Original Article

A K-nearest Neighbor Model to Predict Early Recurrence of Hepatocellular Carcinoma After Resection

Chuanli Liu[1#], Hongli Yang[2#], Yuemin Feng[3#], Cuihong Liu[4], Fajuan Rui[1], Yuankui Cao[5], Xinyu Hu[2], Jiawen Xu[6], Junqing Fan[5], Qiang Zhu[3] and Jie Li[2,7,8*]

[1]Department of Infectious Disease, Shandong Provincial Hospital Affiliated to Shandong Frist Medical University, Ji'nan, Shandong, China; [2]Department of Infectious Disease, Shandong Provincial Hospital, Cheeloo College of Medicine, Shandong University, Ji'nan, Shandong, China; [3]Department of Gastroenterology, Shandong Provincial Hospital Affiliated to Shandong Frist Medical University, Ji'nan, Shandong, China; [4]Department of Ultrasound Diagnosis and Treatment, Shandong Provincial Hospital Affiliated to Shandong Frist Medical University, Ji'nan, Shandong, China; [5]School of Computer Science, China University of Geosciences, Wuhan, Hubei, China; [6]Department of Pathology, Shandong Provincial Hospital Affiliated to Shandong Frist Medical University, Ji'nan, Shandong, China; [7]Department of Infectious Diseases, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing, Jiangsu, China; [8]Institute of Viruses and Infectious Diseases, Nanjing University, Nanjing, Jiangsu, China

Abstract

Background and Aims: Patients with hepatocellular carcinoma (HCC) surgically resected are at risk of recurrence; however, the risk factors of recurrence remain poorly understood. This study intended to establish a novel machine learning model based on clinical data for predicting early recurrence of HCC after resection. **Methods:** A total of 220 HCC patients who underwent resection were enrolled. Classification machine learning models were developed to predict HCC recurrence. The standard deviation, recall, and precision of the model were used to assess the model's accuracy and identify efficiency of the model. **Results:** Recurrent HCC developed in 89 (40.45%) patients at a median time of 14 months from primary resection. In principal component analysis, tumor size, tumor grade differentiation, portal vein tumor thrombus, alpha-fetoprotein, protein induced by vitamin K absence or antagonist-II (PIVKA-II), aspartate aminotransferase, platelet count, white blood cell count, and HBsAg were positive prognostic factors of HCC recurrence and were included in the preoperative model. After comparing different machine learning methods, including logistic regression, decision tree, naïve Bayes, deep neural networks, and k-nearest neighbor (K-NN), we choose the K-NN model as the optimal prediction model. The accuracy, recall, precision of the K-NN model were 70.6%, 51.9%, 70.1%, respectively. The standard deviation was 0.020. **Conclusions:** The K-NN classification algorithm model performed better than the other classification models. Estimation of the recurrence rate of early HCC can help to allocate treatment, eventually achieving safe oncological outcomes.

Introduction

Primary liver cancer is the sixth most common cancer worldwide and the fourth leading cause of cancer-related death, accounting for ~90% of primary liver cancers.[1] Hepatocellular carcinoma (HCC) is the most frequent histologic type of liver cancer. The most effective first-line treatment is surgical resection for selected patients and is widely recommended by current guidelines.[2,3] However, patients with surgically resected HCC are still at risk of recurrence, with an annual rate of ≥ 10% and a recurrence rate of 70–80% after 5 years.[4] In addition, the reasons for postsurgical recurrence and how to prevent recurrence are unresolved. Therefore, identification of potentially curable patients at high risk for postoperative recurrence is critical to improve long-term survival after HCC resection.

HCC recurrence is the main postoperative complication, which is generally considered either early (less than 2 years) or late (more than 2 years).[5] However, early recurrence occurs in 30–50% of patients and accounts for more than 70% of tumor recurrences, and is the major cause of mortality. Previous studies have shown that early recurrence of HCC is usually related to aggressive tumor pathological features, such as large tumor size, multiple tumors,
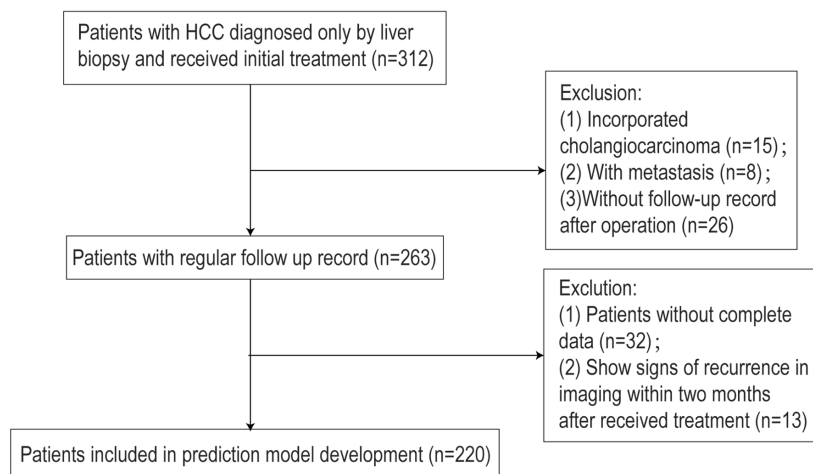
Fig. 1. Study cohort selection.

