Current updates in machine learning in the prediction of therapeutic outcome of hepatocellular carcinoma: what should we know?

Zhi-Min Zou1, De-Hua Chang2, Hui Liu1 and Yu-Dong Xiao1*

Abstract

With the development of machine learning (ML) algorithms, a growing number of predictive models have been established for predicting the therapeutic outcome of patients with hepatocellular carcinoma (HCC) after various treatment modalities. By using the different combinations of clinical and radiological variables, ML algorithms can simulate human learning to detect hidden patterns within the data and play a critical role in artificial intelligence techniques. Compared to traditional statistical methods, ML methods have greater predictive effects. ML algorithms are widely applied in nearly all steps of model establishment, such as imaging feature extraction, predictive factor classification, and model development. Therefore, this review presents the literature pertaining to ML algorithms and aims to summarize the strengths and limitations of ML, as well as its potential value in prognostic prediction, after various treatment modalities for HCC.

Keywords: Hepatocellular carcinoma, Machine learning, Predictive, Modality

Key points

1. To highlight the effectiveness of machine learning algorithm on the prediction of therapeutic outcome for hepatocellular carcinoma after various treatment modalities
2. To illustrate the advantages and disadvantages of each machine learning algorithm
3. To familiarize the challenges of selecting a machine learning algorithm when creating a model

Introduction

Hepatocellular carcinoma (HCC) is an aggressive tumor which remains the second-most frequent cause of cancer death worldwide [1–3]. According to the different statuses of patients with HCC, several guidelines [4–7] recommend various treatment strategies. Due to the aggressive biological behavior of HCC, recurrence is not uncommon. Therefore, it is essential to predict therapeutic outcomes prior to treatment so that physicians can design a personalized therapeutic strategy for each patient. The conventional process of model establishment is selecting the appropriate predictors, utilizing them for statistical analysis and ultimately deriving a multivariate predictive model [8–12]. However, predictive models developed by traditional statistical methods, such as the logistic regression (LR) model and Cox proportional hazards model, are not reliable because the factors included in the models are too simple and utilize a low evidence level. Machine learning (ML) is a powerful tool
for generating high-level medical features or combining quantitative radiomic parameters with efficient algorithms [13–16]. ML algorithms simulate human learning to detect hidden patterns within HCC therapeutic data that are clearer than those derived from traditional statistical methods. With this in mind, ML algorithm has been used in many studies to predict the therapeutic outcome of HCC patients. Thus, in this review, the advantages and disadvantages of each ML algorithm are clarified, and relevant literature on the prediction of therapeutic outcomes after various treatment modalities for HCC is described.

Advantages and disadvantages of the ML algorithm
ML algorithms have several advantages over traditional statistical methods. First, traditional statistical methods can only process the variables that have a linear relationship with the outcome [12], whereas ML algorithms have the ability to process nonlinear data. Second, ML algorithms can learn from existing data to find novel patterns between variables and generate predictions [17–20]. Third, the ML model may contain more variables [21, 22] since the variables do not simply rely on the selection of traditional statistical methods [23–25]. Last, ML methods can process big data at a high speed.

Although ML algorithms are increasingly used, the disadvantages of ML algorithms should be kept in mind. First, the current ML methods are still not readily available for clinical practice, and the design of the ML model is not standard. Second, the lack of perfect generalization capability is still a common issue in clinical practice. The detailed advantages and disadvantages of the ML algorithm are listed in Table 1.

ML models in the prediction of therapeutic outcomes for HCC
With the development of ML algorithms, a growing number of studies have developed prognostic predictive models for HCC using the ML method. Therefore, understanding how ML works is essential. In this section, various ML models are introduced.

Neural networks
Neural networks are a classic ML method that simulates human brain neural networks. The most widely used neural networks are artificial neural networks (ANNs) and deep neural networks (DNNs). ANN [26] is one of the earliest neural network models and can be divided into three components: an input layer, a hidden layer, and an output layer. The ANN model can include a perceptron or a multilayer perceptron (MLP) (Fig. 1), with or without a hidden layer. However, ANNs cannot directly deal with medical imaging. With the development of deep learning, DNNs are widely used in establishing models [27]. Convolutional neural networks (CNNs) [28, 29] are one of the most common DNNs that can automatically identify and segment medical imaging. Another type of DNN is the recurrent neural network (RNN). However, RNNs are limited in HCC prognostic studies because the RNN algorithm cannot process data over a large time span.

Support vector machines
A support vector machine (SVM) [30] is a type of two-category model aimed at finding the optimal separating hyperplane with the largest distance to the support vector of any class (Fig. 2). Due to the hyperplane concept, SVM is often used for the selection of parameters for which the parameters are selected by the correlation to the results. However, SVMs are only applied in studies with small sample sizes, as the number of support vectors in large datasets is still very big, which may increase the complexity and training time of SVM algorithms.

