0.62 (0.55 – 0.68)	> -2.54
	APRI (388)	73.7%	60.3%	65.0	69.6	1.86	0.44	0.71 (0.66 – 0.75)	> 0.46
	AST:ALT (389)	60.4%	59.4%	60.4	59.4	1.49	0.67	0.60 (0.55 – 0.65)	> 0.7
	BARD (261)	62.6%	47.9%	58.8	51.9	1.20	0.78	0.57 (0.50 – 0.63)	> 1
	ML Algorithm	71.4%	65.7%	68.4	68.8	2.08	0.44	0.74 (0.65 – 0.85)	-

NIT – non-invasive tests; PPV – positive predictive value; NPV – negative predictive value; +LR – positive likelihood ratio; -LR – negative likelihood ratio; AUROC – area under the receiver operating curve; NFS – NAFLD fibrosis score; APRI – AST to platelet ratio index.