poor cell differentiation, and macroscopic or microscopic vascular invasion.[6] Other risk factors for HCC recurrence are cirrhosis, tumor size of > 5 cm, or portal vein invasion.[7]

The prognosis of HCC has traditionally been assessed by staging, such as the tumor-node-metastasis (TNM), Barcelona clinic liver cancer and Hong Kong liver cancer systems.[8–11] However, staging systems are not available to patients after surgical treatment and therefore do not predict postoperative recurrence. A few models including the Singapore liver cancer recurrence score and surgery-specific cancer of the liver Italian program (SS-CLIP),[12] have been developed specifically to detect tumor recurrence after surgical resection but none of them have been externally validated.[13]

Machine learning (ML) algorithms are techniques for data mining that use artificial intelligence to evaluate and analyze data, and can generate predictive models more efficiently and effectively than conventional methods by detecting hidden patterns within large data sets. Recent advances in ML models have helped to learn about features represented in data and to improve model performance in different HCC domains, including disease prediction, disease classification, and clinical practice.[14] Various types of model architectures have been used, such as logistic regression, k-nearest neighbor (K-NN), decision trees, naïve Bayes (NB), and deep neural networks (DNN).[15] Several examples of prognosis prediction methods using ML approaches based on pathological information to evaluate micro (mi)RNA expression in exosomes, circulating miRNA information, and to incorporate radiomics have been described,[16–19] but which tumor markers should be included in a surveillance program remains controversial. A more precise prognostic and recurrent prediction model is urgently needed.

In this study, we enrolled pathologically confirmed HCC patients to investigate the factors that are associated with tumor recurrence and to develop a prognostic model to improve the predictive accuracy for HCC recurrence using ML. We hope the model will provide clinicians with an appropriate surveillance tool for early detection of HCC recurrence and treatment.

Methods

Patient population

Of the 312 HCC patients diagnosed between September 2016 and June 2018 at Shandong Provincial Hospital, 220 patients recruited in this retrospective study. Patients (1) with HCC diagnosed by liver biopsy, (2) without other tumors on preoperative CT evaluation and, (3) receiving initial treatment were eligible for inclusion. Patients (1) with cholangiocarcinoma, or (2) metastasis, (3) without post-surgical follow-up; (4) younger than 18 years of age, and (5) with imaging evidence of recurrence within 2 months after treatment were excluded. All patients with HCC enrolled in this study were diagnosed by pathological evaluation. The study was approved by local Hospital Ethics Committee and patient informed consent was waived when data were collected. Figure 1 is a flow chart of patient selection. Patients were divided into two study groups by HCC recurrence and followed-up until recurrence of HCC, death, study conclusion on August 31, 2019. HCC recurrence of HCC was defined by clinical, radiological, and/or pathological diagnosis.

Dataset preparation

We collected patient-related clinical, laboratory, and radiological information from medical records and at follow-up visits. (Tables 1 and 2). Thirty-seven patient characteristics were collected, including. age, etiology, treatment strategy, degree of tumor differentiation, tumor size, number of tumors, platelet count (PLT), alkaline phosphatase (ALP), total bilirubin, prothrombin time (PT), alpha-fetoprotein (AFP), aspartate aminotransferase (AST), white blood cell (WBC) count, protein induced by vitamin K absence or antagonist-II (PIVKA-II), HBsAg, and others.

Evaluation metrics

We used logistic regression, K-NN, decision tree, NB, and DNN models to predict the recurrence of HCC from the patient information. The training cohort included 176 of the 220 patients; the testing cohort included the remaining 44. The training set contains a learned output that the model generalizes to new data. The algorithm flow is shown in Figure 2. The performance of the prediction results was evaluated by introducing four metrics, accuracy (Acc), precision (Prc), recall rate (TPR), and standard deviation (SD).

Acc was the ratio of the number of correctly classified samples and the total number of samples:

Table 1. Patient characteristics

Characteristics	All patients (N=220)	Patients with recurrence (N=89)	Patients without recurrence (N=131)	*p*-value
Age	56.65±10.39	55.89±10.63	57.16±10.23	0.37
Sex				
Male	192 (87.27%)	76 (85.39%)	116 (88.55%)	0.49
Female	28 (12.73%)	13 (14.61%)	15 (11.45%)	
Follow-up time	7.64±8.04	9.71±7.97	14.0±6.36	< 0.001
Hypertension	56 (25.45%)	24 (26.97%)	32 (24.42%)	0.67
Diabetes	27 (12.27%)	12 (13.48%)	15 (11.45%)	0.65
Fatty liver	9 (4.09%)	2 (2.25%)	7 (5.34%)	0.25
Cirrhosis	186 (84.55%)	70 (78.65)	96 (73.28%)	0.36
Family history of liver cancer	14 (6.37%)	7 (7.86%)	7 (5.34%)	0.45
Etiology				
Alcohol	8 (3.64%)	3 (3.37%)	5 (3.82%)	0.83
HBV	131 (59.55%)	54 (60.67%)	77 (58.78%)	
HCV	5 (2.28%)	1 (1.12%)	4 (3.05%)	
Alcohol and HBV	64 (29.09%)	25 (28.09%)	39 (29.77%)	
Others	12 (5.45%)	6 (6.74%)	6 (4.58%)	
Treatment strategy				
Tumor resection	131 (59.55%)	47 (52.80%)	84 (64.12%)	0.09
Resection and TACE	89 (40.45%)	42 (47.20%)	47 (35.88%)	
Portal vein tumor thrombus				
With	41 (18.64%)	26 (29.21%)	15 (11.45%)	< 0.001
Without	179 (81.36%)	63 (70.79%)	116 (88.55%)	
Degree of tumor differentiation				
Poorly differentiated	39 (17.72%)	22 (24.72%)	17 (12.98%)	0.08
Moderately differentiated	162 (73.64%)	60 (67.42%)	102 (77.86%)	
Well differentiated	19 (8.64%)	7 (7.86%)	12 (9.16%)	
Tumor size				
≤5cm	133 (60.45%)	42 (47.19%)	91 (69.47%)	< 0.001
>5cm	87 (39.55%)	47 (52.81%)	40 (30.53%)	
Number of tumors				
Solitary	186 (84.55%)	73 (82.02%)	113 (86.26%)	0.50
2–3	34 (15.45%)	16 (17.98%)	18 (13.74%)	

