Research article JHEP|Reports

Deep learning model for prediction of hepatocellular carcinoma in patients with HBV-related cirrhosis on antiviral therapy

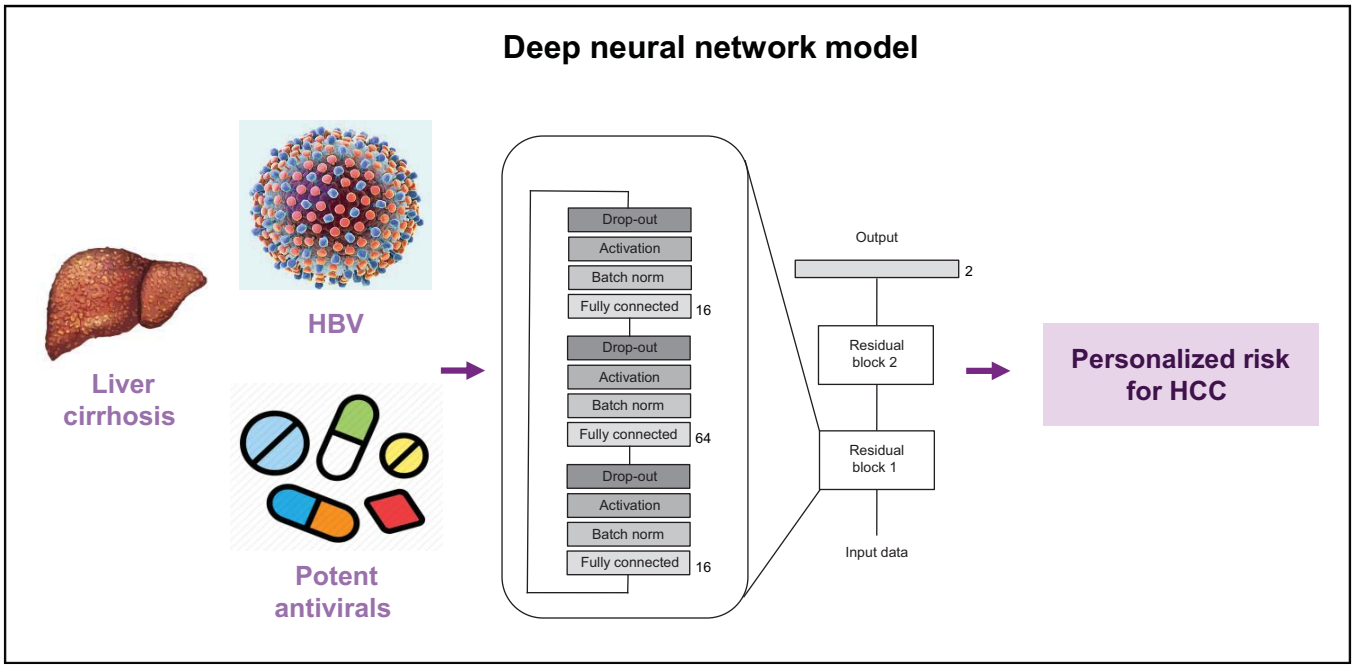
Authors

Joon Yeul Nam, Dong Hyun Sinn, Junho Bae, Eun Sun Jang, Jin-Wook Kim, Sook-Hyang Jeong

Correspondence

jsh@snubh.org (S.-H. Jeong).

Graphical abstract



Highlights

- Personalised risk stratification of HCC among at-risk patients is important for surveillance.
- Deep-learning-based HCC prediction model performed better than the previous models, which were based on conventional statistics.
- Deep learning method can provide continuous probability results, not a binary result.

Lay summary

For early detection of hepatocellular carcinoma, it is important to maintain regular surveillance. However, there is currently no standard prediction model for risk stratification that can be used to establish a personalised surveillance strategy. We develop and validate a deep-learning-based model that showed better performance than previous models.

https://doi.org/10.1016/j.jhepr.2020.100175

Research article

JHEP|Reports

Deep learning model for prediction of hepatocellular carcinoma in patients with HBV-related cirrhosis on antiviral therapy

Joon Yeul Nam,[1,†] Dong Hyun Sinn,[2,†] Junho Bae,[3] Eun Sun Jang,[1] Jin-Wook Kim,[1] Sook-Hyang Jeong[1,*]

[1]*Department of Internal Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea;* [2]*Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea;* [3]*DEEPNOID Inc., Seoul, Republic of Korea*

JHEP Reports **2020.** https://doi.org/10.1016/j.jhepr.2020.100175

Background & Aims: Personalised risk prediction of the development of hepatocellular carcinoma (HCC) among patients with liver cirrhosis on potent antiviral therapy is important for targeted screening and individualised intervention. This study aimed to develop and validate a new model for risk prediction of HCC development based on deep learning, and to compare it with previously reported risk models.

Methods: A novel deep-learning-based model was developed from a cohort of 424 patients with HBV-related cirrhosis on entecavir therapy with 2 residual blocks, including 7 layers of a neural network, and it was validated using an independent external cohort (n = 316). The deep-learning-based model was compared to 6 previously reported models (platelet, age, and gender-hepatitis B score [PAGE-B], Chinese University HCC score [CU-HCC], HCC-Risk Estimating Score in CHB patients Under Entecavir [HCC-RESCUE], age, diabetes, race, etiology of cirrhosis, sex, and severity HCC score [ADRESS-HCC], modified PAGE-B score [mPAGE], and Toronto HCC risk index [THRI]) using Harrell's concordance (*c*)-index.

