Machine learning models for estimating risks of hepatic complications in patients with nonalcoholic fatty liver disease: risk stratification and treatment recommendation (abridged secondary publication)

TCF Yip *, GLH Wong †, VWS Wong, HLY Chan, YK Tse, PC Yuen, QX Tan

KEY MESSAGES

1. A deep learning model incorporating domain adaptation techniques was developed using clinical, laboratory, and medication data to predict liver-related complications in patients with nonalcoholic fatty liver disease (NAFLD).
2. The deep learning model demonstrated accuracy in both a development cohort of patients with NAFLD and an independent validation cohort of patients with type 2 diabetes and probable NAFLD.
3. The deep learning model can guide referrals, further assessments, and intervention recommendations in patients with type 2 diabetes at risk of NAFLD.

Hong Kong Med J 2025;31(Suppl 7):S4-8
HMRF project number: 05190057

[1] **TCF Yip**, [1] **GLH Wong**, [1] **VWS Wong**, [2] **HLY Chan**, [1] **YK Tse**, [3] **PC Yuen**, [4] **QX Tan**

[1] *Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong SAR, China*
[2] *Department of Internal Medicine, Union Hospital, Hong Kong SAR, China*
[3] *Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China*
[4] *Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China*

* Principal applicant: tcfyip@cuhk.edu.hk
† Corresponding author: wonglaihung@cuhk.edu.hk

Introduction

Nonalcoholic fatty liver disease (NAFLD), also known as metabolic dysfunction-associated steatotic liver disease, affects approximately 30% of the population worldwide. Cirrhosis remains the most important risk factor for long-term liver-related complications including hepatic decompensation and hepatocellular carcinoma (HCC) in patients with NAFLD.[1] NAFLD is closely related to type 2 diabetes mellitus (T2DM), which itself is a risk factor for HCC. Given that patients with NAFLD are commonly managed in primary care or endocrine clinics, it is important to stratify patients according to their risk of liver-related complications to enable timely referral, intervention, and HCC surveillance.[2,3] This study aimed to develop a model for stratifying patients with T2DM who are at risk of NAFLD using deep learning for clinical, laboratory, and medication data.

Methods

The International Classification of Diseases, Ninth Revision, Clinical Modification was used for coding diagnoses and procedures.[4] Consecutive patients with a diagnosis of fatty liver were included in the development cohort, whereas consecutive patients with T2DM were included in the validation cohort. Baseline was defined as the earliest date of diagnosis. Patients were excluded if they were aged <18 years, had incomplete demographic data, lacked hepatitis B surface antigen test results, were infected with hepatitis B, C, and/or D, were infected with HIV, had excessive alcohol use or alcohol-related liver disease, had other coexisting autoimmune or metabolic liver diseases, developed liver-related complications within 6 months of baseline, or had <6 months of follow-up. Duplicate cases were excluded from the validation cohort to ensure independence.

The NAFLD ridge score of ≥0.44 was indicative of at risk for NAFLD. Patients not at risk of NAFLD or with type 1 diabetes were excluded. Patients were followed up until the diagnosis of liver-related complications, death, last follow-up (31 July 2021), or 15 years of follow-up, whichever occurred first. The primary endpoint was liver-related events including ascites, spontaneous bacterial peritonitis, variceal bleeding, hepatic encephalopathy, hepatorenal syndrome, HCC, liver transplantation, and liver-related mortality. Baseline laboratory parameters including liver and renal biochemistries, lipid profile, complete blood picture, and blood glucose were collected, as were other relevant diagnoses, procedures, concomitant drugs, and laboratory parameters.

Five domain adaptation models were examined, including adversarial discriminative domain adaptation (ADDA), conditional domain adversarial networks (CDAN), confidence-regularised self-training (CRST), unsupervised domain adaptation by backpropagation (DANN), and maximum classifier discrepancy (MCD) for unsupervised domain adaptation. A total of 39 features including patients' demographics, medication use, comorbidities, and laboratory tests was included. Missing data were imputed using an autoencoder. Feature importance was determined using the permutation method.