Decision tree and random forest
The decision tree (DT) [31] is easy to understand and adopts a form of yes or no question and comprises a root node, parent node and leaf node/terminal node. Unfortunately, with the increasing complexity of the

| Table 1 | Advantages and disadvantages of ML algorithm |
|---------|---------------------------------------------|
| **Advantages** | It can process a big data<br>It can process a nonlinear data<br>It can be used to select meaningful predictive variables and extract radiomic variables<br>It usually has higher predictive performance than traditional statistical model |
| **Disadvantages** | The accurate selection of ML algorithms is a challenge to establish a predictive model<br>The design of ML predictive models lacks standards<br>It is difficult to identify the process of ML model development due to the existence of “black box”<br>The generalization ability of the established model needs to be further confirmed in validation cohort |

ML machine learning
Bayesian networks

Bayesian networks (BNs) [35] are different from most ML algorithms. A BN is an extension of Bayes’ theorem and presents the causality under each variable via a directed acyclic graph (Fig. 4). Therefore, those algorithms can visualize information. BNs have been applied to analyze predictors for survival in postsurgical HCC patients through conditional probability tables (CPTs) [36]. However, the relationship between each variable in the BN model is not always clear, which leads to low accuracy.

Methods

A search of PubMed was conducted for a prognostic predictive model for HCC published from January 1995 to May 2020. The following search algorithm was created: “hepatocellular carcinoma” and “model” and “predict” or “prognostic/prognosis” and “machine learning” or “neural network” or “support vector machine” or “decision tree” or “random forest” or “Bayesian network”. Initially, a total of 291 relevant research articles were searched, and the literature selection process is shown in Fig. 5. Ultimately, 29 articles were enrolled in the final analysis.

Prediction of therapeutic outcomes by various treatment modalities

There are various treatment options for HCC. Surgical resection, ablative therapy, and liver transplantation (LT) are potentially curative treatments, and transarterial chemoembolization (TACE) and sorafenib are palliative treatments [37–40]. Due to the poor prognosis of HCC patients, it is essential to create a suitable predictive model for predicting therapeutic outcomes prior to treatment. In this section, the current updates of ML algorithms are reviewed for various treatment modalities in HCC patients.

Surgical resection

Partial hepatectomy remains the mainstay of curative treatment in the early stage of HCC. Intrahepatic recurrence of HCC after surgical resection is the major cause of death, as the incidence rate is approximately 70% at 5 years [41, 42]. Therefore, an accurate prediction of prognosis prior to resection is crucial.

In previous studies, the authors developed a predictive ANN model [43–47] to predict therapeutic outcomes after surgical resection, and the ANN model was verified to be superior to the LR model and Cox proportional hazards regression model. Unfortunately, the ANN model cannot be used to select variables, which may decrease the predictive accuracy when some potentially clinically meaningful variables are overlooked. Similar to ANN, the BN model [19, 36] also cannot be used to select...
variables. The predictive variables within the BN model are based on the clinician’s experience and knowledge, and the associated relationship between variables and outcome is not always clear; therefore, the performance of the BN model is generally confusing. Unlike ANN and BN, RF and SVM can be used to either select variables or develop models [48–54]. Wang et al. [52] used the RF algorithm to select 30 radiomic features from 3144 MR texture features and developed a predictive RSF model for the 5-year survival of HCC following surgical resection with an area under curve (AUC) of 0.980. In addition, Liao et al. developed an RF model [53] based on 46 features from whole slide images (WSIs), and the results showed comparable accuracy to the TNM staging system in predicting the prognosis of HCC patients after surgical resection. However, it should be noted that the sample size in the clinical study of HCC is usually small, and the SVM model is theoretically more suitable than other
models. Xu et al. used an immunohistochemistry (IHC)-based SVM algorithm [48] to predict the recurrence of 336 HCC patients after surgical resection. The SVM model finally selected 8 features from 49 features and had an accuracy of 82.1%. In comparison to the above-mentioned ML algorithms, the CNN algorithm has great convenience in establishing predictive models because they can be used not only to segment imaging but also to select parameters and to develop models [23, 55]. Wang et al. [23] used the CNN algorithm to extract high-level temporal and spatial features from multiphase CT imaging using an automatic mode, which showed high efficacy with an AUC of 0.825 for predicting the early recurrence of HCC. Nevertheless, the automatic mode based on the CNN algorithm requires high computational power and thus has limited use. The relevant papers are listed in Table 2.

**LT**

LT is regarded as an effective therapy in HCC patients who are within the Milan criteria, with a recurrence rate of 10–15% [38, 56, 57]. Once post-LT HCC recurrence occurs, the prognosis is poor. Therefore, it is necessary to accurately identify HCC patients who will benefit from LT, thereby optimizing donor-recipient matching.