Results: During a median 5.2 yr of follow-up (inter-quartile range 2.8–6.9 yr), 86 patients (20.3%) developed HCC. The deep-learning-based model had a Harrell's *c*-index of 0.719 in the derivation cohort and 0.782 in the validation cohort. Goodness of fit was confirmed by the Hosmer-Lemeshow test (p >0.05). Moreover, this model in the validation cohort had the highest *c*-index among the 6 previously reported models: PAGE-B (0.570), CU-HCC (0.548), HCC-RESCUE (0.577), ADRESS-HCC (0.551), mPAGE (0.598), and THRI (0.587) (all p <0.001). The misclassification rate of this model was 23.7% (model accuracy: 76.3%) in the validation group.

Conclusions: The deep-learning-based model had better performance than the previous models for predicting the HCC risk in patients with HBV-related cirrhosis on potent antivirals.

Introduction

Hepatocellular carcinoma (HCC) is the fourth most common cause of death from cancer worldwide.[1] Liver cirrhosis is the leading cause of HCC, and HBV infection is the first major cause of cirrhosis worldwide.[2] Current international guidelines recommend surveillance for HCC in at-risk patients to detect HCC at an earlier stage, and thereby to improve their outcomes.[3,4] However, even in patients with cirrhosis, the risk of HCC development is variable, and it is necessary to define the strategy of HCC surveillance according to risk stratifications.

Therefore, several HCC prediction models were developed for patients with chronic liver disease, including platelet, age, and gender-hepatitis B score (PAGE-B),[5] Chinese University HCC score (CU-HCC),[6] HCC-Risk Estimating Score in CHB patients Under Entecavir (HCC-RESCUE),[7] age, diabetes, race, etiology of cirrhosis, sex, and severity HCC score (ADRESS-HCC),[8] modified PAGE-B score (mPAGE-B),[9] and Toronto HCC risk index (THRI).[10] Each model includes 3 to 6 factors showing a significant association with HCC development, which are easy to get at the initial hospital visit[5–10]: age, sex, platelet count, serum albumin, serum bilirubin, HBV DNA titre, presence of cirrhosis, presence of diabetes, and race. When applying these factors in the models, a certain cut-off value for each factor is determined by conventional statistical method that results in the factor being binary data instead of a continuous value, which originally it is. Although these models include a small number of factors and a binary data type is easy to use, none of these models are widely accepted yet because they have a rather poor performance in real-world practice.

To increase the model performance, it is necessary to include all possible related factors with continuous values as intended. Recently, advancements in the deep learning methodology have provided a more effective and accurate way to manage large-scale data compared with conventional statistical methods.[11] The aims of this study were to develop and validate a

Keywords: Hepatitis B virus; Hepatocellular carcinoma; Cirrhosis; Prediction model; Convolutional neural network.

Received 9 April 2020; received in revised form 14 August 2020; accepted 18 August 2020Available online 30 August 2020

† Joon Yeul Nam and Dong Hyun Sinn contributed equally to this work as first authors. Current affiliation of Joon Yeul Nam is Department of Internal Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea.

* Corresponding author. Address: Department of Internal Medicine, Seoul National University Bundang Hospital, Gumi-dong 300, Bundang-gu, Seongnam-si, Gyeonggi-do, Republic of Korea. Tel.: +82-31-787-7029; Fax: +82-31-787-4052.

E-mail address: jsh@snubh.org (S.-H. Jeong).

deep-learning-based prediction model of HCC and to compare its performance with previously reported models in patients with HBV-related cirrhosis on antiviral therapy.

Patients and methods

Study design, setting, and participants

The study was conducted at 2 tertiary hospitals (Seoul National University Bundang Hospital [SNUBH], Bundang, Gyeonggi-do, Republic of Korea and Samsung Medical Center [SMC], Seoul, Republic of Korea), using electronic medical records. We included consecutive patients meeting all of the following inclusion criteria:

- chronic HBV infection was defined by the presence of the HBsAg for more than 6 months or relevant clinical history;
- cirrhosis was defined as
  - a platelet count of <100,000/ml and image diagnosis (either cirrhosis ultrasonography [US] or computed tomography) that included a blunted liver edge with splenomegaly (>12 cm);
  - the presence of portal hypertension, such as oesophageal or gastric varices; or
  - features of decompensation, such as ascites or hepatic encephalopathy[12]; and
- treatment-naive patients who started entecavir therapy between March 2007 and June 2013.

Among them, we excluded patients if they met any of the following criteria:

- age <18 yr;
- co-infection with other hepatitis viruses (*i.e.* HCV or HDV) or HIV; and
- previous and current malignancies, to identify adult patients with HBV mono-infected cirrhosis without malignancy at baseline who started entecavir therapy.

