A Hybrid Machine Learning Model Based on Semantic Information Can Optimize Treatment Decision for Naïve Single 3–5-cm HCC Patients

Wenzhen Ding  Zhen Wang  Fang-Yi Liu  Zhi-Gang Cheng  Xiaoling Yu  Zhiyu Han  Hui Zhong  Jie Yu  Ping Liang

Department of Interventional Ultrasound, The First Center of Chinese PLA General Hospital, Beijing, PR China

Keywords
Hepatocellular carcinoma · Laparoscopic hepatectomy · Microwave ablation · Treatment decision

Abstract
Background: Tumor recurrence is an abomination for hepatocellular carcinoma (HCC) patients receiving local treatment. Purpose: The aim of the study was to build a hybrid machine learning model to recommend optimized first treatment (laparoscopic hepatectomy [LH] or microwave ablation [MWA]) for naïve single 3–5-cm HCC patients based on early recurrence (ER, ≤2 years) probability. Methods: This retrospective study collected 20 semantic variables of 582 patients (LH: 300, MWA: 282) from 13 hospitals with at least 24 months follow-up. Both groups were divided into training, validation, and test set, respectively. Five algorithms (logistic regression, random forest, neural network, stochastic gradient boosting, and eXtreme Gradient Boosting [XGB]) were used for model building. A model with highest area under the receiver operating characteristic curve (AUC) in a validation set of LH and MWA was selected to connect as a hybrid model which made decision based on ER probability. In the comprehensive set, a treatment confusion matrix was established based on recommended and actual treatment. The predicted ER probabilities were comparable with the actual ER rates for various types of patients in matrix (p > 0.05). ER rate of patients whose actual treatment was consistent with recommendation was lower than that of inconsistent patients (LH: 21.2% vs. 46.2%, p = 0.042; MWA: 26.3% vs. 54.1%, p = 0.048). By recommending optimal treatment, the hybrid model can significantly reduce ER probability from 38.2% to 25.6% for overall patients (p < 0.001). Conclusions: The hybrid model can accurately predict ER probability of different treatments and thereby provide reliable evidence to make optimal treatment decision for patients with single 3–5-cm HCC.
patient’s treatment confidence. The overall survival rate of patients with ER is significantly less than that of patients without ER. Appropriate treatment can significantly reduce ER and prevent patients from falling into the deadly cycle of treatment-recurrence-retreatment-recurrence [2–4]. Therefore, optimizing treatment decision-making for HCC patients is an urgent clinical problem.

With increasing requirements for and interest in minimally invasive treatments, laparoscopic hepatectomy (LH) has gradually become mainstream in the field of surgical resection [5, 6]. Advances in medical equipment have also made microwave ablation (MWA) become the latest and most effective technique in the field of thermal ablation [7, 8]. For very early-stage HCC patients, the newest guidelines and several studies have demonstrated that there is no significant difference in therapeutic effect between these two treatment methods. However, for 3–5-cm HCC, both treatments can offer their own unique advantages; thereby, the decision of optimal treatment has a crucial influence on reducing ER. When clinicians are faced with the question about which treatment (LH or MWA) could maximize the effective ER-free benefit to patients, access to an effective and reliable treatment decision-making tool would be invaluable [1, 9].

The rapid development of artificial intelligence and its wide application in medical world give rich and powerful weapons for clinicians when they face problems such as diagnosing disease or making treatment decision [10–12]. Machine learning based on the new advanced algorithm can break the chain of traditional generalized linear model, they are not only more flexible in mining data value, but also more acceptable in data integrity and data distribution. By taking the full advantage of these features, machine learning could achieve better predictive ability and adaptability than the traditional model. Therefore, the purpose of our study is to build a hybrid machine learning model to recommend optimal first treatment for naïve single 3–5-cm HCC patients.

Materials and Methods

Patients
This retrospective multicenter study included patients with HCC from 13 hospitals who were treated with ultrasound-guided MWA or LH from 2006 to 2018, and this study was approved by the Ethics Committees of 13 hospitals (name of hospitals in online suppl. material; for all online suppl. material, see www.karger.com/doi/10.1159/000522123). Inclusion criteria were as follows: (1) diagnosis of HCC with pathological confirmation; (2) no treatment history; (3) 3–5-cm single tumor; (4) more than 24 months follow-up; (5) for MWA group, lesions were clearly visible on ultrasound or contrast-enhanced ultrasound. Exclusion criteria were as follows: (1) obvious extrahepatic metastasis on ultrasound, computed tomography, or magnetic resonance imaging; (2) palliative treatment; (3) Child-Pugh class C; (4) infiltrative HCC (Fig. 1).