To our knowledge, ML-based analysis in predicting therapeutic outcomes for HCC after LT is rather limited. Marsh et al. [58, 59] developed an ANN model using seven clinical factors to predict the recurrence risk in HCC patients after LT, and the results showed that the discriminatory power was 70%. However, in the combination of this model and other variables, such as genotyping for microsatellite mutations/deletions (TM-GTP), the predictive performance increased from 70 to 85%. Rodriguez-Luna et al. [60] externally validated the ANN/TM-GTP model, and the discriminatory power was 89.5%, while the sample size in the external validation cohort was too small, comprising only 19 patients; therefore, the predictive performance was less convincing. A multicenter study conducted by Nam et al. showed more convincing results [24] because they developed a DNN model and included a relatively large sample size, in which the training cohort was 563 and the validation cohort was 214. Nevertheless, the predictive model should be based not only on the characteristics of receipts but also on donors. Therefore, precise receipt-donor matching is crucial to develop a predictive model. Zhang et al. [61] established an MLP model by including 14 characteristics of donors as well as recipients. The results showed that the c-statistics of the specific MLPs at 1, 2, and 5 years were 0.909, 0.888, and 0.845, respectively. However, the main weakness of this MLP model is the lack of external validation, and the generalization of this model needs to be further confirmed. The relevant papers are listed in Table 3.
| Author          | Study type | No. of patients | Model     | Outcomes    | AUC/C-index | Conclusion                                                                 |
|-----------------|------------|----------------|-----------|-------------|-------------|-----------------------------------------------------------------------------|
| Hamamoto [43], 1995 | Retrospective Single center | 65 | ANN | Death | – | In the study for predicting the died of hepatic dysfunction, ANN predicted the outcome of 11 patients in the validation group and achieved the accuracy of 100% |
| Ho [44], 2012    | Retrospective Multi-center | 427 | ANN and DT | 1,3,5-year DFS | D: 0.977 and 0.734 (1-year) 0.989 and 0.825 (3-year) 0.963 and 0.675 (5-year) V: 0.777 and 0.718 (1-year) 0.774 and 0.561 (3-year) 0.864 and 0.627 (5-year) | The ANN outperforms DT in predicting DFS in post-surgical HCC patients |
| Xu [48], 2012    | Retrospective Multi-center | 336 | SVM | RR | – | The SVM based on IHC features could identify HCC patients who are easily recurrence after surgery, and the predictive accuracy of SVM was 66.5% |
| Chiu [45], 2013  | Retrospective Multi-center | 434 | ANN | 1,3,5-year survival | D: 0.980 (1-year) 0.989 (3-year) 0.993 (5-year) V: 0.875 (1-year) 0.798 (3-year) 0.810 (5-year) | The ANN model can process a greater number of predictors and had better accuracy than the traditional LR model |
| Qiao [46], 2014  | Prospective Multi-center | 725 | ANN | 5-year survival | D: 0.855 V: 0.829 | The ANN model outperforms both Cox and other staging systems in predicting survival in HCC patients who have received surgical resection |
| Cai [36], 2015   | Retrospective Single center | 299 | BN | 10-month survival | – | The BN model had 67.2% of accuracy to classify the survival time of post-surgical HCC patients |
| Akai [49], 2018  | Retrospective Single center | 127 | RSF | DFS, OS | 0.611 0.701 | RSF can predict the individual risk for each patient on DFS and OS |
| Wang [23], 2019  | Retrospective Single center | 167 | DCNN | RR | 0.825 | Combined clinical information and radiomics features can effectively predict early recurrence of HCC patients |
| Author          | Study type     | No. of patients | Model       | Outcomes       | AUC/C-index | Conclusion                                                                 |
|-----------------|----------------|-----------------|-------------|----------------|-------------|----------------------------------------------------------------------------|
| Kim [50], 2019  | Retrospective  | 167             | RSF1*       | Early recurrence | 0.671(RSF1)  | Compared to another two RSF models, combined clinicopathologic-radiomic RSF model achieved the highest predictive power for the recurrence within 2 years after surgery of HCC, and has fair predictive performance for lately recurrence |
|                 | Single center  |                 | RSF2**      | Lately recurrence | 0.679(RSF2)  |                                                                            |
|                 |                |                 | RSF3***     |                | 0.707(RSF3)  |                                                                            |
|                 |                |                 |             |                | 0.737(RSF1)  |                                                                            |
|                 |                |                 |             |                | 0.622(RSF2)  |                                                                            |
|                 |                |                 |             |                | 0.716(RSF3)  |                                                                            |
| Xu [19], 2019   | Retrospective  | 1139            | SVM         | RR             | 0.46         | The accuracy of SVM, RF and BN model was 0.46, 0.48 and 0.56, respectively, in validation group form another independent institution. The BN model could contribute to HCC recurrence research |
|                 | Multi-center   |                 | RF          |                | 0.48         |                                                                            |
|                 |                |                 | BN          |                | 0.56         |                                                                            |
| Mai [47], 2020  | Retrospective  | 353             | ANN         | PHLF           | 0.880(D)     | The risk of severe PHIF in HCC patients after surgery based on ANN model, can be accurately divided into 3 groups |
|                 | Single center  |                 |             |                | 0.876(V)     |                                                                            |
| Saillard [55], 2020 | Retrospective Multi-center | 522         | CNN1#       | OS             | D: 0.75(CNN1) | Two CNN models based on histological features form WSIs performed well for predicting OS of HCC patients after surgery, and both CNN models outperformed the CS that the score included the relevant clinical, biological and pathological features |
|                 |                |                 | CNN2##      |                | 0.78(CNN2)   |                                                                            |
|                 |                |                 |             |                | V: 0.68(CNN1) |                                                                            |
|                 |                |                 |             |                | 0.70(CNN2)   |                                                                            |
| Schoenberg [51], 2020 | Retrospective Single center | 180         | RF          | DFS            | D: 0.766(0.627–0.904) | RF model based on clinical and laboratory variables, can accurately predict DFS after surgery of HCC |
|                 |                |                 |             |                | V: 0.788(0.658–0.919) |                                                                            |
| Wang [52], 2020 | Retrospective Multi-center | 201         | RF          | S-year survival | 0.980        | RAD model integrated with RF in a valid method to predict S-year survival of post-operative HCC patients |
|                 |                |                 |             |                | V: 0.758     |                                                                            |
| Liao [53], 2020 | Retrospective Multi-center | 645         | RF          | 1,3,5-Y survival | V1: 0.626(1-year) | RF model based on 46 histopathological features, was able to stratify post-surgical patients of HCC into long and short-term groups. And the RF model showed similar accuracy with TNM staging systems |
|                 |                |                 |             |                | 0.658(3-year) |                                                                            |
|                 |                |                 |             |                | 0.581(5-year) |                                                                            |
|                 |                |                 |             |                | V2: 0.600(1-year) |                                                                            |
|                 |                |                 |             |                | 0.595(3-year) |                                                                            |
|                 |                |                 |             |                | 0.566(5-year) |                                                                            |
Local ablation