There were 430 eligible patients at the SNUBH cohort and 324 eligible patients at the SMC cohort. Among them, those who developed HCC within 6 months from cohort entry were excluded (SNUBH cohort [n = 6] and SMC cohort [n = 8]). The SNUBH cohort was used to evaluate the 6 previous models and to develop the deep-learning-based model. The SMC cohort was used to validate the novel model and to compare it with the 6 previous models. This study complied with the Declaration of Helsinki. The study protocol was approved by the Institutional Review Board of SNUBH and SMC, and the requirement for informed consent from patients was waived.

Variables, data sources, and measurement

The index visit was defined as the first day of entecavir prescription. Patients were monitored for a 1- to 6-month interval on a regular basis by their attending physicians, who were all experienced hepatologists. HCC surveillance was done using US and serum alpha-fetoprotein at a 6-month interval. In some cases, computed tomography was used instead of US. HCC was diagnosed clinically or histologically based on regional HCC diagnosis guidelines during the study periods.[13,14] The follow-up period was defined from the day of the initiation of the entecavir treatment to HCC diagnosis or last hospital visit, whichever comes first. In case of death or liver transplantation before HCC development, data were censored at the time of death or transplantation. The reference date was June 30, 2017.

The following variables were collected by reviewing the electronic medical records of each patient: age, sex, height, weight, presence of diabetes, platelet count, serum bilirubin, serum albumin, and HBV DNA levels at the index visit. Cirrhosis was defined as the aforementioned criteria, and the presence of diabetes was determined by medical record review of the glucose level and by prescription history of anti-diabetic medications.

Development of the deep-learning-based novel prediction model

Using deep neural network technology, most of the reported factors significantly associated with the development of HCC from previous studies were selected: age, sex, platelet count, serum albumin, serum bilirubin, HBV DNA titre, presence of diabetes, and observation period from the index visit. Three factors among all the factors (presence of cirrhosis, aetiology, and race) from the 6 previous models were excluded in the deep-learning-based model because all of our derivation cohorts were Asian patients with HBV-related cirrhosis.

Among the selected factors, sex and the presence of diabetes were binary data types, and the others were continuous numerical data types. The numerical data were applied to the new model as a continuous value as is rather than categorising them with certain cut-off values to maintain and maximise the influence of each numerical factor. Because the related factors consisted of 2 data types, we used a deep neural network and arranged that each input variable was allocated to its own input node, not to be affected by the other data types. Numerical variables were normalised before entering them into the neural network, which could then be trained faster and reduce the chances of getting stuck in local optima.

To improve the model predictability, we used the residual learning framework of the ResNet architecture (Microsoft Research, Redmond, WA, USA), which has shown good performance in previous image recognition.[15] We adopted residual learning for every stacked layer in this model. Residual learning is established through the connection of stacked layers. By integrating the following input and output variables of each layer, it provides additional non-linearity and reduces the additionally generated weight to increase the learning performance. We applied shortcuts (or skip connection) to improve the learning performance by minimising the data loss for the centring layer responses, gradients, and propagated errors, implemented by shortcut connections.

The deep neural network was developed based on TensorFlow (version 1.13; Google, Mountain View, CA, USA), and the Adam optimization algorithm was used to optimise the model. We implemented the weighted cross entropy as the loss function to control the class imbalance in the derivation cohort. The parametric rectified linear unit was used as an activation function, and we adopted batch normalisation with a minimal dropout. In addition, the neural network was not deeper, considering the small amount of data, and the learning rate was 1×10^{-5}.

Statistical analyses

Baseline characteristics were presented as the mean ± SD for normally distributed continuous variables and as the median with inter-quartile ranges for continuous variables with a skewed distribution. Discrete variables were summarised by the

Table 1. Baseline characteristics of patients of HBV-related cirrhosis on antiviral therapy.

Patient characteristics	Derivation set (n = 424)	Validation set (n = 316)	*p* value
Age, mean ± SD (yr)	52.7 ± 10.1	51.9 ± 9.2	0.084*
Sex, male, n (%)	270 (63.7)	204 (64.6)	0.806†
BMI	24.2 ± 3.1	24.9 ± 3.1	0.003*
Platelet ($\times 10^9$/L)	124.6 ± 50.7	115.2 ± 54.2	0.016*
Albumin (g/dl)	3.9 ± 0.6	3.8 ± 0.6	0.106*
Total bilirubin (mg/dl)	1.5 ± 2.0	1.3 ± 1.4	0.265*
HBV DNA ($\log_{10}$ IU/ml)	6.7 ± 1.3	6.0 ± 1.3	<0.001*
FIB-4	3.4 (2.2–5.8)‡	3.8 (2.4–6.8)	0.082*
DM, n (%)	75 (17.7)	56 (17.7)	0.991†

Data are expressed as n (%) or mean ± SD.
BMI, body mass index; DM, diabetes mellitus; FIB-4, fibrosis-4.
* By Student's *t* test.
† By Pearson's chi-square test.
‡ Inter-quartile range.