TACE, trans arterial chemoembolization.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}.$$

In the confusion matrix of classification results TP represents the positive samples that are predicted to be positive by the model, FP represents the negative samples that are predicted to be positive by the model, FN represents the positive samples that are predicted to be negative by the model, and TN represents the negative samples that are predicted to be negative by the model. Prc is the ratio of the number of correctly classified positive instances and the number of instances classified as positive:

$$Prc = \frac{TP}{TP + FP}.$$

TPR was the proportion of the number of positive cases correctly classified to the actual number of positive cases:

$$TPR = \frac{TP}{TP + FN}.$$

The SD is the extent of dispersion of the accuracy of random tests:

Table 2. Patient laboratory findings

Variables	All patients (N=220)	Patients with recurrence (N=89)	Patients without recurrence (N=131)	*p*-value
White blood cell count, ×10⁹/L	5.1 (2–82)	5.2 (2.1–82)	5.1 (2–15)	0.62
Red blood cell count,	4.7 (1.7–5.8)	4.7 (1.7–5.8)	4.7 (3.1–5.6)	0.44
Hemoglobin, g/L	14 (6–84)	15 (10–84)	14 (6–82)	0.44
Platelet count, ×10⁹/L	175.30±82.63	184.79±81.72	168.86±82.93	0.16
Alanine aminotransferase, U/L	30.5 (10–581)	37 (12–581)	36 (10–209)	0.64
Aspartate aminotransferase, U/L	38 (9–317)	38.0 (9–317)	38.0 (16.00–249.00)	0.29
Alkaline phosphatase, U/L	76.5 (12–968)	94 (23–968)	61 (12–807)	0.005
γ-glutamyl transpeptadase, U/L	104 (14–619)	106 (14–427)	103 (19–619)	0.06
Total bilirubin, m/L	17 (7–74)	16 (7–47)	18 (7–74)	0.06
Direct bilirubin, um/L	3 (1–97)	3 (1–97)	3 (1–64)	0.77
Indirect bilirubin, μm/L	13 (5–61)	13 (5–61)	14 (5–56)	0.06
ALB, g/L	41.59±5.18	41.55±4.41	41.85±5.65	0.38
Glucose, mmol/L	5.0 (2–14)	5.0 (4–13)	5.0 (2–14)	0.41
Cholesterol	4.39±1.39	4.60±1.37	4.23±1.39	0.27
Triglycerides, mmol/L	0.88 (0.3–2.79)	0.77 (0.3–1.8)	0.9 (0.42–2.79)	0.04
High-density lipoprotein, mmol/L	1.21 (0.37–4.06)	1.25 (0.4–4.06)	1.20 (0.37–3.19)	0.95
Low-density lipoprotein, mmol/L	2.59±0.93	2.73±0.97	2.49±0.89	0.30
PT, s	13 (10–18)	13 (10–17)	13 (10–18)	0.58
PTA, %	85.45±13.36	85.23±13.62	85.61±13.23	0.84
Alpha-fetoprotein, ng/mL	27.0 (1.1–998.0)	59 (1.5–919.0)	15.0 (1.1–998.0)	0.001
PIVKA-II, ng/mL	604.81 (9.38–75,000)	1,519.5 (16.00–75,000)	355.29 (9.38–75,000)	0.001
Fibrosis-4 (FIB-4)				
Low (<1.45)	43 (19.55%)	20 (22.47%)	23 (17.56%)	0.41
Intermediate (1.45–3.25)	110 (50.0%)	46 (51.69%)	64 (48.85%)	
High (>3.25)	47 (21.35%)	23 (25.84%)	44 (33.59%)	
HBsAg, IU/mL	5,790.5 (0.39–8,724.0)	5,828.0 (0.41–8,002.0)	5,122.0 (0.39–8,724)	0.78

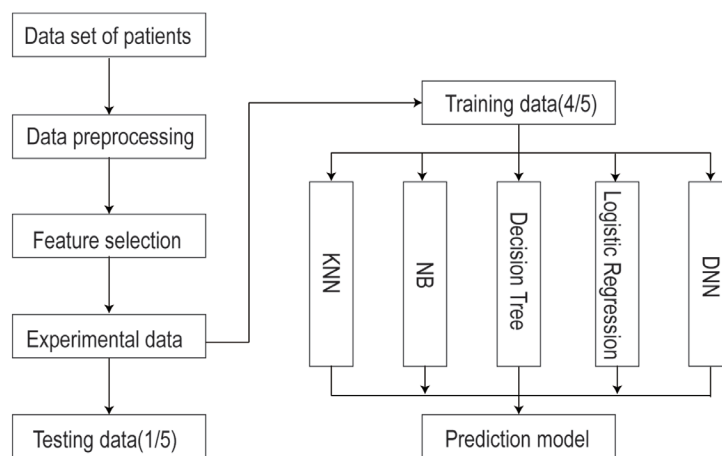PIVKA-II, protein induced by vitamin K absence or antagonist-II.