A string of image-guided percutaneous ablations encompasses a great variety of techniques, including radiofrequency ablation (RFA), microwave ablation (MWA), ethanol injection, and cryoablation [62–65]. As RFA is the most frequently used ablation modality for HCC [66, 67], the main topic addressed in this section is RFA. Although RFA has shown good feasibility in local tumor control for HCC, complete ablation is slightly idealistic, and the relapse rate ranges from 49 to 63% [68]. When a recurrence of HCC after ablation arises, the proliferation and invasive ability of tumors are markedly increased.

Very few studies have used ML models to predict therapeutic outcomes in HCC patients in the setting of RFA. In a small sample analysis of 83 HCC patients [69], an SVM model was used to analyze the relationship...
clinical features and early post-RFA recurrence, and the results showed that the model had an AUC of 0.69. However, the predictive performance of the SVM model may decrease when a large number of variables are inputted. Therefore, it is essential to select the variable prior to establishing the SVM model. Conversely, the ANN algorithm cannot be used to select variables, while the key advantage of the ANN model is that it can process data with a large number of variables and samples. In the study of Wu et al. [70], a total of 15 variables were inputted into two ANN models for the prediction of 1-year disease-free survival (DFS) and 2-year DFS, and the performances of these two models were both excellent, with AUCs of 0.964 and 0.974, respectively. Unfortunately, both SVM and ANN models are immature because they lack external validation. The relevant papers are listed in Table 4.

TACE
Most HCC patients are typically diagnosed at intermediate or advanced stages when curative treatments cannot be applied [71, 72]. According to the Barcelona Clinic Liver Cancer (BCLC) staging system [73] and several treatment guidelines [6, 7], TACE is the gold standard for patients with intermediate-stage HCC. Since not all HCC patients can benefit from TACE [74, 75], a predictive model providing therapeutic outcome estimation prior to the procedure is urgently needed for clinical decision making.

Previous studies have constructed a variety of ML models with clinical and radiological variables for predicting the therapeutic outcome of HCC patients after TACE [25, 76–79]. Mähringer-Kunz et al. [76] used traditional imaging features, such as tumor size and tumor number, and other clinical variables to create an ANN model for predicting 1-year survival after TACE. Further, the results demonstrated that the predictive performance of this model was 0.77 in the training cohort and 0.83 in the validation cohort. However, imaging features are not always visible to the naked eye, and some tiny imaging features may be overlooked. Radiomics is an emerging discipline that can extract invisible imaging features such as statistic, shape and texture features from medical images. Abajian et al. [77] established an RF model and used semiautomatic 3D tumor segmenting software to extract several statistic and shape features from MR imaging. The results of their study demonstrated that the most valuable predictor of treatment response following TACE was relative tumor signal intensity on pre-TACE MR images, and the highest predictive accuracy of the RF model achieved 78%. However, the quantitative imaging features in Abajian’s study are too simple and cannot provide adequate information to predict the therapeutic outcome. Liu et al. [78] developed an SVM model with complex radiomic features. These complex radiomic features were first extracted by manual segmentation from static B-mode images, which included 181 statistic features, 13 tumor shape features, and 740 texture features. After extraction, the meaningful radiomic features were selected by the gradient boosted regression trees (GBRT) algorithm [78], and finally, the SVM model was established with an AUC of 0.81 in the internal validation cohort. Indeed, radiomic features can not only be extracted by manual or semiautomatic segmentation tools but can also be extracted automatically by CNN algorithms [25, 78, 79]. Morshid et al. [79] used a CNN-based segmentation protocol to extract a large number of shape and texture features from portal venous phase CT images. Based on these imaging features, an RF model was established, and the results showed that the RF model could accurately distinguish TACE-refractory patients with an AUC of 0.7331 [79]. In addition to extracting imaging features, the CNN algorithm can also be used to establish the predictive model [25, 78]. Peng et al. [25] used the CNN algorithm to automatically extract the imaging features of HCC from CT images and established a predictive model of tumor response after TACE. Their study showed that the CNN models could