The cumulative incidence function of liver-related events, with adjustment for the competing risk of non–liver-related causes, was estimated using Gray's method. The discriminatory performance of the models was assessed using the area under the receiver operating characteristic curve (AUROC) with competing risk. Model calibration was assessed using calibration curves. The optimal single cut-off was determined using Youden's index. The performances of existing scores including the Fibrosis-4 index and NAFLD outcomes score were analysed using established cut-offs.[5] The time-dependent sensitivity, specificity, and positive and negative predictive values of the optimal cut-offs of the five models were reported. All statistical tests were two-sided. A P value of <0.05 was considered statistically significant.

Results

In total, 26 993 patients with NAFLD (mean age, 56.3 years; 46.1% male; 0.6% with cirrhosis) in the development cohort and 411 395 patients with T2DM (mean age, 61.8 years; 50.7% male; 0.4% with cirrhosis) in the validation cohort were included in the analysis (Table).

Respectively in the development and validation cohorts, 281 (1.0%) and 5984 (1.5%) patients developed liver-related events at a median follow-up of 1.9 (range, 1.4-10.3) years and 11.1 (range, 7.4-15.0) years; the respective 15-year cumulative incidences of liver-related events were 3.0% (95% confidence interval [CI]=2.6%-3.4%) and 1.9% (95% CI=1.8%-1.9%).

In the development cohort, the MCD and CRST models had better sensitivity, with a 15-year AUROC of 0.898 and 0.886, respectively, whereas the 15-year AUROCs of the ADDA, CDAN, and DANN models were 0.844, 0.830, and 0.840, respectively (Fig 1). All five domain adaptation models outperformed the Fibrosis-4 index and NAFLD outcomes score. All five models demonstrated reasonable calibration, with slight risk overestimation among high-risk patients. The MCD model performed best in calibration. At the optimal cut-off of 0.197, the MCD model achieved 82.3% (95% CI=80.3%-84.3%) sensitivity, 81.9% (95% CI=81.3%-82.5%) specificity, 12.2% (95% CI=12.1%-12.4%) positive predictive value, and 99.3% (95% CI=99.2%-99.4%) negative predictive value.

Among the 411 395 patients with T2DM in the validation cohort, 26 993 were randomly selected to fine-tune the model hyperparameters through domain adaptation algorithms. The remaining 384 402 patients were used for validation. The AUROCs for sensitivity of MCD, CRST, ADDA, CDAN, and DANN were 0.865, 0.858, 0.855, 0.849, and 0.843, respectively (Fig 1). All five domain adaptation models outperformed the Fibrosis-4 index and NAFLD outcomes score. The MCD model achieved the highest discriminatory power in both cohorts. All five models had good calibration, with subtle risk overestimation among high-risk patients. The MCD model performed best in calibration. Using the optimal cut-off, the MCD model achieved 68.7% (95% CI=68.2%-69.1%) sensitivity, 84.8% (95% CI=84.7%-84.8%) specificity, 7.9% (95% CI=7.9%-7.9%) positive predictive value, and 99.3% (95% CI=99.3%-99.3%) negative predictive value.

The 10 most important features in the MCD model were age, alanine aminotransferase, diabetes, gamma-glutamyl transferase, cirrhosis, aspartate aminotransferase, creatinine, use of thiazide diuretics, platelets, and use of sulphonylureas (Fig 2). Age, cirrhosis, diabetes, gamma-glutamyl transferase, and creatinine were selected by all five models.

Discussion

Although the incidence of liver-related complications was lower in patients with NAFLD than in those with chronic viral hepatitis, the high prevalence of NAFLD was due to a substantial total number of patients at risk of liver-related complications. Identifying patients with NAFLD at risk of developing liver-related complications may facilitate timely referrals and interventions. In clinical practice, patients with NAFLD are managed in primary care or endocrine clinics. Clinicians need to identify those who would benefit from referral to secondary care. Experts have suggested various care pathways for risk stratification and management of such patients. These pathways recommend screening for NAFLD with advanced fibrosis in at-risk patients including those with T2DM. The Fibrosis-4 index, an extensively validated serum fibrosis score, is used as a screening tool to exclude advanced fibrosis in primary care and endocrine clinics.[3] However, its lower AUROC indicates less accuracy in detecting advanced fibrosis in patients with T2DM. We therefore developed deep learning models to predict the risk of developing liver-related complications in patients with NAFLD. Through domain adaptation, TABLE. Baseline characteristics of patients in the development and validation cohorts.