number of subjects with percentages. To compare the baseline characteristics between the cohorts, we used Student's *t* test properly. Distribution of categorical variables was compared using the chi-square test. Survival analysis was performed using the Kaplan-Meier analysis, which was used to estimate the cumulative incidence rate of the HCC rate, according to the risk groups. To compare the prognostication power between the models, we conducted both discrimination and calibration performance. It was evaluated with the concordance (*c*)-index for the discrimination function and the Hosmer-Lemeshow test for the calibration function. The performances of the 6 previous models, which were CU-HCC, HCC-RESCUE, ADRESS-HCC, PAGE-B, mPAGE-B, and THRI, were evaluated and compared using the SNUBH cohort. Six prediction models were selected according to the characteristic of each cohort, which should include patients with potent antiviral treatment and some portion of patients with cirrhosis. The prognostic factors that were included in each model were as follows: CU-HCC[6] (age, albumin, bilirubin, HBV DNA titre, and presence of cirrhosis), HCC-RESCUE[7] (age, sex, and presence of cirrhosis), ADRESS-HCC[8] (age, diabetes, race, aetiology of cirrhosis, sex, and hepatic function severity), PAGE-B[5] (age, sex, and platelet), mPAGE-B[9] (age, sex, platelet, and albumin), and THRI[10] (age, sex, aetiology, and platelet). In the original studies, all the previous models established 2 cut-off values, which divide the subjects into 3 categories according to the risk stratification (low, intermediate, and high risk). Because enrolled patients in this study cohort had cirrhosis and none of the patients were in the low-risk group of each model, we evaluated and compared the model performance with a higher cut-off value for each model. The statistical analyses were performed using SPSS 22.0 (SPSS Inc., Chicago, IL, USA) and the SAS version 9.4 software (SAS Institute, Cary, NC, USA).

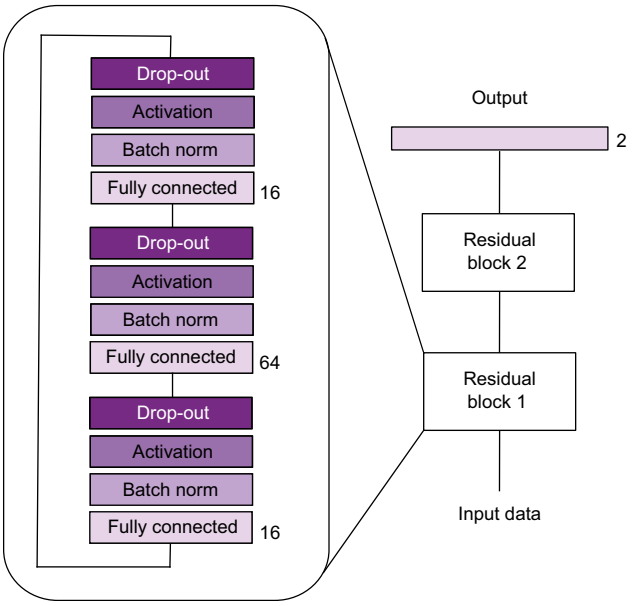


Fig. 1. Network architecture. The optimal model was established and had 2 residual blocks, including 7 layers of a neural network.

Table 2. Comparison of HCC development among the predictive models.

	Derivation set of this study			
	5-yr HCC incidence		3-yr HCC incidence	
	Value	95% CI	Value	95% CI
PAGE-B				
<18	0.126	0.080–0.169	0.085	0.048–0.121
≥18	0.236	0.161–0.304	0.142	0.083–0.196
CU-HCC				
<19	0.160	0.042–0.263	0.062	0.001–0.127
≥19	0.172	0.129–0.214	0.115	0.080–0.150
HCC-RESCUE				
<85	0.112	0.062–0.160	0.072	0.032–0.110
≥85	0.217	0.156–0.273	0.137	0.089–0.183
ADRESS-HCC				
<4.71	0.084	0.001–0.160	0.039	0.001–0.090
≥4.71	0.191	0.148–0.241	0.127	0.088–0.164
mPAGE-B				
<13	0.112	0.064–0.158	0.063	0.028–0.093
≥13	0.233	0.167–0.294	0.157	0.102–0.209
THRI				
<240	0.085	0.035–0.131	0.051	0.013–0.087
≥240	0.218	0.161–0.270	0.140	0.094–0.183

ADRESS-HCC, age, diabetes, race, etiology of cirrhosis, sex, and severity HCC score; CU-HCC, Chinese University HCC score; HCC-RESCUE, HCC-Risk Estimating Score in CHB patients Under Entecavir; mPAGE-B, modified platelet, age, and gender-hepatitis B score; PAGE-B, platelet, age, and gender-hepatitis B score; THRI, Toronto HCC risk index.

Results

Development of the optimal deep-learning-based model

The baseline characteristics of the derivation cohort are summarised in Table 1. During a median 5.2 yr of follow-up (interquartile range 2.8–6.9 yr), 86 patients (20.3%) developed HCC. The baseline characteristics of the validation cohort are also shown in Table 1. During a median 6.4 yr of follow-up (interquartile range 2.4–9.0 yr), 68 patients (21.5%) developed HCC. HCC was mainly diagnosed according to radiological criteria of qualified dynamic computed tomography or magnetic resonance imaging. Histological confirmation was performed in the absence of typical image finding of HCC. When compared with the derivation cohort, the validation cohort had a higher body mass index, lower platelet count, and lower HBV DNA levels. The optimal model was established and had 2 residual blocks, including 7 layers of a neural network, as the data size of the derivation cohort was relatively small. The deep-learning-based model did not present a binary result according to a certain cut-off. Instead, it presented the results (the probability of HCC development) as a continuous probability from 0 to 1 (Fig. 1). The *c*-index of the deep-learning-based model was 0.719 (95% CI