Fig. 2. Algorithm flow. K-NN, k-nearest neighbor; NB, naïve Bayes; DNN, deep neural networks.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i - \mu},$$

where x_1, x_2, …, xn are real numbers, μ is the arithmetic mean, and σ is the SD.

Statistical analysis

Continuous variables were reported as means (SD) if they were normally distributed or a medians (IQR) if they were not. Categorical variables were reported as numbers and percentages (%). We assessed differences between severe and nonsevere patients with two-sample *t*-tests or the Wilcoxon rank-sum test depending on parametric or nonparametric data for continuous variables and Fisher's exact test for categorical variables. A two-sided α of less than 0.05 was considered statistically significant. The statistical analysis was performed with SPSS version 26.0 (IBM Corp., Armonk, NY, USA).

In the building of the predictive models, the Pearson correlation coefficient was used to find the independent predictors of severity of disease from 37 vectors. The predictive models were built based on five ML classification algorithms, i.e. logistic regression, K-NN, decision tree, NB and DNN model by using Python programming software version 3.6.5.

1. Pearson correlation coefficient and feature selection by univariate analysis were used. The Pearson coefficient between each patient characteristic and recurrence was calculated separately, and the characteristics with significant correlations were selected. The specific steps were as follows: To make the characteristics in the dataset $D = \{x_1, x_2, …, x_m, y\}$ numerically comparable, the absolute values, maxima and minima of each were mapped to [0, 1];

$$x_i \leftarrow \frac{x_i - \min_{x_i}}{\max_{x_i} - \min_{x_i}}, y \leftarrow \frac{y - \min_y}{\max_y - \min_y}, D \leftarrow \{x_1, x_2, \cdots, x_m, y\}.$$

2. The correlation between each feature and the tag value $p(x_i, y)$

$$p(x_i, y) = \frac{\sum_{n=1}^{n}\left(x_i^k - \overline{x}_i\right)\left(y^k - \overline{y}\right)}{\sqrt{\sum_{k}^{n}\left(x_i^k - \overline{x}\right)^2 \sum_{k}^{n}\left(x_i^k - \overline{y}_i\right)^2}}$$

where x_i^k, y^k represent the value of the k-th sample of the characteristic, and $\overline{x}_i$, $\overline{y}$ represent the sample mean value of the two characteristics, represents the total number of characteristics in the patient data.

3. To calculate the eigenvectors and eigenvalues of the covariance matrix the features with large influencing factors were selected as the optimal feature subset. The final data set was constructed based on the feature subset.
The K-NN algorithm was constructed as follows:

1. For data set , the distance from each sample $d_i = (x_i, y_i)$ to be classified x to all known samples, $L(d_i, d_j)$;

$$L\left(d_i, d_j\right) = \left(\sum_{l=1}^{m}\left|x_i^{(l)} - x_j^{(l)}\right|^2 + \left|y_i - y_j\right|^2\right)^{1/2}$$

was constructed.

2. Adjacent values of each sample were sorted in descending order according to the distance.

3. The k-nearest neighbors of each sample are obtained by determining the K value. According to the majority voting rule of the following formula, the sample x to be classified is classified into the category with the largest number of samples:

$$C_x = \mathrm{argmax}_{j \in I} \Sigma_{y=} xkI(C_y = j)$$

Where j represents the tag values of different categories, and Y represents the k-nearest neighbors of sample x to be classified.

Results

Patient characteristics

The clinical characteristics of the patients are shown in Table 1. Most patients were men (192/220, 82.27%), and the mean age was 56.65 (SD = 10.39) years. Of the 220 HCC patients, 89 (40.5%) were recrudescent and 131(59.5%) were nonrecrudescent. The mean time from surgery to recurrence was 14 (SD = 6.36) months. Patients with recurrent HCC were more likely to have larger tumors (> 5 cm diameter, 52.81% vs. 30.53%, $P < 0.001$) and portal vein tumor thrombus (29.21% vs. 11.45%, $P < 0.001$). Some differences in the laboratory values of patients with recurrent and nonrecurrent HCC obtained on admission (Table 2) were significant.

