Machine learning in predicting hepatocellular carcinoma in patients with chronic viral hepatitis in Hong Kong: abridged secondary publication

GLH Wong *, HLY Chan, YK Tse, PC Yuen †, VWS Wong

KEY MESSAGES

1. Novel machine-learning models generate accurate risk scores for hepatocellular carcinoma (HCC) in patients with chronic viral hepatitis.
2. HCC ridge score is consistently more accurate than existing HCC risk scores.
3. These machine-learning models may be incorporated into electronic health systems to guide cancer surveillance strategies and reduce cancer death.

Hong Kong Med J 2023;29(Suppl 1):S14-7
HMRF project number: 07180216

[1] GLH Wong, [1] HLY Chan, [2] YK Tse, [2] PC Yuen, [1] VWS Wong

[1] *Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong SAR, China*
[2] *Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China*

* Principal applicant: wonglaihung@cuhk.edu.hk
† Corresponding author: pcyuen@comp.hkbu.edu.hk

Introduction

The Hong Kong Viral Hepatitis Action Plan 2020-2024 aims to reduce the burden of chronic viral hepatitis (CVH) through effective prevention, treatment, and control of viral hepatitis. A comprehensive review of the disease burden of CVH and accurate prediction of hepatocellular carcinoma (HCC) may help guide strategies and action plans and ultimately eliminate viral hepatitis. Most HCC risk prediction models are developed using regression analysis.[1,2] Machine learning is a comprehensive tool for model development.[3,4] It enables direct selection of predicting parameters without subjective preselection, maximising data use while minimising bias. This study aims to develop prediction models using machine-learning algorithms to define the risk levels of HCC in patients with CVH. These models can be incorporated into electronic health systems to facilitate clinical assessment and risk stratification of HCC in patients with CVH.

Methods

This territory-wide registry cohort study was conducted using data from the Hospital Authority Data Collaboration Laboratory, which provides anonymised and de-identified data from all public hospitals and clinics in Hong Kong, including demographics, inpatient admissions, transfers and discharges, outpatient appointments, diagnosis, procedures, medications, laboratory tests and results, radiology examinations, clinical notes and summaries, and radiology reports and radiology images.

Data of patients with CVH (chronic hepatitis B [CHB] and chronic hepatitis C [CHC]) between 1 January 2000 and 31 December 2018 were retrieved. CHB/CHC was defined by positive hepatitis B/C surface antigen for ≥6 months, and/or by the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis codes, and/or by use of antiviral treatment for CHB/CHC.

Baseline date was defined as the date of first diagnosis of CHB or CHC by viral markers, ICD-9-CM codes, or antiviral drugs, whichever came first. Liver biochemistries and haematological and virological parameters were collected. Antiviral treatment included oral nucleos(t)ide analogues for CHB as well as (pegylated)-interferon with or without ribavirin and direct-acting antivirals for CHC. Medication use was defined as those prescribed for ≥4 weeks. The severity of liver fibrosis was assessed with serum formulae, namely the aspartate transaminase to platelet ratio index, Fibrosis-4 index, and Forns index. Advanced liver fibrosis was defined as the aspartate transaminase to platelet ratio index of ≥2, Fibrosis-4 index of ≥3.25, or Forns index of ≥8.4.

Patients with HCC were identified by diagnosis codes (155.0 for hepatocellular carcinoma and 155.2 for carcinoma of liver) or procedure codes for HCC treatment. The use of single ICD-9-CM codes for diagnosis was 99% accurate when referenced to clinical, laboratory, imaging, and endoscopy results from electronic medical records.