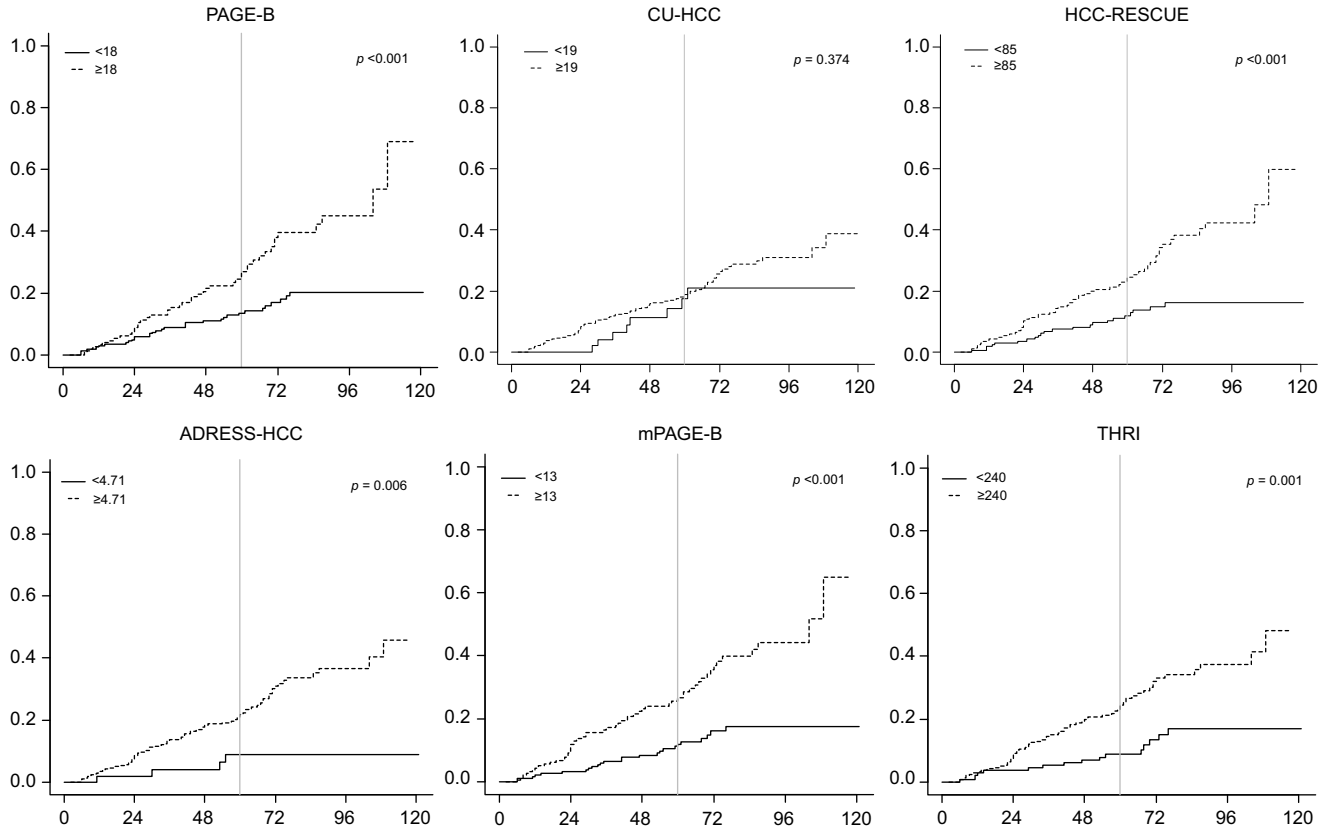


Fig. 2. Comparison of the previous models in the derivation cohort. The HCC incidence was significantly different between the high-risk and low-risk groups in 5 previous models (HCC-RESCUE, ADRESS-HCC, PAGE-B, mPAGE-B, and THRI) except for CU-HCC (by Kaplan-Meier analysis). ADRESS-HCC, age, diabetes, race, etiology of cirrhosis, sex, and severity HCC score; CU-HCC, Chinese University HCC score; HCC, hepatocellular carcinoma; HCC-RESCUE, HCC-Risk Estimating Score in CHB patients Under Entecavir; mPAGE-B, modified platelet, age, and gender-hepatitis B score; PAGE-B, platelet, age, and gender-hepatitis B score; THRI, Toronto HCC risk index.

Table 3. Comparison of previous HCC prediction models with DNN model with validation cohort.

Model	*c*-Index	95% CI		*p* value*
		Lower	Upper	
DNN	0.782	0.734	0.830	–
PAGE-B	0.570	0.514	0.626	<0.001
CU-HCC	0.548	0.491	0.604	<0.001
HCC-RESCUE	0.577	0.520	0.632	<0.001
ADRESS-HCC	0.551	0.495	0.607	<0.001
mPAGE-B	0.598	0.542	0.653	<0.001
THRI	0.587	0.530	0.641	<0.001

ADRESS-HCC, age, diabetes, race, etiology of cirrhosis, sex, and severity HCC score; *c*-index, concordance index; CU-HCC, Chinese University HCC score; DNN, deep neural network; HCC-RESCUE, HCC-Risk Estimating Score in CHB patients Under Entecavir; mPAGE-B, modified platelet, age, and gender-hepatitis B score; NPV, negative predictive value; PAGE-B, platelet, age, and gender-hepatitis B score; PPV, positive predictive value; THRI, Toronto HCC risk index.
* Compare with the *c*-index of the DNN model.