Characteristic	Development cohort (n=26 993)*	Validation cohort (n=411 395)*	P value
Age, y	56.3±13.5	61.8±12.4	<0.001
Male sex	12 432 (46.1)	208 589 (50.7)	<0.001
Cirrhosis	172 (0.6)	1723 (0.4)	<0.001
Diabetes mellitus	11 906 (44.1)	411 395 (100)	<0.001
Hypertension	17 733 (65.7)	195 977 (47.6)	<0.001
Dyslipidaemia	22 547 (83.5)	172 963 (42.0)	<0.001
Renal replacement therapy	198 (0.7)	1207 (0.3)	<0.001
Alanine aminotransferase, U/L	44 (27-69)	23 (16-33)	<0.001
Missing	0	0	
Aspartate aminotransferase, U/L	32 (23-47)	22 (17-31)	<0.001
Missing	42.9	56.2	
Fibrosis-4 index	1.1 (0.7-1.7)	1.2 (0.8-1.9)	<0.001
Missing	43.9	56.2	
Albumin, g/L	41.7±4.7	40.6±5.0	<0.001
Missing	0	9.7	
Total bilirubin, μmol/L	12.1±8.4	11.9±9.8	<0.001
Missing	0	0.1	
Haemoglobin, g/dL	13.7±1.8	13.5±1.8	<0.001
Missing	54.8	0	
White blood cells, $\times 10^9$/L	7.8±2.6	8.7±3.8	<0.001
Missing	54.8	0	
Platelets, $\times 10^9$/L	255.7±74.4	250.2±76.9	<0.001
Missing	3.6	0.1	
Gamma-glutamyl transferase, U/L	57 (34-107)	50 (26-119)	<0.001
Missing	41.1	86.4	
Alpha-fetoprotein, μg/L	2.9 (2.0-4.0)	2.4 (1.7-3.5)	<0.001
Missing	26.2	74.5	
Creatinine, μmol/L	73 (61.3-87.4)	81 (67-97)	<0.001
Missing	0.2	0	
International normalised ratio	1.0±0.2	1.0±0.3	<0.001
Missing	29.6	23.8	
Triglycerides, mmol/L	1.9±1.5	2.0±1.9	<0.001
Missing	2.0	0	
Total cholesterol, mmol/L	4.9±1.1	5.1±1.2	<0.001
Missing	2.0	0	
High-density lipoprotein cholesterol, mmol/L	1.2±0.3	1.2±0.3	<0.001
Missing	2.2	0	
Low-density lipoprotein cholesterol, mmol/L	2.8±1.0	3.0±1.0	<0.001
Missing	2.3	0.1	
Haemoglobin A1c, %	6.7±1.4	8.0±2.0	<0.001
Missing	14.0	0	
Fasting glucose, mmol/L	6.5±2.1	8.5±3.1	<0.001
Missing	1.9	1.1	
Follow-up duration, y	1.9 (1.4-10.3)	11.1 (7.4-15.0)	<0.001
Medication			
Angiotensin-converting enzyme inhibitor/angiotensin receptor blocker	10 932 (40.5)	72 769 (17.7)	<0.001
Aspirin	4629 (17.1)	67 798 (16.5)	<0.001
Beta blocker	10 296 (38.1)	113 874 (27.7)	<0.001
Calcium channel blocker	13 845 (51.3)	100 313 (24.4)	<0.001
Thiazide diuretics	3545 (13.1)	41 334 (10.0)	<0.001
Lipid-lowering agents	11 890 (44.0)	72 783 (17.7)	<0.001
Statin	10 983 (40.7)	56 086 (13.6)	<0.001
Other lipid-lowering agents	2343 (8.7)	22 548 (5.5)	<0.001
Oral hypoglycaemic agents	9390 (34.8)	199 067 (48.4)	<0.001
Metformin	8908 (33.0)	121 459 (29.5)	<0.001
Sulphonylureas	5946 (22.0)	135 122 (32.8)	<0.001
Dipeptidyl peptidase 4 inhibitor	1066 (3.9)	995 (0.2)	<0.001
Sodium-glucose co-transporter-2 inhibitor	235 (0.9)	21 (0.0)	<0.001
Thiazolidinedione	924 (3.4)	603 (0.1)	<0.001
Glucagon-like peptide-1 receptor agonist	38 (0.1)	15 (0.0)	<0.001
Others	329 (1.2)	2945 (0.7)	<0.001
Insulin	1995 (7.4)	18 404 (4.5)	<0.001