Table 4 Characteristics of ML-based predictive model after RFA

| Author            | Study type          | No. of patients | Modality | Model   | Outcomes          | AUC     | Conclusion                                                                 |
|-------------------|---------------------|-----------------|----------|---------|-------------------|---------|-----------------------------------------------------------------------------|
| Liang [69], 2014  | Retrospective Single center | 83              | US guided | SVM     | RR                | 0.69    | The SA+ RF SVM method had the best accuracy for predicting high-risk recurrent patients |
| Wu [70], 2017     | Retrospective Single center | 431             | CT guided | MLP     | 1,2-year DFS      | D: 0.94 (1-year) 0.88 (2-year) V: 0.77 (1-year) 0.72 (2-year) | The MLP-based model with 15 clinical HCC relevant features achieved satisfactory predictive performance for 1-year DFS |

ML, machine learning; RFA, radiofrequency ablation; AUC, area under the curve; US, ultrasound; SVM, support vector machine; RR, recurrence rate; SA, simulated annealing algorithm; RF, random forest; CT, computed tomography; MLP, multilayer perceptron; DFS, disease-free survival; D, development cohort; V, validation cohort.
predict the complete response (CR), partial response (PR), stable disease (SD) and progressive disease (PD) of HCC lesions with AUCs of 0.97, 0.96, 0.95, and 0.96, respectively. Similarly, Liu et al. [78] developed a CNN model to extract imaging features from dynamic contrast-enhanced ultrasound (CEUS) images and predict the objective response of 130 HCC patients after TACE with an AUC of 0.93. The relevant papers are listed in Table 5.

**Sorafenib**

Sorafenib is the standard treatment for advanced-stage HCC. The median OS of sorafenib-treated HCC was 10.7 months and 6.5 months in two previous representative randomized controlled trials [80, 81]. Because of the high cost and modest efficacy, a reliable predictive tool is necessary to assist clinicians in adjusting the daily management of sorafenib for such patients. ML methods are not routinely used for predicting therapeutic outcomes in the treatment of sorafenib for HCC. Choi et al. [22] collected clinical and radiological data from 480 sorafenib-treated patients, and the important variable scores were used to select final parameters based on the RF algorithm. They found that the established model had a better predictive performance in time to progression (TTP) and overall survival (OS) than those of the Child–Pugh and Model for End-Stage Liver Disease (MELD) scores (0.746 vs 0.686 and 0.545 for TTP, 0.875 vs 0.777 and 0.682 for OS). However, this study lacks independent external validation. The relevant papers are listed in Table 6.

| Author                  | Study type       | No. of patients | Model | Outcomes                  | AUC         | Conclusion                                                                 |
|-------------------------|------------------|-----------------|-------|---------------------------|-------------|---------------------------------------------------------------------------|
| Abajian [77], 2018      | Retrospective    | 36              | RF    | Responders or non-responders | –           | RF model combined with MRI parameters may be predicted tumor response of post-TACE HCC |
| Morshid [79], 2019      | Retrospective    | 105             | RF    | TACE-susceptible or TACE-refractory | 0.733       | The accuracy of RF model using a combination of clinical parameters plus quantitative image features was higher than the RF model based on the clinical parameters alone, in the study of predicting HCC response to TACE |
| Mähringer-Kunz [76], 2020 | Retrospective    | 282             | ANN   | 1-year survival           | V: 0.77±0.13 | The ANN model had a promising performance at predicting HCC patient survival after TACE and outperformed the traditional scoring systems |
| Peng [25], 2020         | Retrospective    | 798             | CNN   | CR, PR, SD, PD            | D: 0.97 (CR) | The CNN model presented a good performance for predicting the outcome of TACE |
| Liu [78], 2020          | Retrospective    | 138             | CNN   | ORR                       | D: 0.98 (CNN) | CNN is better in predicting treatment response over SVM in HCC patients treated with TACE |

* SVM1: radiomics-based time-intensity curve of CEUS model using SVM; SVM2: radiomics-based B-Mode images model using SVM

ML: machine learning, HCC: hepatocellular carcinoma, TACE: transarterial chemoembolization, AUC: area under the curve, RF: random forest, MRI: magnetic resonance imaging, ANN: artificial neural network, D: development cohort, V: validation cohort, CNN: convolutional neural network, CR: complete response, PR: partial response, SD: stable disease, PD: progressive disease, SVM: support vector machine, ORR: objective response rate, CEUS: contrast-enhanced ultrasound.
Future perspectives in ML for the prognostic study of HCC

Currently, the ML model for predicting the therapeutic outcome of HCC is usually based on multivariate predictors, such as demographic, clinical, radiological, pathologic and genetic parameters. Selecting the final predictors is a considerable challenge in traditional statistical models because traditional statistical methods may lose some important information. The ML model can include more variables, and it may become a promising protocol over the traditional statistical model. In addition, the ML algorithm can extract and select radiomic features that are invisible to the naked eye, and those novel variables may provide promising predictive value compared with simple radiological parameters (tumor size and tumor number, etc.).