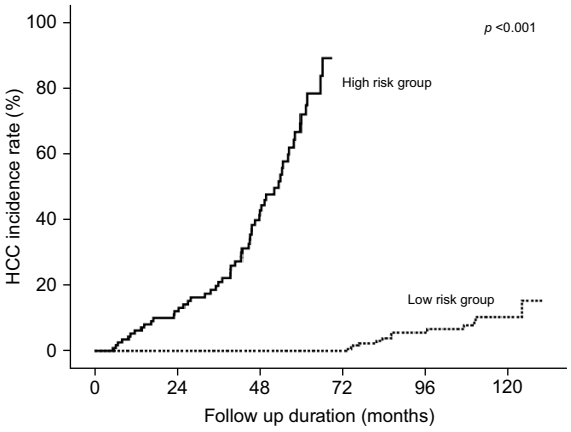
0.680–0.758) in the derivation cohort. Goodness of fit was confirmed by the Hosmer-Lemeshow test (p >0.05).

Performance evaluation of a deep-learning-based model and comparison of 6 previous models with the derivation cohort

The 5- and 3-yr HCC incidences of the 6 previous models (CU-HCC, HCC-RESCUE, ADRESS-HCC, PAGE-B, mPAGE-B, and THRI) were calculated and compared with our derivation cohort (Table 2). When we applied the derivation cohort to the 6 previous models, the discrimination of the 5-yr HCC cumulative incidence between the high- and low-risk groups was maintained along with the results of the original articles. However, the difference between the HCC cumulative incidences between the risk groups was reduced compared with the original data, which were reported in each article. In the survival analysis, the HCC incidence was significantly higher in the high-risk group than in the low-risk group among 5 previous models (HCC-RESCUE, ADRESS-HCC, PAGE-B, mPAGE-B, and THRI) except CU-HCC (Fig. 2).

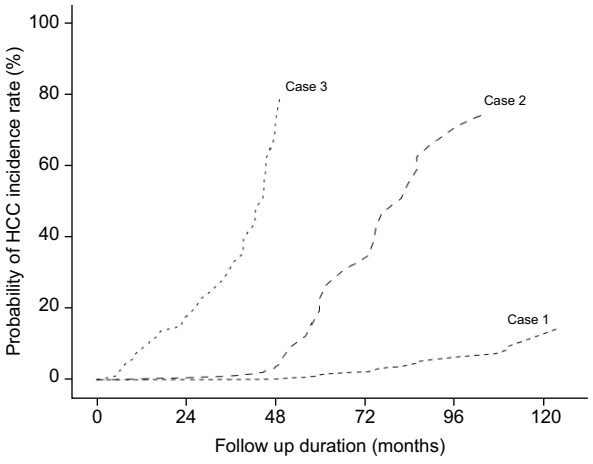
Performance evaluation of the deep-learning-based model and comparison of the 6 previous models in the validation cohort

The *c*-index of the deep-learning-based model was 0.782 (95% CI 0.734–0.830) in the validation cohort. The performance of the new model was compared with the previous models. The *c*-indexes of the previous models were below 0.6 with the validation cohort (PAGE-B [*c*-index 0.570; 95% CI 0.514–0.626], CU-HCC [*c*-index 0.548; 95% CI 0.491–0.604], HCC-RESCUE [*c*-index 0.577; 95% CI 0.520–0.632], ADRESS-HCC [*c*-index 0.551; 95% CI 0.495–0.607], mPAGE [*c*-index 0.598; 95% CI 0.542–0.653], and THRI [*c*-index 0.587; 95% CI 0.530–0.640]). However, the deep-learning-based model showed a significantly better performance than that of the 6 previous models (all p <0.001). The *c*-index of the deep-learning-based model was 0.782 (95% CI 0.734–0.830) in the validation cohort (Table 3). To evaluate the



N° at risk						
High risk	117	85	37			
Low risk	199	199	199	190	87	32

Fig. 3. Evaluation of the deep-learning-based model performance according to the risk groups in the validation cohort (cut-off value: 0.5). In the survival analysis between 2 groups, the high-risk group presented a significantly higher HCC incidence than the low-risk group in the validation cohort (p <0.001; by Kaplan-Meier analysis). HCC, hepatocellular carcinoma.



Case	Age	Sex	Diabetes	Platelet (x10^3/ml)	Albumin (g/dl)	Bilirubin (mg/dl)	HBV DNA (log_{10})
1	51	Female	0	187	3.9	1.2	6.0
2	54	Male	0	116	3.7	1.7	6.2
3	55	Male	1	107	3.0	2.2	5.8

Fig. 4. Expected HCC incidence rate of 3 hypothetical patients. The expected HCC incidences of 3 hypothetical patients were presented, according to baseline clinical and laboratory data (by prediction probabilities of deep neural network). HCC, hepatocellular carcinoma.

performance of the model, we categorised the risk group using a cut-off probability of 0.5 in the new model. In the survival analysis between the 2 groups, the high-risk group (≥0.5) had a significantly higher HCC incidence than that of the low-risk group (<0.5) in the validation cohort (p <0.001) (Fig. 3). The expected HCC incidences of 3 hypothetical patients were calculated and presented in Fig. 4. The probabilities of HCC development were discriminated according to the baseline clinical and laboratory data.