Performance comparison

In principal component analysis, we found nine key factors affecting the recurrence of HCC, including tumor size, tumor grade differentiation, portal vein tumor thrombus, PLT, AFP, PIVKA-II, AST, WBC, and HBsAg (Fig. 3). Tumor size, tumor differentiation grade, portal vein tumor thrombus, PLT, AFP, PIVKA-II, AST, WBC, HBsAg, and recurrence results of 176 patients in the training cohort were formed into a data set. The data sets were input into different ML algorithms (i.e. logistic regression, K-NN, decision tree, naïve Bayes, and DNN) to form the ML model. Then the data of 44 patients in the testing cohort were input into the five ML models for prediction. The prediction results from different models were evaluated by comparing the model performance metrics. The accuracies of the K-NN (70.6%), NB (60.9%), decision tree (57.5%), logistic regression (67.9%), and DNN (64.9%) models is reported in Figure 4. After comparing different ML methods, we choose the K-NN model as the optimal prediction model. In terms of accuracy and precision, K-NN algorithm was superior to other algorithms. It had 70.6% Acc and 70.1% Prc. The TPR was 51.9% and the SD was 0.02.

Discussion

The ideal resection index is early solitary HCC, regardless of tumor size, and preserved liver function. Unfortunately, the rate of disease recurrence remains high, with early relapses considered to be "true relapses" and "relapses" afterward assumed to be mainly caused by *de novo* tumors. However, there is no reliable prediction tool for early HCC recurrence. In this study, we retrospectively evaluated 89 patients with early recurrence of HCC, which had different clinical characteristics and laboratory parameters compared with non-recurrent patients. Using Pearson analysis, we discovered that early recurrence was mainly determined by aggressive characteristics of the primary (resected) tumor, including size, grade, differentiation, and higher serum AFP, PIVKA-II, PLT, AST, WBC, and HBsAg levels.

Currently, we can only use tumor markers such as AFP and PIVKA-II to determine HCC recurrence, because there is no useful postoperative recurrence marker. AFP is the most commonly used clinical biomarker of HCC, but its sensitivity and specificity are not ideal. AFP is a risk fac-
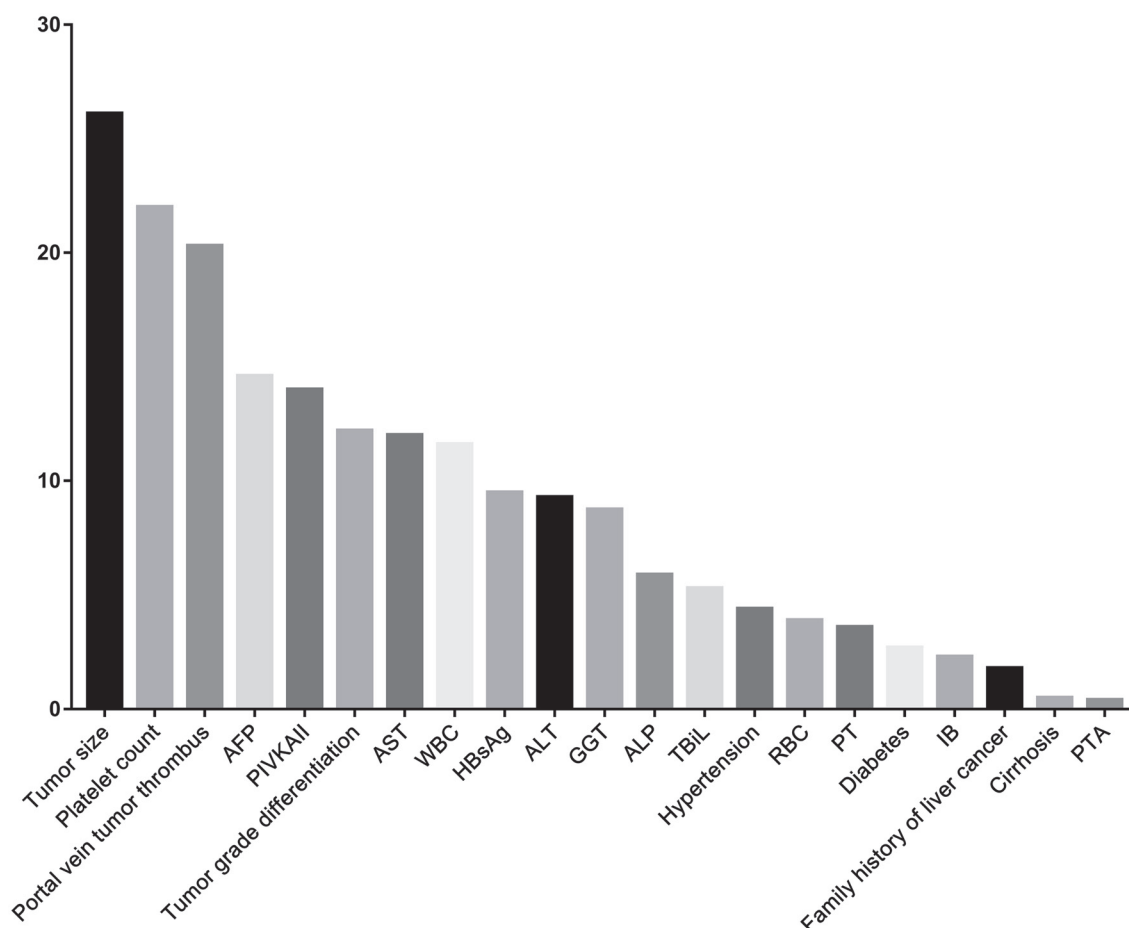
Fig. 3. Variable importance plot for predicting tumor recurrence showing absolute values of Spearman correlation coefficients between markers and HCC recurrence. AFP, alpha-fetoprotein; PIVKA-II, protein induced by vitamin K absence or antagonist-II; AST, aspartate amino transferase; WBC, white blood cells; HBsAg, hepatitis B surface antigen; ALT, alanine aminotransferase; GGT, gamma-glutamyl transpeptidase; ALP, alkaline phosphatase; TBiL, total bilirubin; RBC, red blood cell; PT, prothrombin time; IB, indirect bilirubin; PTA, prothrombin activity.