* Data are presented as mean±standard deviation, No. (%) of participants, or % of participants.

the MCD model demonstrated accuracy in both the development cohort of patients with NAFLD and the validation cohort of patients with T2DM at risk of NAFLD. The MCD model may be a useful alternative for identifying high-risk individuals with T2DM who require referral.

Our study had several limitations. First,

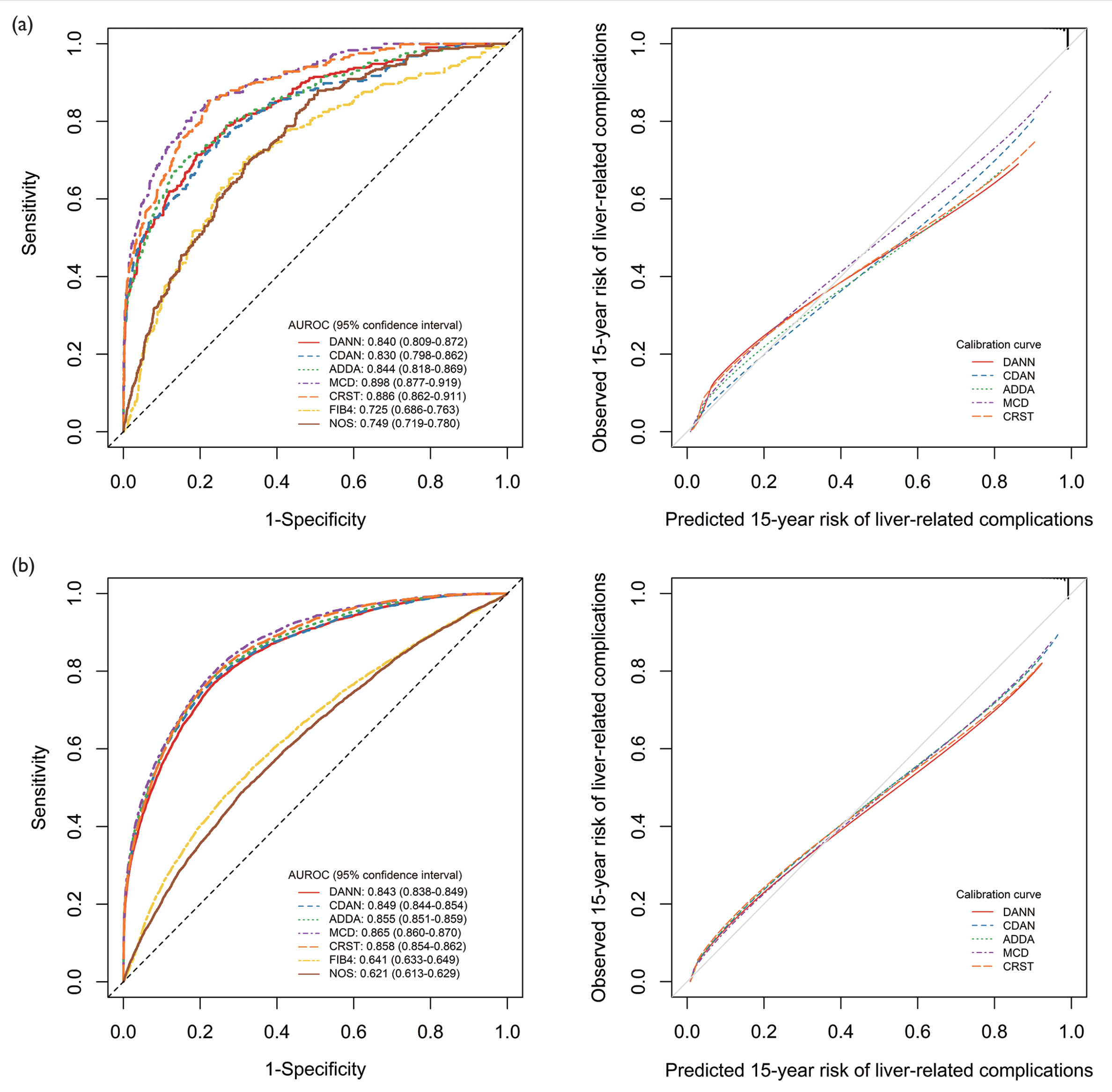

FIG 1. Area under the receiver operating characteristic curves (AUROCs) and calibration curves for predicting liver-related events within 15 years in patients with nonalcoholic fatty liver disease (NAFLD) in the (a) development and (b) validation cohorts.
Abbreviations: ADDA=adversarial discriminative domain adaptation, CDAN=conditional domain adversarial networks, CRST=confidence-regularised self-training, DANN=domain adaptation by back-propagation, FIB4=fibrosis-4 index, MCD=maximum classifier discrepancy, and NOS=NAFLD outcomes score

substantial proportions of laboratory data were missing. The imputation autoencoder was incorporated into the model architecture to minimise bias due to missing data. Second, there was no external validation of the model. Future validation studies should assess the model's generalisability in other populations of patients at risk of NAFLD. Nonetheless, our study had important strengths. In particular, our cohorts included a broader range of patients including those with multiple comorbidities, who are often excluded from other trials. Our findings are therefore more applicable to routine clinical practice. Additionally, we implemented strict exclusion criteria to minimise bias.

Conclusion

We developed and validated a deep learning model based on clinical, laboratory, and medication data to predict liver-related complications in patients with NAFLD and in patients with T2DM and probable NAFLD.

Funding

This study was supported by the Health and Medical Research Fund, Health Bureau, Hong Kong SAR Government (#05190107). The full report is available from the Health and Medical Research Fund website (https://rfs2.healthbureau.gov.hk).