Discussion

In this study, we developed a novel deep-learning-based model, which can predict HCC development in patients with HBV-related cirrhosis, which showed a better performance than the previous 6 reported models. As far as we know, this is the first HCC prediction model by adapting the deep learning method. We confirmed the model performance by applying an independent validation cohort and compared the c-indexes with 6 previously developed models. This deep-learning-based model had the best performance among all the models. Moreover, the new model is promising because it can evolve its performance through further training with new data sets.

Prediction of HCC development among at-risk patients is important to establish an individualised surveillance strategy. However, the risk of HCC development varies according to age, race, and aetiologies.[16] Previous HCC prediction models were mostly developed with patients who were enrolled from the same race, same aetiologies, or same continents. The models for PAGE-B or mPAGE-B were only for patients with chronic hepatitis B. The models for CU-HCC, mPAGE-B, and HCC-RESCUE were developed from Asian patients, while the models for PAGE-B or ADRESS-HCC and THRI were from Caucasian patients. Thus, these models, which were developed based on homogeneous patients, had a low expandability and lower performances when they were applied to other patients of different races or continents. Although the new model was also based on Asian patients with chronic hepatitis B, it can have high expandability through training with additional data set from other races, continents, or aetiologies. The deep-learning-based model can upgrade its performance accumulatively and continuously, even after it is released. This is the most important difference between the deep-learning-based model and the previous models.

When evaluating the performance of the previous models of at-risk patients by applying our cohort, we identified the risk group for HCC development along with the original published data. Although the performances of the previous models were different, predictions were available in the previous models. Therefore, we found that the factors that consisted of the models were actually effective. Among the prognostic factors, including the 6 previous models, age was included in all models, sex was included in 4 models, and platelet count was included in 3 models. The deep-learning-based model included 7 possible factors: age, gender, platelet, albumin, bilirubin, HBV DNA titre, and presence of diabetes.[5–10] Three factors among all the factors of the 6 previous models were excluded in the deep-learning-based model. The presence of cirrhosis, aetiology, and race could not be included in this model, because our derivation cohort included Asian patients with HBV-related cirrhosis.

Previously, there was no effective deep-learning-based prediction model that had been trained from numerical data, not image data. Through this model, we could confirm the potential of the deep-learning-based prediction model. First, the deep-learning-based model performed better than the previous models, which were based on conventional statistics. Moreover, this better performance could be achieved with a relatively small amount of data, while usually the deep learning method may require a large amount of data at least, generally.[17] We showed that only hundreds of pieces of data would be enough to train a deep-learning-based prediction model, if the potential factors for the new model were well identified, such as age, sex, bilirubin, albumin, platelet, presence of diabetes mellitus, and HBV DNA in this model. Second, the deep learning method can develop a more accurate model and derive continuous probability results

(from 0 to 1), not a binary result. Previous models, which had been developed from conventional statistics, usually established a scoring system. The scoring system divided continuous variables into 2 or 3 categories with a certain cut-off value. The purpose of these scoring systems is to distinguish risk stratification. However, the deep-learning-based model could use the factors as an original continuous value without categorisation. Thus, the new model could result in more accurate continuous probabilities. Third, we found that training with a small number of data points (424 data points in our derivation cohort) was not a time-consuming task, and it could replace conventional statistical methods without any time delays, which is 1 of the general concerns of the deep learning method. In addition, we also found that training with a relatively small number of numerical data points, not image data sets, can be performed with a common laptop computer without requiring multiple graphics processing units. Fourth, the deep-learning-based model could evolve with data accumulation.[18] Previous conventional models have not changed and have the same performance despite further data accumulation. However, the deep-learning-based model can upgrade its performance through additional data training.

This study has several limitations. First, our model was developed based on patients with cirrhosis. Because the previous models included patients with chronic hepatitis B and analysed them, the generalisability of our model was limited compared with the previous models. However, it can be overcome through additional data-set training of patients with chronic hepatitis B. Second, the previous model performances were poor in our validation cohort, while external validation studies of the models showed a fair performance for each model.[9,19,20] This result might be caused by the relatively high risk for HCC development in our cohort, which consisted of patients with cirrhosis. Third, although the deep-learning-based model has shown its potential for HCC prediction, it does not represent an intuitive formula or scoring system for medical decisions. However, it can be overcome through a web-based application, which could be easily acceptable. Fourth, our model was developed in an Asian cohort with cirrhosis; therefore, additional validation study will be needed in non-Asian cohorts and cohorts with advanced fibrosis before cirrhosis.

In conclusion, the deep-learning-based model made in this study has better performance in HCC prediction of at-risk patients compared with the previous models, and its performance can be progressively improved with further data accumulation.

Abbreviations

ADRESS-HCC, age, diabetes, race, etiology of cirrhosis, sex, and severity HCC score; *c*-index, concordance index; CU-HCC, Chinese University HCC score; HCC, hepatocellular carcinoma; HCC-RESCUE, HCC-Risk Estimating Score in CHB patients Under Entecavir; mPAGE-B, modified platelet, age, and gender-hepatitis B score; PAGE-B, platelet, age, and gender-hepatitis B score; SMC, Samsung Medical Center; SNUBH, Seoul National University Bundang Hospital; THRI, Toronto HCC risk index; US, ultrasonography.