tor for the recurrence of HCC after radical treatment, and has been considered as a better prognostic predictor than cancer morphology alone.[20,21] PIVKA-II may play a role in the progression of HCC and is associated with HCC size, microvascular invasion, metastatic dissemination, and recurrence after tumor ablation. In fact, AFP levels are high in 40–60% of HCC patients and in 10–20% of early-stage tumors. It may also be elevated in many benign tumors.[22–24] Other studies have shown that the performance of PIVKA-II in HBV-related HCC varies across populations, with a sensitivity of 44–91% and specificity of 68–99% at a cutoff values between 40 and 150 mAU/mL.[25] The evidence supports the need for more sensitive and specific HCC markers, no method to predict the recurrence of surgically resected HCC is currently available.

Given the validated, good discriminatory performance of AFP and PIVKA-II prediction models, we studied a novel predictor of HCC recurrence based on the AFP model and including 36 additional serological, pathological, and radiological patient features. Nine features, tumor size, tumor grade differentiation, portal vein tumor thrombus, PLT, AFP, PIVKA-II, AST, WBC, and HBsAg were found influence the recurrence of HCC. The accuracy, recall, and precision of the model were 70.6%, 51.9%, 70.1%, respectively. The inclusion of more clinical markers might further improve the diagnostic accuracy.

In recent years, ML has developed rapidly, and has contributed to outstanding achievements in disease prediction and clinical practice. ML algorithms can be used to predict the outcome of a new observation, based on a training dataset containing previous observations where the outcome is known. It can detect complex nonlinear relationships between numerous variables that are useful in predictive applications.[26,27] Many research results show that prediction models based on ML significantly improve the accuracy of cancer diagnosis and prognosis prediction.[28–30] In this study, after data training and performance comparisons, we found a novel, sensitive, and stable K-NN model to predict the recurrence of HCC after surgery. We believe that it can help to identify individuals who are at high risk of early recurrence after tumor resection. K-NN algorithms are very effective nonparametric models that are widely used for classification, regression, and pattern recognition. It is highly appropriate to use the K-NN method to predict HCC recurrence of HCC, especially using a large chronic liver disease, tumor characteristics, and hepatic function dataset. The K-NN model was the optimal prediction model, with 70.6% accuracy. When developing the model to predict the risk of patient recurrence, we input nine key factors, tumor size, grade, and differentiation; portal vein tumor thrombus, PLT, AFP, PIVKA-II, AST, WBC, and HBsAg in the K-NN algorithm, which then was able to automatically estimate the HCC recurrence risk
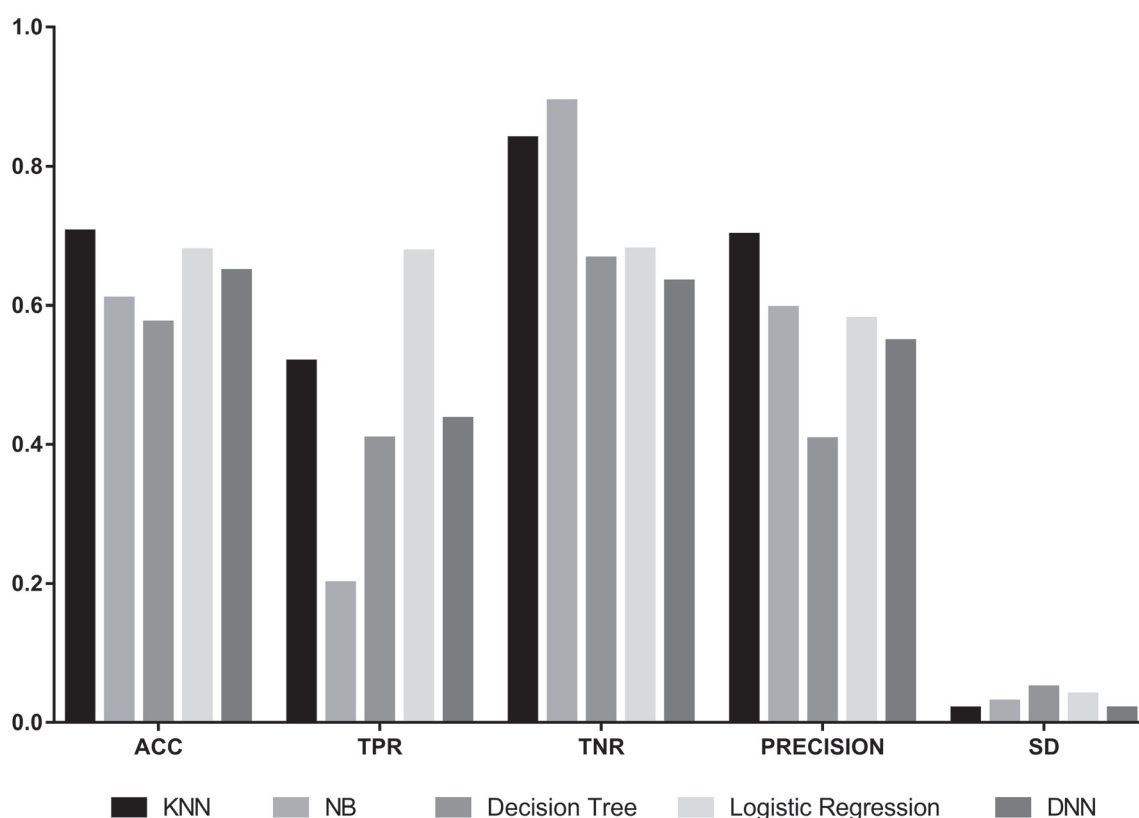
Fig. 4. Accuracy, recall rate, true negative rate, precision, and standard deviation of different algorithms. K-NN, k-nearest neighbor; NB, naïve Bayes; DNN, deep neural networks; ACC, accuracy; TPR, recall rate; TNR, true negative rate; SD, standard deviation.