The most important challenge in the ML approach is the accurate selection of algorithms to create the predictive model with external validation for the model. On the one hand, certain types of ML models are favored for specific types of data, such as CNNs for imaging data and SVMs for small sample size data. The ML model should be selected by a thoughtful study design. On the other hand, as there is a need for clinical reality in the future, appropriate external validation should be used to confirm the generalization ability. Due to the lack of a commonly accepted design of ML predictive models for the prognostic study of HCC, it may be possible that the current ML model is not the best one available.

Conclusion

ML algorithms can automatically extract imaging features and identify optimal subsets of features from large data sets, particularly when combined with radiomics analysis. Relative to traditional statistical models, ML models demonstrate improved predictive performance in the prognostic study of HCC. Regrettably, most existing ML predictive models lack external validation, which is an obstacle to serving HCC patients as personalized predictive tools. Although most current ML algorithms are preliminary, this promising method will be widely accepted in clinical practice in the future.
11. Schobert IT, Savic LJ, Chapiro J et al (2020) Neutrophil-to-lymphocyte ratios as predictors of tumor response in hepatocellular carcinoma after DEB-TACE. Eur Radiol 30:5663–5673

12. Nam JY, Choie AR, Sinn DH et al (2020) A differential risk assessment and decision model for Transarterial chemoembolization in hepatocellular carcinoma based on hepatic function. BMC Cancer 20:504

13. Erickson BJ, Korfiatis P, Akkus Z, Kline TL (2017) Machine learning for decision model for Transarterial chemoembolization in hepatocellular carcinoma. J Hepatol 69:182–236

14. Heimbach JK, Kulik LM, Finn RS et al (2018) AASLD guidelines for the management of hepatocellular carcinoma. J Hepatol 69:182–236

15. Omata M, Cheng AL, Kokudo N et al (2017) Asia-Pacific clinical practice update. Hepatol Int 11:317–370

16. Wakabayashi T, Ouhmich F, Gonzalez-Cabrera C et al (2019) Radiomics in hepatocellular carcinoma recurrences: a data-driven multiclass classification method proposed by the Japan Society of Hepatology (JSH) 2010 update. Dig Dis 29:339–364

17. Santos MS, Abreu PH, García-Laencina PJ, Simão A, Carvalho A (2015) A new cluster-based oversampling method for improving survival prediction of hepatocellular carcinoma patients. J Biomed Inform 58:49–59

18. Divya R, Radha P (2019) An optimized HCC recurrence prediction using APO algorithm multiple time series clinical liver cancer dataset. J Med Syst 43:193

19. Xu D, Sheng JQ, Hu PJ, Huang TS, Lee WC (2019) Predicting hepatocellular carcinoma recurrences: a data-driven multiclass classification method incorporating latent variables. J Biomed Inform 96:103257

20. Liu X, Hou Y, Wang X et al (2020) Machine learning-based development and validation of a scoring system for progression-free survival in liver cancer. Hepatol Int 14:567–576

21. Ji GW, Zhu FP, Xu Q et al (2019) Machine-learning analysis of contrast-enhanced CT radiomics predicts recurrence of hepatocellular carcinoma after resection. J Biomed Imaging 25:115–165

22. Choi GH, Han S, Shim JH et al (2017) Prognostic scoring models for patients undergoing sorafenib treatment for advanced stage hepatocellular carcinoma in real-life practice. Am J Clin Oncol 40:167–174

23. Wang W, Chen Q, Iwamoto Y et al (2019) Deep learning-based radiomics models for early recurrence prediction of hepatocellular carcinoma with multi-phase CT images and clinical data. Ann Int Conf IEEE Eng Med Biol Soc 2019:4881–4884

24. Nam JY, Lee JH, Bae J et al (2020) Novel model to predict HCC recurrence after liver transplantation obtained using deep learning: a multicenter study. Cancers (Basel). 12,E2791

25. Peng J, Kang S, Ning Z et al (2020) Residual convoluted neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging. Eur Radiol 30:413–424

26. Cross SS, Harrison RF, Kennedy RL (1995) Introduction to neural networks. Lancer 546:1075–1079

27. Lee JG, Jun S, Cho YW et al (2017) Deep learning in medical imaging: general overview. Korean J Radiol 18:570–584

28. Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK (2018) Medical image analysis using convolutional neural networks: a review. J Med Syst. 42:226

29. Breher R, Mitrean DA, Vanea C et al (2020) Comparison of deep-learning and conventional machine-learning methods for the automatic recognition of the hepatocellular carcinoma areas from ultrasound images. Sensors (Basel). 20:3085

30. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20:273–297

31. Podgorver V, Kokol P, Stiglic B, Rozman I (2002) Decision trees: an overview and their use in medicine. J Med Syst 26:445–463

32. Breiman L (2001) Random forests. Mach Learn 45:5–32

33. Uddin S, Khan A, Hossain ME, Moni MA (2019) Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 19:281

34. Ishiwara H, Gerds TA, Kagubare UB, Moore RD, Gange SJ, Lau BM (2014) Random survival forests for competing risks. Biostatistics 15:757–773

35. Weber P, Medina-Oliva G, Simon C, Lung B (2012) Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas. Eng Appl Artif Intell 25:671–682