Data Collection
Clinical and laboratory parameters of patients were retrieved from 13 hospital databases. Information includes 4 aspects and a total of 20 variables: (1) basic condition: age, gender, body mass index, Charlson comorbidity index [13]; (2) the condition of liver disease: the Child-Pugh stage, liver cirrhosis, HBV, HCV; (3) laboratory parameters: platelet, alanine aminotransferase, aspartate aminotransferase, total protein, serum albumin (ALB), total bilirubin, direct bilirubin, γ-glutamyl transpeptidase, prothrombin activity, international normalized ratio, alpha fetoprotein (AFP); (4) the parameters of tumor: tumor size. Finally, we observed and recorded whether patients occurred tumor recurrence during the 24 months following treatment (explanation of variables in online suppl. material).

Study Design
The study design flowchart is shown in Figure 1. About 20% of patients in the LH group and MWA group were randomly selected and combined into a comprehensive test set. The remaining patients in 2 groups were divided into LH training set, LH validation set and MWA validation set, respectively. In the respective training sets, ER-related variables were selected from 20 semantic information. Multiple models were built through logistics regression (LR), random forest (RF), neural network (NNT), stochastic gradient boosting (SGB), and eXtreme Gradient Boosting (XGB) to predict ER after the first treatment. A model with the highest area under the receiver operating characteristic curve (AUC) in each validation set (Best LH-model and Best MWA model) would be connected in parallel to build the hybrid model. The hybrid model would be used to predict the ER probability of taking LH as the first treatment and the ER probability of taking MWA as the first treatment for every patient in comprehensive test set. If the patient’s ER probability of MWA was higher than that of LH, he would be recommended LH as the first treatment. Otherwise, he would be recommended MWA as the first treatment.

Therapy Process
The processes of MWA and LH were carried out in accordance with previous research [14, 15]. The detailed process was in online supplementary material.

Follow-Up
All patients underwent regular follow-up assessments according to hospital practice. The follow-up included routine outpatient physical examinations, serum alpha fetoprotein, and contrast-enhanced US or CT/MRI at 1 and 3 months after treatment and then at 6-month intervals. If any examination was abnormal, a variety of imaging would be taken to confirm whether the tumor had recurred.

Variable Selection and Model Building
In the training set, Student’s t test, χ² tests, or Fisher’s exact test was used to compare continuous and categorical variables between ER patients and without ER patients to determine the variables that

DOI: 10.1159/000522123

Liver Cancer 2022;11:256–267

257
influenced ER. Variables with \( p < 0.1 \) would be included into model building. Ten-fold cross-validation was used in the model training process. For all algorithms, model performance was primarily evaluated by the AUC. Prediction accuracy was also recorded as basis for adjusting the weight of each variable and main parameters of model. The initial parameters of each model were set according to default values, and then main parameters, such as iteration, learning rate, and maximum depth, were adjusted through grid search with cross-validation according to the AUC and accuracy.

The algorithms used in model building include LR, RF, NNT, SGB, and XGB. LR is a classic binary classifier that fits a linear combination of original features onto a sigmoid curve and then outputs the probabilities for each class. NNET is a computing model composed of a large number of interconnected neurons. Each neuron represents an activation function, and each connection between two neurons represents the weight of the signal passing through the connection. The output is depending on the connection mode of network, the weight, and the activation function. RF is an ensemble learning algorithm based on multiple decision trees that use random subsets of features and training samples to independently perform binary predictions. Gradient boosting constructs additive regression models by sequentially fitting a simple parameterized function to current "pseudo"-residuals by least squares at each iteration. SGB can substantially improve both the approximation accuracy and execution speed of gradient boosting by incorporating randomization into the procedure. Based on gradient boosting, XGB can take the difference value of the second-order Taylor formula of loss function as learning objective and optimize the regularization technology to improve efficiency and reduce overfitting. All of these algorithms were implemented in R software (version 4.0.0) using the “caret” package.

### Inclusion Criteria:
1. Diagnosis of HCC with pathological confirmation
2. No treatment history
3. 3–5 cm single tumor
4. More than 24 months follow-up
5. For MWA group, lesions were clearly visible on ultrasound or contrast-enhanced ultrasound

### Exclusion Criteria:
1. Extrahepatic metastasis on ultrasound, CT and MRI
2. Palliative treatment
3. Child-Pugh class C
4. Infiltrative CC

### Fig. 1. Flowchart. HCC, hepatocellular carcinoma; MWA, microwave ablation; LH, laparoscopic hepatectomy; RF, random forest; NNT, neural network; SGB, stochastic gradient boosting; XGB, eXtreme gradient boosting.