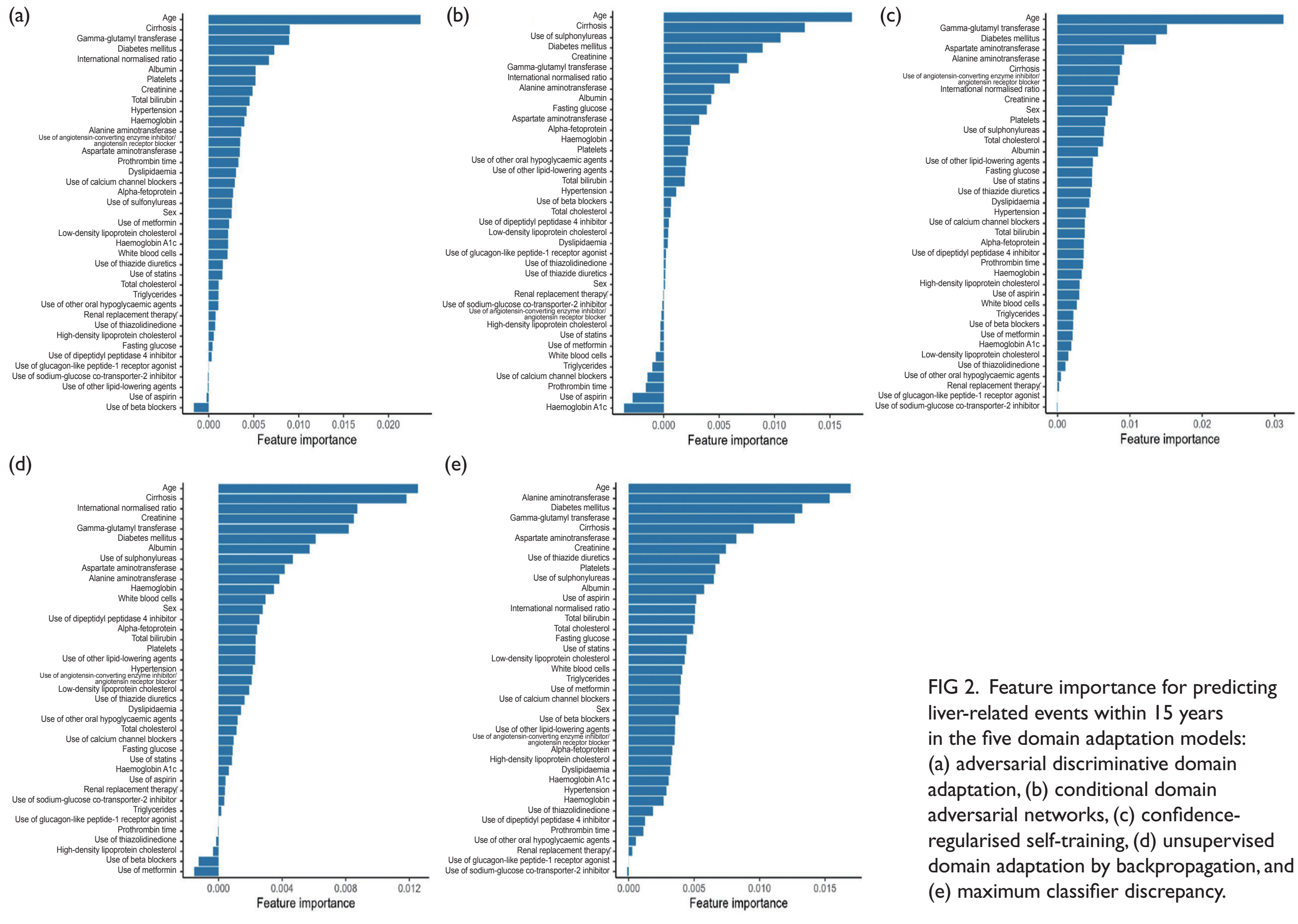

FIG 2. Feature importance for predicting liver-related events within 15 years in the five domain adaptation models: (a) adversarial discriminative domain adaptation, (b) conditional domain adversarial networks, (c) confidence-regularised self-training, (d) unsupervised domain adaptation by backpropagation, and (e) maximum classifier discrepancy.

Disclosure

The results of this research have been previously published in:

1. Zhang X, Yip TC, Tse YK, et al. Trends in risk factor control and treatment among patients with non-alcoholic fatty liver disease and type 2 diabetes between 2000 and 2020: a territory-wide study. Aliment Pharmacol Ther 2023;57:1103-16.
2. Yip TC, Hui WY, Ho WW, et al. Laboratory parameter-based model for predicting hepatic complications in non-alcoholic fatty liver disease: a territory-wide cohort study. Gut 2023;72(Suppl 1):A146-A148.
3. Yip TC, Yam TF, Lai MS, et al. A laboratory parameter-based risk score for predicting liver-related complications in patients with non-alcoholic fatty liver disease. Hepatol Int 2023;17(Supp 1):S47-48.
4. Lin H, Yip TC, Zhang X, et al. Age and the relative importance of liver-related deaths in nonalcoholic fatty liver disease. Hepatology 2023;77:573-84.