36. Cai ZQ, Si SB, Chen C et al (2015) Analysis of prognostic factors for survival after hepatectomy for hepatocellular carcinoma based on a Bayesian network. PLoS ONE 10:e0120805

37. Keating GM (2017) Sorafenib: a review in hepatocellular carcinoma. Target Oncol 12:243–253

38. Berumen J, Hemming A (2018) Liver transplantation for hepatocellular carcinoma. Abdom Radiol (NY) 43:185–192

39. Narsinh KH, Duncan DP, Newton KG, Minincha J, Rose SC (2018) Liver-directed therapy for hepatocellular carcinoma. Abdom Radiol (NY) 43:203–217

40. Langenbach MC (2019) RFA vs resection of HCC: exploring the past to improve the future. Eur Radiol 29:2677–2678

41. Bruix J, Reig M, Sherman M (2016) Evidence-based diagnosis, staging, and treatment of patients with hepatocellular carcinoma. Gastroenterology 150:835–853

42. Kulik L, El-Seraq HB (2019) Epidemiology and management of hepatocellular carcinoma. Gastroenterology 156:477–491

43. Hamamoto I, Okada S, Hashimoto T, Wakabayashi H, Maeba T, Maeta H (2016) Evidence-based diagnosis, staging, and conventional machine-learning methods for the automatic prediction of the early prognosis of the hepatectomized patient. J Hepatol 65:1–15

44. Cai ZQ, Si SB, Chen C et al (2015) Analysis of prognostic factors for survival after hepatectomy for hepatocellular carcinoma based on a Bayesian network. PLoS ONE 10:e0120805

45. Chiu HC, Ho TW, Lee KT, Chen HY, Ho WH (2013) Mortality predicted accuracy for hepatocellular carcinoma patients with hepatic resection using artificial neural network. Sci World J 2013:201976

46. Qiao G, Li J, Huang A, Yan Z, Lau WY, Shen F (2014) Artificial neural networking model for the prediction of post-hepatectomy survival of patients with early hepatocellular carcinoma. J Biomed Imaging 20:273–297

47. Xu J, Ding T, He Q et al (2012) An in situ molecular signature to predict disease-free survival after hepatic resection in hepatocellular carcinoma patients: a prediction approach using artificial neural network. PLoS ONE 7:e29179

48. Chiu HC, Ho TW, Lee KT, Chen HY, Ho WH (2013) Mortality predicted accuracy for hepatocellular carcinoma patients with hepatic resection using artificial neural network. Sci World J 2013:201976

49. Ohrui T, Ishigaki M, Hasegawa K et al (2019) Artificial neural network model for the prediction of post-hepatectomy survival of patients with early hepatocellular carcinoma. J Gastroenterol Hepatol 29:2014–2020

50. Mai RY, Lu HZ, Bai T et al (2020) Artificial neural network model for preoperative prediction of severe liver failure after hemihepatectomy in patients with hepatocellular carcinoma. Surgery 168:643–652

51. Xu J, Ding T, He Q et al (2012) An in situ molecular signature to predict early recurrence in hepatitis B virus-related hepatocellular carcinoma. J Gastroenterol Hepatol 57:313–321

52. Aka H, Yasaka K, Kunimatsu A et al (2018) Predicting prognosis of resected hepatocellular carcinoma by radiomics analysis with random survival forest. Diagn Interv Imaging 99:643–651

53. Kim S, Shin J, Kim DY, Choi GH, Kim MJ, Choi JY (2019) Radiomics on gadoteric acid-enhanced magnetic resonance imaging for prediction of postoperative early and late recurrence of single hepatocellular carcinoma. Clin Cancer Res 25:3847–3855

54. Schoenber MB, Bucher JN, Koch D et al (2020) A novel machine learning algorithm to predict disease free survival after resection of hepatocellular carcinoma. Ann Transl Med 8:434

55. Wang XH, Long LH, Cui Y et al (2020) MRI-based radiomics model for preoperative prediction of 5-year survival in patients with hepatocellular carcinoma. Br J Cancer 122:978–985

56. Liao H, Xiong T, Peng J et al (2020) Classification and prognosis prediction from histopathological images of hepatocellular carcinoma by a fully automated pipeline based on machine learning. Ann Surg Oncol 27:2359–2369