Data were analysed using SPSS (Windows version 25; IBM Corp, Armonk [NY], US), SAS (9.4; SAS Institute, Cary [NC], US), and R software (3.5.1; R Foundation for Statistical Computing, Vienna, Austria). The cohort was randomly split into training and validation cohorts in a 7:3 ratio. Additional external validation was performed in an independent cohort of Korean patients. Five popular machine-

learning models (logistic regression, ridge regression, AdaBoost, decision tree, and random forest) were compared to determine the best prediction model. Accuracy of the models was assessed by the area under the receiver operating characteristic curve (AUROC). Dual cutoffs were selected to achieve 90% sensitivity and 90% specificity to rule out and rule in patients with HCC, while maximising the corresponding specificity and sensitivity, respectively. The model with the highest AUROC in the validation cohort was considered the most predictive model, which was compared with other HCC risk scores: CU-HCC (Chinese University-HCC), GAG-HCC (guide with age, gender, hepatitis B virus DNA, central promoter mutations and cirrhosis-HCC), REACH-B (risk estimate for hepatocellular carcinoma in chronic hepatitis B), PAGE-B (platelets, age, gender, and hepatitis B virus), and REAL-B (real-world effectiveness from the Asia Pacific rim liver consortium for hepatitis B virus). All tests were two-sided. A P value of <0.05 was considered statistically significant. P values for pairwise comparison were adjusted by Bonferroni correction.

Results

Of 266 017 patients with viral hepatitis identified, 117 640 were excluded and 148 377 with CVH (CHB=126 890, CHC=16 811, and both=4676) were included in analysis. The cohorts were predominantly male, and most patients had compensated liver disease. The prevalence of comorbidities generally increased over time.

51 572 (>40%) patients with CHB had received antiviral treatment by 2018. The cumulative treatment uptake increased from 12.05% during 2005-2009 to 17.76% during 2010-2013 to 40.64% during 2014-2018. Of them, 99.3% received nucleos(t)ide analogues and 1.9% received conventional or pegylated interferon. 5660 (>30%) patients with CHC had received antiviral treatment by 2018. Of them, 92.2% received conventional or pegylated interferon and ribavirin and 7.8% received direct-acting antivirals, which became available in Hong Kong in late 2013.

A total of 124 006 patients were included in developing machine-learning models to predict HCC, with inclusion of all 46 parameters at baseline, in which 36 or 20 selected parameters had best predictive power (Table 1). In the training cohort (n=86 804, HCC=6821), random forest, decision tree, and ridge regression performed the best with inclusion of all 46 parameters (AUROC=0.992, 0.800, and 0.842, respectively), 36 selected parameters (AUROC=0.991, 0.884, and 0.839, respectively), and 20 selected parameters (AUROC=0.987, 0.877, and 0.817, respectively). In the validation cohort (n=37 202, HCC=2875), ridge regression had

TABLE 1. Parameters used in machine-learning models

Parameter	All 46 parameters	36 selected parameters	20 selected parameters
Male sex	✓	✓	✓
Age	✓	✓	✓
Platelet	✓	✓	✓
Albumin	✓	✓	✓
Total bilirubin	✓	✓	✓
Alanine aminotransferase	✓	✓	✓
Aspartate aminotransferase	✓	-	-
Alpha-fetoprotein	✓	-	-
International normalised ratio	✓	-	-
Creatinine	✓	-	-
Gamma glutamyl transferase	✓	-	-
Total cholesterol	✓	-	-
Glycated haemoglobin	✓	-	-
Fasting glucose	✓	-	-
Hepatitis B virus DNA	✓	-	-
Positive hepatitis B e-antigen	✓	-	-
Cirrhosis	✓	✓	✓
Cardiovascular disease	✓	✓	-
Colorectal cancer	✓	✓	-
Lung cancers	✓	✓	-
Urinary/renal malignancies	✓	✓	-
Cervical cancer	✓	✓	-
Breast cancer	✓	✓	-
Lymphoma	✓	✓	-
Chronic kidney disease	✓	✓	✓
Osteopenia	✓	✓	-
Osteoporosis	✓	✓	-
Diabetes mellitus	✓	✓	✓
Hypertension	✓	✓	✓
Anticoagulants	✓	✓	-
Angiotensin-converting-enzyme inhibitor / angiotensin receptor blocker	✓	✓	✓
Antiplatelet agents	✓	✓	✓
Beta blockers	✓	✓	✓
Histamine-2 receptor antagonist	✓	✓	-
Insulin	✓	✓	✓
Immunosuppressant	✓	✓	-
Loop diuretics	✓	✓	-
Metformin	✓	✓	✓
Nonsteroidal anti-inflammatory drug	✓	✓	-
Other lipid lowering agents	✓	✓	✓
Other oral hypoglycaemic agents	✓	✓	✓
Proton pump inhibitor	✓	✓	✓
Potassium sparing diuretics	✓	✓	-
Statins	✓	✓	✓
Sulphonylurea	✓	✓	✓
Thiazides	✓	✓	-