of each patient.

This study has several limitations. It was limited by the small sample size and retrospective method. Some cases had incomplete documentation of laboratory testing, and most of the HCC patients included in our study had chronic hepatitis B infection. The limitations might have result in some bias in our general understanding of the disease. In addition, early and late recurrence were not distinguished in this study because of the relatively short follow-up. The two problems mentioned above can be resolved by additional study. The main limitations of ML algorithms are that they are best suited to predicting outcomes in the environment from which they are derived. Conversely, this limitation is also its strength, in that it is highly specific to the peculiarities of a particular center, enabling the best decision for each individual patient.

In conclusion, used ML to develop a K-NN model for predicting HCC recurrence that included a comprehensive evaluation of serological, pathological, and radiological features. The accuracy of this model was about 70.6%, which is much better than the models using only clinical or serological data. This K-NN model was sensitive and stable when used to predict the recurrence of HCC in patient after surgical resection.

Conflict of interest

JL has been an editorial board member of *Journal of Clinical and Translational Hepatology* since 2021. The other authors have no conflict of interests related to this publication.

Author contributions

Guarantor of article (JL), study concept and study supervision (JL), data collection and/or data interpretation (CL, HY, YF, YC), data analysis (JL), manuscript drafting (CL, HY, YF). All authors read and revised the manuscript.

Data sharing statement

All data are available upon request.

References

[1] 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(1):358–380. doi:10.1002/hep.29086.
[2] EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. J Hepatol 2018;69(1):182–236. doi:10.1016/j.jhep.2018.03.019.
[3] Marrero JA, Kulik LM, Sirlin CB, Zhu AX, Finn RS, Abecassis MM, *et al*. Diagnosis, Staging, and Management of Hepatocellular Carcinoma: 2018 Practice Guidance by the American Association for the Study of Liver Diseases. Hepatology 2018;68(2):723–750. doi:10.1002/hep.29913.
[4] 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(6):394–

424. doi:10.3322/caac.21492.

[5] Zheng Q, Xu J, Gu X, Wu F, Deng J, Cai X, *et al*. Immune checkpoint targeting TIGIT in hepatocellular carcinoma. Am J Transl Res 2020;12(7):3212–3224.

[6] Yang Y, Ying G, Wu S, Wu F, Chen Z. In vitro inhibition effects of hepatitis B virus by dandelion and taraxasterol. Infect Agent Cancer 2020;15:44. doi:10.1186/s13027-020-00309-4.

[7] Zhou QH, Wu FT, Pang LT, Zhang TB, Chen Z. Role of γδT cells in liver diseases and its relationship with intestinal microbiota. World J Gastroenterol 2020;26(20):2559–2569. doi:10.3748/wjg.v26.i20.2559.

[8] Vauthey JN, Lauwers GY, Esnaola NF, Do KA, Belghiti J, Mirza N, *et al*. Simplified staging for hepatocellular carcinoma. J Clin Oncol 2002;20(6):1527–1536. doi:10.1200/jco.2002.20.6.1527.

[9] Llovet JM, Brú C, Bruix J. Prognosis of hepatocellular carcinoma: the BCLC staging classification. Semin Liver Dis 1999;19(3):329–338. doi:10.1055/s-2007-1007122.

[10] Kudo M, Chung H, Haji S, Osaki Y, Oka H, Seki T, *et al*. Validation of a new prognostic staging system for hepatocellular carcinoma: the JIS score compared with the CLIP score. Hepatology 2004;40(6):1396–1405. doi:10.1002/hep.20486.

[11] Okuda K, Ohtsuki T, Obata H, Tomimatsu M, Okazaki N, Hasegawa H, *et al*. Natural history of hepatocellular carcinoma and prognosis in relation to treatment. Study of 850 patients. Cancer 1985;56(4):918–928. doi:10.1002/1097-0142(19850815)56:4<918::aid-cncr2820560437>3.0.co;2-e.

[12] Chen P, Wang YY, Chen C, Guan J, Zhu HH, Chen Z. The immunological roles in acute-on-chronic liver failure: An update. Hepatobiliary Pancreat Dis Int 2019;18(5):403–411. doi:10.1016/j.hbpd.2019.07.003.

[13] Deng JW, Yang Q, Cai XP, Zhou JM, E WG, An YD, *et al*. Early use of dexamethasone increases Nr4a1 in Kupffer cells ameliorating acute liver failure in mice in a glucocorticoid receptor-dependent manner. J Zhejiang Univ Sci B 2020;21(9):727–739. doi:10.1631/jzus.B2000249.