5. Yip TC, Li G, Wong GL, Tse YK, Wong VW. Accuracy of administrative coding for nonalcoholic fatty liver disease and related hepatic complications: a single-center validation study. Hepatology 2021;74(S1):928A-929A.

Acknowledgements

We thank Mr Yee-Kit Tse and Ms Mandy Sze-Man Lai for formatting large datasets and performing advanced statistical analyses.

References

1. Sanyal AJ, Van Natta ML, Clark J, et al. Prospective study of outcomes in adults with nonalcoholic fatty liver disease. N Engl J Med 2021;385:1559-69.
2. Kanwal F, Shubrook JH, Adams LA, et al. Clinical care pathway for the risk stratification and management of patients with nonalcoholic fatty liver disease. Gastroenterology 2021;161:1657-69.
3. Zhang X, Yip TC, Wong GL, et al. Clinical care pathway to detect advanced liver disease in patients with type 2 diabetes through automated fibrosis score calculation and electronic reminder messages: a randomised controlled trial. Gut 2023;72:2364-71.
4. Yip TC, Li G, Wong GL, Tse YK, Wong VW. Accuracy of administrative coding for nonalcoholic fatty liver disease and related hepatic complications: a single-center validation study. Hepatology 2021;74(S1):928a-9a.
5. Calzadilla-Bertot L, Jeffrey GP, Wang Z, et al. Predicting liver-related events in NAFLD: a predictive model. Hepatology 2023;78:1240-51.