consistently higher accuracy with inclusion of all 46 parameters (AUROC=0.844), 36 selected parameters (AUROC=0.840), and 20 selected parameters (AUROC=0.821) [Table 2]. Sensitivity, specificity, and positive and negative predictive values of these five machine-learning models in training and validation cohorts are shown in Table 3. Dual cut-offs approach was applicable in >60% of patients in most models; the applicability was particularly high with random forest in the training cohort (96.6%) but not in the validation cohort (59.5%).

As the ridge regression model achieved consistently good performance in training and validation cohorts with all or selected parameters, the HCC ridge score was developed for comparison with other HCC risk scores. The low cut-off was set at <0.1 to achieve high sensitivity (≥90%) and between 0.1 and 0.2 to achieve high specificity (≥90%). The AUROC of the CU-HCC score, GAG-HCC score, REACH-B score, PAGE-B score, and REAL-B score was 0.672, 0.745, 0.671, 0.748, and 0.712, respectively (Table 4). Using dual cut-offs, the low cut-off of the REAL-B score (<4) had highest sensitivity (96.0%) but was applicable only to 17.6% of patients. The

TABLE 2. Area under the receiver operating characteristic curve (AUROC) of five machine-learning models in predicting hepatocellular carcinoma (HCC)

Machine-learning model	Training cohort (n=86 804, HCC=6821)			Validation cohort (n=37 202, HCC=2875)		
	20 selected parameters	36 selected parameters	All 46 parameters	20 selected parameters	36 selected parameters	All 46 parameters
	AUROC					
Logistic regression	0.814±0.006	0.829±0.006	0.825±0.006	0.818±0.009	0.832±0.009	0.829±0.009
Ridge regression	0.817±0.005	0.839±0.005	0.842±0.005	0.821±0.009	0.840±0.009	0.844±0.009
AdaBoost	0.822±0.006	0.828±0.006	0.828±0.006	0.824±0.009	0.833±0.009	0.832±0.009
Decision tree*	0.877±0.005	0.884±0.005	0.800±0.005	0.802±0.010	0.819±0.010	0.818±0.010
Random forest†	0.987±0.003	0.991±0.003	0.992±0.003	0.807±0.010	0.821±0.010	0.821±0.010

* AUROC higher than that of logistic regression and AdaBoost in both cohorts, P<0.05
† AUROC higher than that of decision tree in validation cohort, P<0.05

TABLE 3. Accuracy of five machine-learning models in predicting hepatocellular carcinoma