[14] Ngiam KY, Khor IW. Big data and machine learning algorithms for healthcare delivery. Lancet Oncol 2019;20(5):e262–e273. doi:10.1016/s1470-2045(19)30149-4.

[15] Spann A, Yasodhara A, Kang J, Watt K, Wang B, Goldenberg A, *et al*. Applying Machine Learning in Liver Disease and Transplantation: A Comprehensive Review. Hepatology 2020;71(3):1093–1105. doi:10.1002/hep.31103.

[16] Itami-Matsumoto S, Hayakawa M, Uchida-Kobayashi S, Enomoto M, Tamori A, Mizuno K, *et al*. Circulating Exosomal miRNA Profiles Predict the Occurrence and Recurrence of Hepatocellular Carcinoma in Patients with Direct-Acting Antiviral-Induced Sustained Viral Response. Biomedicines 2019;7(4):87. doi:10.3390/biomedicines7040087.

[17] Yamamoto Y, Kondo S, Matsuzaki J, Esaki M, Okusaka T, Shimada K, *et al*. Highly Sensitive Circulating MicroRNA Panel for Accurate Detection of Hepatocellular Carcinoma in Patients With Liver Disease. Hepatol Commun 2020;4(2):284–297. doi:10.1002/hep4.1451.

[18] Corredor G, Wang X, Zhou Y, Lu C, Fu P, Syrigos K, *et al*. Spatial Architecture and Arrangement of Tumor-Infiltrating Lymphocytes for Predicting Likelihood of Recurrence in Early-Stage Non-Small Cell Lung Cancer. Clin Cancer Res 2019;25(5):1526–1534. doi:10.1158/1078-0432.ccr-18-2013.

[19] Ji GW, Zhu FP, Xu Q, Wang K, Wu MY, Tang WW, *et al*. Machine-learning analysis of contrast-enhanced CT radiomics predicts recurrence of hepatocellular carcinoma after resection: A multi-institutional study. Ebiomedicine 2019;50:156–165. doi:10.1016/j.ebiom.2019.10.057.

[20] Hakeem AR, Young RS, Marangoni G, Lodge JP, Prasad KR. Systematic review: the prognostic role of alpha-fetoprotein following liver transplantation for hepatocellular carcinoma. Aliment Pharmacol Ther 2012;35(9):987–999. doi:10.1111/j.1365-2036.2012.05060.x.

[21] Mazzaferro V, Droz Dit Busset M, Bhoori S. Alpha-fetoprotein in liver transplantation for hepatocellular carcinoma: The lower, the better. Hepatology 2018;68(2):775–777. doi:10.1002/hep.29835.

[22] Hakamada K, Kimura N, Miura T, Morohashi H, Ishido K, Nara M, *et al*. Des-gamma-carboxy prothrombin as an important prognostic indicator in patients with small hepatocellular carcinoma. World J Gastroenterol 2008;14(9):1370–1377. doi:10.3748/wjg.14.1370.

[23] Nanashima A, Morino S, Yamaguchi H, Tanaka K, Shibasaki S, Tsuji T, *et al*. Modified CLIP using PIVKA-II for evaluating prognosis after hepatectomy for hepatocellular carcinoma. Eur J Surg Oncol 2003;29(9):735–742. doi:10.1016/j.ejso.2003.08.007.

[24] Kim DY, Paik YH, Ahn SH, Youn YJ, Choi JW, Kim JK, *et al*. PIVKA-II is a useful tumor marker for recurrent hepatocellular carcinoma after surgical resection. Oncology 2007;72(Suppl 1):52–57. doi:10.1159/000111707.

[25] Loglio A, Iavarone M, Facchetti F, Di Paolo D, Perbellini R, Lunghi G, *et al*. The combination of PIVKA-II and AFP improves the detection accuracy for HCC in HBV caucasian cirrhotics on long-term oral therapy. Liver Int 2020;40(8):1987–1996. doi:10.1111/liv.14475.

[26] Giger ML. Machine Learning in Medical Imaging. J Am Coll Radiol 2018;15(3 Pt B):512–520. doi:10.1016/j.jacr.2017.12.028.

[27] Venkatesh R, Balasubramanian C, Kaliappan M. Development of Big Data Predictive Analytics Model for Disease Prediction using Machine learning Technique. J Med Syst 2019;43(8):272. doi:10.1007/s10916-019-1398-y.

[28] Montazeri M, Montazeri M, Montazeri M, Beigzadeh A. Machine learning models in breast cancer survival prediction. Technol Health Care 2016;24(1):31–42. doi:10.3233/thc-151071.

[29] Hasnain Z, Mason J, Gill K, Miranda G, Gill IS, Kuhn P, *et al*. Machine learning models for predicting post-cystectomy recurrence and survival in bladder cancer patients. PLoS One 2019;14(2):e0210976. doi:10.1371/journal.pone.0210976.

[30] Kim I, Choi HJ, Ryu JM, Lee SK, Yu JH, Kim SW, *et al*. A predictive model for high/low risk group according to oncotype DX recurrence score using machine learning. Eur J Surg Oncol 2019;45(2):134–140. doi:10.1016/j.ejso.2018.09.011.