Continuous data were expressed as means ± SD and evaluated by the Student’s \( t \) test. Categorical variables were analyzed using \( \chi^2 \) tests or Fisher’s exact test. One-sample \( t \) test was used to compare the actual ER rate with the predicted ER probability. Paired sample \( t \) test was used to compare the predicted ER probability of actual treatment with that of recommended treatment. The data segmentation, statistical analysis, and model building were completed by using the “caret” package and “tableone” package in R software (version 4.0.0), and SPSS (version 26.0).
Results

Patient Baseline

A total of 582 eligible patients (LH group, n = 300; MWA group, n = 282; Table 1) enrolled for our study. The baseline and ER% of training set, verification set, and test set were comparable in each group (online suppl. Table 1, 2). Pairwise comparisons were shown in online supplementary Tables 3–6. A total of 115 patients were included in the comprehensive test set (LH group, n = 59; MWA group, n = 56; Table 2).

Variable Selection

Among the 20 semantic variables, 4 variables in LH group were selected as meaningful variable of ER, including ALB (p = 0.002), total bilirubin (p = 0.079), AFP (p = 0.013), and tumor size (p = 0.004). These variables were included in the process of building LH model. Similarly,
4 variables in MWA group were selected as meaningful variable of ER, including alanine aminotransferase ($p = 0.002$), total protein ($p = 0.001$), ALB ($p = 0.025$), and AFP ($p = 0.001$). These variables were included in the process of building MWA model (Table 3).

**Model Building**

In training process, LR, RF, NNT, SGB, and XGB were used to build multiple models, and the performance of each model was evaluated by AUC in validation set. By comparing the AUC of five models in the LH validation set, the XGB model had the best performance with AUC = 0.744 (Fig. 2). In the MWA group, the SGB model had the best performance with AUC = 0.750 (Fig. 2). Therefore, the LH-XGB model and the MWA-SGB model were connected in parallel to build the hybrid model.

Table 2. Baseline of test set

| Parameter                  | LH (n = 59)   | MWA (n = 56)  | $p$ value |
|----------------------------|--------------|--------------|-----------|
| Basic condition            |              |              |           |
| Age, year                  | 56.10±12.32  | 58.98±11.20  | 0.193     |
| Sex                        |              |              |           |
| Female                     | 10 (16.9)    | 11 (19.6)    | 0.895     |
| Male                       | 49 (83.1)    | 45 (80.4)    |           |
| BMI                        | 25.24±3.45   | 24.78±3.15   | 0.46      |
| CCI                        |              |              |           |
| 0                          | 45 (76.3)    | 36 (64.3)    |           |
| 1                          | 12 (20.3)    | 12 (21.4)    |           |
| 2                          | 1 (1.7)      | 5 (8.9)      | 0.295     |
| 3                          | 1 (1.7)      | 2 (3.6)      |           |
| 4                          | 0 (0.0)      | 1 (1.8)      |           |
| The condition of liver disease |          |              |           |
| The Child-Pugh             |              |              |           |
| A                          | 58 (98.3)    | 54 (96.4)    | 0.963     |
| 1                          | 1 (1.7)      | 2 (3.6)      |           |
| Liver cirrhosis            | 47 (79.7)    | 46 (82.1)    | 0.92      |
| HBV                        | 36 (61.0)    | 39 (69.6)    | 0.438     |
| HCV                        | 11 (18.6)    | 5 (8.9)      | 0.217     |
| Laboratory parameters      |              |              |           |
| PLT, 10^9/L                | 177.02±55.97 | 139.38±63.57 | 0.001     |
| ALT, U/L                   | 34.19±24.35  | 31.15±18.16  | 0.452     |
| AST, U/L                   | 26.76±13.54  | 26.68±12.53  | 0.974     |
| TP, g/L                    | 68.81±6.37   | 64.88±5.55   | 0.001     |
| ALB, g/L                   | 39.69±3.91   | 39.00±4.81   | 0.401     |
| TB, μmol/L                 | 17.73±10.33  | 18.31±4.92   | 0.708     |
| DB, μmol/L                 | 6.49±3.22    | 6.40±3.59    | 0.888     |
| γ-GT, U/L                  | 62.07±38.99  | 52.21±38.65  | 0.176     |
| PTA                        | 90.01±12.76  | 88.34±15.35  | 0.525     |
| INR                        | 1.15±0.10    | 1.10±0.13    | 0.007     |
| AFP                        |              |              |           |
| ≤20                        | 38 (64.4)    | 35 (62.5)    | 0.985     |
| >20                        | 21 (35.6)    | 21 (37.5)    |           |
| The condition of tumor     |              |              |           |
| Tumor size, cm             | 3.79±0.62    | 3.66±0.57    | 0.253     |
| ER                         | 19 (32.2)    | 25 (44.6)    | 0.238     |