Machine-learning model	Dual Cut-offs	No. (%) of patients with <lower cut-off and ≥upper cut-off	Sensitivity	Specificity	Positive predictive value	Negative predictive value
			% (95% confidence interval)			
Training cohort (n=86 804)*						
Logistic regression	0.18	43 951 (50.6)	0.90 (0.89-0.91)	0.54 (0.542-0.545)	0.143 (0.141-0.146)	0.985 (0.983-0.985)
	0.29	11 527 (13.3)	0.52 (0.51-0.53)	0.90 (0.898-0.902)	0.307 (0.300-0.315)	0.956 (0.955-0.958)
Ridge regression	0.07	48 341 (55.7)	0.90 (0.89-0.91)	0.596 (0.593-0.599)	0.160 (0.156-0.164)	0.986 (0.985-0.987)
	0.15	11 506 (13.3)	0.52(0.51-0.53)	0.900 (0.898-0.902)	0.307 (0.300-0.315)	0.956 (0.955-0.958)
AdaBoost	0.42	43 298 (49.9)	0.91 (0.90-0.92)	0.533 (0.529-0.536)	0.142 (0.140-0.145)	0.985 (0.984-0.986)
	0.45	10 363 (11.9)	0.48 (0.46-0.49)	0.911 (0.909-0.913)	0.313 (0.308-0.320)	0.953 (0.952-0.954)
Decision tree	0.04	48 765 (56.2)	0.92 (0.92-0.93)	0.603 (0.599-0.606)	0.166 (0.163-0.170)	0.990 (0.989-0.991)
	0.17	12 029 (13.9)	0.63 (0.62-0.64)	0.903 (0.902-0.905)	0.356 (0.349-0.363)	0.966 (0.965-0.967)
Random forest	0.45	71 804 (82.7)	0.90 (0.90-0.91)	0.998 (0.997-0.998)	0.976 (0.973-0.979)	0.992 (0.991-0.992)
	0.10	12 074 (13.9)	0.97 (0.96-0.97)	0.932 (0.930-0.933)	0.547 (0.539-0.557)	0.997 (0.997-0.998)
Validation cohort (n=37 202)						
Logistic regression	0.18	19 448 (52.3)	0.90 (0.89-0.91)	0.568 (0.553-0.565)	0.146 (0.142-0.151)	0.985 (0.984-0.987)
	0.29	4930 (13.3)	0.52 (0.50-0.54)	0.900 (0.896-0.902)	0.304 (0.293-0.317)	0.957 (0.955-0.959)
Ridge regression	0.07	20 816 (56.0)	0.90 (0.89-0.91)	0.598 (0.593-0.603)	0.158 (0.152-0.164)	0.986 (0.985-0.988)
	0.15	4932 (13.3)	0.52 (0.50-0.54)	0.900 (0.897-0.903)	0.304 (0.291-0.317)	0.957 (0.955-0.960)
AdaBoost	0.42	18 725 (50.3)	0.91 (0.90-0.92)	0.538 (0.532-0.543)	0.142 (0.137-0.146)	0.987 (0.985-0.988)
	0.45	4377 (11.8)	0.47 (0.45-0.49)	0.912 (0.909-0.914)	0.310 (0.297-0.323)	0.954 (0.952-0.956)
Decision tree	0.02	17 689 (47.6)	0.90 (0.89-0.91)	0.507 (0.501-0.511)	0.133 (0.129-0.137)	0.983 (0.982-0.985)
	0.17	4987 (13.4)	0.54 (0.52-0.56)	0.900 (0.897-0.904)	0.312 (0.302-0.330)	0.959 (0.957-0.961)
Random forest	0.01	17 561 (47.2)	0.90 (0.89-0.91)	0.503 (0.496-0.508)	0.132 (0.127-0.137)	0.984 (0.982-0.986)
	0.20	4561 (12.3)	0.52 (0.50-0.53)	0.910 (0.907-0.913)	0.326 (0.312-0.341)	0.957 (0.955-0.959)

* Dual cut-offs are selected to achieve >90% sensitivity and specificity

TABLE 4. Comparison of the hepatocellular carcinoma (HCC) ridge score and other HCC risk scores in predicting HCC in the validation cohort (n=37 202)

Risk score	Area under the receiver operating characteristic curve	Dual Cut-offs	No. (%) of patients with <lower cut-off and ≥upper cut-off	Sensitivity	Specificity	Positive predictive value	Negative predictive value
				% (95% confidence interval)			
HCC ridge score	0.840	0.07	20 816 (56.0)	90.0 (89.0-0.91)	59.8 (59.3-60.3)	15.8 (15.2-16.4)	98.6 (98.5-98.8)
		0.15	4932 (13.3)	52.2 (50.3-54.0)	90.0 (89.7-90.3)	30.4 (29.1-31.7)	95.7 (95.5-96.0)
CU-HCC score	0.672	<5	27 083 (72.8)	46.4 (28.6-64.3)	74.0 (69.9-78.4)	10.3 (6.4-14.3)	95.6 (94.2-97.1)
		≥20	7812 (21.0)	32.1 (14.3-50.0)	79.7 (75.9-83.6)	9.1 (4.5-14.0)	94.8 (93.6- 96.2)
GAG-HCC score	0.745	<80	25 781 (69.3)	64.3 (46.4-82.1)	71.5 (67.2-75.6)	12.3 (8.8-15.9)	97.0 (95.5-98.4)
		≥101	2939 (7.9)	28.6 (14.3-46.4)	93.4 (91.1-95.6)	21.1 (10.5-33.3)	95.5 (94.5-96.6)
REACH-B score	0.671	<8	18 601 (50.0)	72.7 (54.6-90.9)	52.8 (45.4-59.7)	16.2 (12.1-20.0)	94.1 (89.8-97.9)
		≥14	558 (1.5)	4.5 (0-13.5)	98.9 (97.4-100)	33.3 (0-100)	89.2 (88.7-90.2)
PAGE-B score	0.748	<10	10 193 (27.4)	95.7 (94.9-96.5)	29.4 (28.9-30.0)	10.7 (10.6-10.9)	98.7 (98.5-99.0)
		≥13	17 969 (48.3)	81.1 (79.4-82.7)	54.6 (54.0-55.3)	13.7 (13.4-14.0)	97.0 (96.8-97.3)
REAL-B score	0.712	<4	6548 (17.6)	96.0 (95.2-96.9)	19.2 (18.5-19.8)	12.0 (11.9-12.2)	97.7 (97.2-98.2)