Financial support

The authors received no specific funding for this work.

Conflicts of interest

The authors declare no conflicts of interest that pertain to this work.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors' contributions

Conception: Joon Yeul Nam, Sook-Hyang Jeong.

Data collection: Joon Yeul Nam, Dong Hyun Sinn, Eun Sun Jang, Jin-Wook Kim, Sook-Hyang Jeong.

Data analysis: Joon Yeul Nam, Junho Bae, Sook-Hyang Jeong.

Article preparation: Joon Yeul Nam, Dong Hyun Sinn, Sook-Hyang Jeong.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhepr.2020.100175.

References

Author names in bold designate shared co-first authorship

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394–424.

[2] World Health Organization. Global Hepatitis Report 2017. 2017. Available at: https://www.who.int/hepatitis/publications/global-hepatitis-report2017/en/. [Accessed 21 April 2020].

[3] Galle PR, Forner A, Llovet JM, Mazzaferro V, Piscaglia F, Raoul J-L, et al. EASL Clinical Practice Guidelines: management of hepatocellular carcinoma. J Hepatol 2018;69:182–236.

[4] Heimbach JK, Kulik LM, Finn RS, Sirlin CB, Abecassis MM, Roberts LR, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. Hepatology 2018;67:358–380.

[5] Papatheodoridis G, Dalekos G, Sypsa V, Yurdaydin C, Buti M, Goulis J, et al. PAGE-B predicts the risk of developing hepatocellular carcinoma in Caucasians with chronic hepatitis B on 5-year antiviral therapy. J Hepatol 2016;64:800–806.

[6] Wong VW-S, Chan SL, Mo F, Chan T-C, Loong HH-F, Wong GL-H, et al. Clinical scoring system to predict hepatocellular carcinoma in chronic hepatitis B carriers. J Clin Oncol 2010;28:1660–1665.

[7] **Sohn W**, **Cho J-Y**, Kim JH, Lee JI, Kim HJ, Woo M-A, et al. Risk score model for the development of hepatocellular carcinoma in treatment-naive patients receiving oral antiviral treatment for chronic hepatitis B. Clin Mol Hepatol 2017;23:170–178.

[8] Flemming JA, Yang JD, Vittinghoff E, Kim WR, Terrault NA. Risk prediction of hepatocellular carcinoma in patients with cirrhosis: the ADRESS-HCC risk model. Cancer 2014;120:3485–3493.

[9] **Kim JH**, **Kim YD**, Lee M, Jun BG, Kim TS, Suk KT, et al. Modified PAGE-B score predicts the risk of hepatocellular carcinoma in Asians with chronic hepatitis B on antiviral therapy. J Hepatol 2018;69:1066–1073.

[10] Sharma SA, Kowgier M, Hansen BE, Brouwer WP, Maan R, Wong D, et al. Toronto HCC risk index: a validated scoring system to predict 10-year risk of HCC in patients with cirrhosis. J Hepatol 2018;68:92–99.

[11] Singal AG, Mukherjee A, Elmunzer BJ, Higgins PD, Lok AS, Zhu J, et al. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. Am J Gastroenterol 2013;108:1723–1730.

[12] **Jung KS**, **Kim SU**, Ahn SH, Park YN, Kim DY, Park JY, et al. Risk assessment of hepatitis B virus-related hepatocellular carcinoma development using liver stiffness measurement (FibroScan). Hepatology 2011;53:885–894.

[13] Korean Liver Cancer Study Group, National Cancer Center, Korea. 2014 KLCSG-NCC Korea practice guideline for the management of hepatocellular carcinoma. Gut Liver 2015;9:267–317.

[14] Korean Liver Cancer Association, National Cancer Center (NCC), Korea. 2018 Korean Liver Cancer Association–National Cancer Center Korea practice guidelines for the management of hepatocellular carcinoma. Gut Liver 2019;13:227–299.

[15] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV: 2016;770–778.

[16] Mittal S, Kramer JR, Omino R, Chayanupatkul M, Richardson PA, El-Serag HB, et al. Role of age and race in the risk of hepatocellular carcinoma in veterans with hepatitis B virus infection. Clin Gastroenterol Hepatol 2018;16:252–259.

[17] Cho J, Lee K, Shin E, Choy G, Do S. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? arXiv preprint arXiv:151106348 2015. Available at: https://arxiv.org/abs/1511.06348. [Accessed 18 August 2020].
[18] Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw 2015;61:85–117.
[19] Lee HW, Kim SU, Park JY, Kim DY, Ahn SH, Han KH, et al. External validation of the modified PAGE-B score in Asian chronic hepatitis B patients receiving antiviral therapy. Liver Int 2019;39:1624–1630.
[20] Yip TC-F, Wong GL-H, Wong VW-S, Tse Y-K, Liang LY, Hui VW-K, et al. Reassessing the accuracy of PAGE-B-related scores to predict hepatocellular carcinoma development in patients with chronic hepatitis B. J Hepatol 2020;72:847–854.