57. Saito A, Toyota H, Kobayashi M et al (2020) Prediction of early recurrence of hepatocellular carcinoma after resection using digital pathology
images assessed by machine learning [published online ahead of print, 2020 Sep 18]. Mod Pathol. https://doi.org/10.1038/s41379-020-00671-z
55. Saillard C, Schmauch B, Lafia O et al (2020) Predicting survival after hepatocellular carcinoma resection using deep-learning on histological slides [published online ahead of print, 2020 Feb 28]. Hepatology. https://doi.org/10.1002/hep.31207
56. Sapisochin G, Bruix J (2017) Liver transplantation for hepatocellular carcinoma: outcomes and novel surgical approaches. Nat Rev Gastroenterol Hepatol 14:203–217
57. Kim B, Kahn J, Terrault NA (2020) Liver transplantation as therapy for hepatocellular carcinoma. Liver Int 40(Suppl 1):116–121
58. Marsh JW, Dvorchik I, Subotin M et al (1997) The prediction of risk of recurrence and time to recurrence of hepatocellular carcinoma after orthotopic liver transplantation: a pilot study. Hepatology 26:444–450
59. Marsh JW, Finkelstein SD, Demetris AJ et al (2003) Genotyping of hepatocellular carcinoma in liver transplant recipients adds predictive power for determining recurrence-free survival. Liver Transpl 9:664–671
60. Rodriguez-Luna H, Vargas HE, Byrne T, Rakela J (2005) Artificial neural network and tissue genotyping of hepatocellular carcinoma in liver-transplant recipients: prediction of recurrence. Transplantation 79:1737–1740
61. Zhang M, Yin F, Chen B et al (2012) Mortality risk after liver transplantation in hepatocellular carcinoma recipients: a nonlinear predictive model. Surgery 151:889–897
62. Ren Y, Cao Y, Ma H et al (2019) Improved clinical outcome using transarterial chemoembolization combined with radiofrequency ablation for patients in Barcelona clinic liver cancer stage A or B hepatocellular carcinoma regardless of tumor size: results of a single-center retrospective case control study. BMC Cancer 19:983
63. Liu B, Long J, Wang W et al (2019) Predictive factors of treatment outcomes after percutaneous ablation of hepatocellular carcinoma in the caudate lobe: a retrospective study. BMC Cancer 19:699
64. Kim R, Kang TW, Cha DI et al (2019) Percutaneous cryoablation for perivascular hepatocellular carcinoma: therapeutic efficacy and vascular complications. Eur Radiol 29:654–662
65. Chai Y, Li K, Zhang C, Chen S, Ma K (2019) The short-term efficacy of no-touch radiofrequency ablation in treating cirrhosis-based small hepatocellular carcinoma. BMC Cancer 19:497
66. Forner A, Reig M, Bruix J (2018) Hepatocellular carcinoma. Lancet 391:1301–1314
67. Lee DH, Lee JM, Kim PN et al (2019) Whole tumor ablation of locally recurred hepatocellular carcinoma including retained iodized oil after transarterial chemoembolization improves progression-free survival. Eur Radiol 29:5052–5062
68. Nault JC, Sutter O, Nahon P, Ganne-Carrié N, Séror O (2018) Percutaneous treatment of hepatocellular carcinoma: state of the art and innovations. J Hepatol 68:783–797
69. Liang JD, Ping XO, Tseng YJ, Huang GT, Lai F, Yang PM (2014) Recurrence predictive models for patients with hepatocellular carcinoma after radiofrequency ablation using support vector machines with feature selection methods. Comput Methods Programs Biomed 117:425–434
70. Wu CF, Wu YJ, Liang PC, Wu CH, Peng SF, Chiu HW (2017) Disease-free survival assessment by artificial neural networks for hepatocellular carcinoma patients after radiofrequency ablation. J Formos Med Assoc 116:765–773
71. Zhang ZS, Li HZ, Ma C, Xiao YD (2019) Conventional versus drug-eluting beads chemoembolization for infiltrative hepatocellular carcinoma: a comparison of efficacy and safety. BMC Cancer 19:1162
72. Liang B, Xiang H, Ma C et al (2020) Comparison of chemoembolization with CalliSpheres® microspheres and conventional chemoembolization in the treatment of hepatocellular carcinoma: a multicenter retrospective study. Cancer Manag Res 12:941–956
73. Llovet JM, Brú C, Bruix J (1999) Prognosis of hepatocellular carcinoma: the BCLC staging classification. Semin Liver Dis 19:329–338
74. Sieghart W, Huckle F, Peck-Radosavljevic M (2015) Transarterial chemoembolization: modalities, indication, and patient selection. J Hepatol 62:1187–1195
75. Lencioni R, de Baere T, Soulen MC, Rilling WS, Geschwind JF (2016) Lipiodol transarterial chemoembolization for hepatocellular carcinoma: a systematic review of efficacy and safety data. Hepatology 64:106–116
76. Mähringer-Kunz A, Wagner F, Hahn F et al (2020) Predicting survival after transarterial chemoembolization for hepatocellular carcinoma using a neural network: a pilot study. Liver Int 40:694–703
77. Abajian A, Murali N, Savic LJ et al (2018) Predicting treatment response to intra-arterial therapies for hepatocellular carcinoma with the use of supervised machine learning an artificial intelligence concept. J Vasc Interv Radiol 29:850–857
78. Liu D, Liu F, Xie X et al (2020) Accurate prediction of responses to transarterial chemoembolization for patients with hepatocellular carcinoma by using artificial intelligence in contrast-enhanced ultrasound. Eur Radiol 30:2365–2376
79. Morshid A, Elsayes KM, Khalaf AM et al (2019) A machine learning model to predict hepatocellular carcinoma response to transcatheter arterial chemoembolization. Radiol Artif Intell 1:e180021
80. Faivre S, Raymond E, Boucher E et al (2009) Safety and efficacy of sunitinib in patients with advanced hepatocellular carcinoma: a multicentre, phase II study. Lancet Oncol 10:794–800
81. Cheng AL, Kang YK, Chen Z et al (2009) Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. Lancet Oncol 10:25–34

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.