Abbreviations: CU-HCC = Chinese University-HCC; GAG-HCC = guide with age, gender, hepatitis B virus DNA, central promoter mutations and cirrhosis-HCC; PAGE-B = platelets, age, gender, and hepatitis B virus; REACH-B = risk estimate for hepatocellular carcinoma in chronic hepatitis B; REAL-B = real-world effectiveness from the Asia Pacific rim liver consortium for hepatitis B virus

high cut-off of the REACH-B score (≥14) had highest specificity (98.9%) but was applicable only to 1.5% of patients. The HCC ridge score achieved larger AUROC (0.840) and higher applicability, with 30.7% of patients falling into the grey zone.

Discussion

Machine-learning models of ridge regression and random forest are accurate to predict HCC in patients with CVH. These models may be used as built-in functional keys or calculators in electronic health systems to facilitate hepatitis elimination. Electronic health records provide robust and comprehensive demographic and laboratory data of patients. Nonetheless, clinical observations and anthropometric measurements may be missing, especially in regions where manual data entry is not available.

Machine-learning models can be applied in managing patients with CVH, which affects >300 million people worldwide. Ridge regression is a technique for analysing multiple regression data that have multicollinearity problems. When multicollinearity occurs, least squares estimates are unbiased, but their variances are large and may deviate far from the true value. Hence, ridge regression is particularly suitable for machine-learning models in clinical medicine, as many parameters included in the models are closely related and with multicollinearity.

Conclusion

The HCC ridge score accurately predicts HCC in CVH patients. Machine-learning models may be developed as built-in functional keys or calculators in electronic health systems to reduce cancer mortality. Studies comparing machine-learning-model-guided HCC surveillance with routine clinical practice for early diagnosis of HCC in CVH patients are warranted.

Funding

This study was supported by the Health and Medical Research Fund, Health Bureau, Hong Kong SAR Government (#07180216). The full report is available from the Health and Medical Research Fund website (https://rfs1.fhb.gov.hk/index.html).

Disclosure

The results of this research have been previously published in:

1. 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 2022;4:100441.

References

1. Yip TC, Hui VW, Tse YK, Wong GL. Statistical strategies for HCC risk prediction models in patients with chronic hepatitis B. Hepatoma Res 2021;7:7.
2. Wong GL, Wong VW, Choi PC, Chan AW, Chan HL. Development of a non-invasive algorithm with transient elastography (Fibroscan) and serum test formula for advanced liver fibrosis in chronic hepatitis B. Aliment Pharmacol Ther 2010;31:1095-103.
3. Wong GL, Yuen PC, Ma AJ, Chan AW, Leung HH, Wong VW. Artificial intelligence in prediction of non-alcoholic fatty liver disease and fibrosis. J Gastroenterol Hepatol 2021;36:543-50.
4. Wong GL, Wong VW, Yuen BW, et al. An aging population of chronic hepatitis B with increasing comorbidities: a territory-wide study from 2000 to 2017. Hepatology 2020;71:444-55.