Values are mean±standard deviation or number (%). BMI, body mass index; CCI, Charlson comorbidity index; PLT, platelets; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TP, total protein; ALB, serum albumin; TB, total bilirubin; DB, direct bilirubin; γ-GT, γ-glutamyl transpeptidase; PTA, prothrombin activity; INR, international normalized ratio; AFP, alpha fetoprotein; ER, early recurrence.
Predictive Performance of the Hybrid Model

In test set, 19 of 59 LH patients occurred ER. The ER rate was 32.2%, which was no different from the ER rate in all LH cases (96/300, ER rate = 32.0%, \( p = 0.991 \)). Twenty-five of fifty-six MWA patients occurred ER. The ER rate was 44.6%, which was also no different from the ER rate in all MWA cases (125/282, ER rate = 44.3%, \( p = 0.874 \)).

Then, we simulated a clinical scenario, which assumed that all patients in test set were in an untreated state. When a hypothetical untreated patient needed to make treatment decision, the hybrid model would be used to predict the ER probability of taking LH or MWA as first treatment for the patient. And the treatment with lower ER probability would be recommended. Therefore, patients were divided into recommended LH group and recommended MWA group. Then we could establish the treatment confusion matrix according to the actual and recommended treatment.

Among the 59 patients who actually received LH, 33 patients were recommended for taking LH as the optimal first treatment by the hybrid model (actually (a)LH-recommended (r)LH group). Seven of 33 patients occurred ER, the ER rate was 21.2%. The other 26 patients were
recommended for taking MWA as the optimal first treatment (aLH-rMWA group). Twelve of 26 patients occurred ER, the ER rate was 46.2%, which was significantly higher than the ER rate of the aLH-rLH group (p = 0.042). Among the 56 patients who actually received MWA, 19 patients were recommended for taking MWA as the optimal first treatment (aMWA-rMWA group). Five of 19 patients occurred ER, the ER rate was 26.3%. The other 37 patients were recommended for taking LH as the optimal first treatment (aMWA-rLH group), 20 of 37 patients occurred ER, the ER rate was 54.1%, which was significantly higher than the ER rate of aMWA-rMWA group (p = 0.048) (Fig. 3a). The ER rate of patients whose actual treatment was consistent with recommend-

Fig. 2. ROC curve of LH training set (a); LH validation set (b); MWA train set (c); MWA validation set (d).
ed treatment was lower than that of the inconsistent patients.

In order to verify the ER prediction accuracy again, the hybrid model was used to predict ER probability according to the actual treatment taken by patients in 4 groups. We compared the predicted ER probability and the actual ER rate in each group. For the aLH-aLH group, the actual ER rate was 21.2%, and the predicted ER probability was 22.7 ± 10.9% (p = 0.438). For aLH-rMWA group, the actual ER rate was 46.2%, and the predicted ER probability was 39.0 ± 18.4% (p = 0.058). For aMWA-rMWA group, the actual ER rate was 26.3%, and the predicted ER probability was 27.8 ± 14.1% (p = 0.655). For aMWA-rLH group, the actual ER rate of was 54.1%, and the predicted ER probability was 56.8 ± 16.4% (p = 0.327) (Fig. 3b, c). There was no statistical difference between the actual ER rate and the predicted ER probability according to the actual treatment taken by patients in all 4 groups.

Clinical Use

For the patients whose actual treatment was not consistent with recommended treatment by the hybrid model, we assumed that these patients had taken the recommended treatment as their first treatment. We used the hybrid model to predict the ER probability according to the recommended treatment and compared the ER probability of recommended treatment with the patients’ actual ER rate and the ER probability of actual treatment. For the 26 patients in aLH-rMWA group, the predicted ER probability of MWA was 23.3 ± 13.4%, which was significantly lower than the actual ER rate (23.3 ± 13.4% vs. 46.2%, p < 0.001) and the predicted ER probability of LH (23.3 ± 13.4% vs. 39.0 ± 18.4%, p < 0.001). For the 37 patients in the aMWA-rLH group, the predicted ER probability of LH was 28.7 ± 14.8%, which was also significantly lower than the actual ER rate (28.7 ± 14.8% vs. 54.1%, p < 0.001) and the predicted ER probability of MWA (28.7 ± 14.8% vs. 56.8 ± 16.4%, p < 0.001). Therefore, for patients who were inconsistent with recommendation (aLH-rMWA and aMWA-rLH group), they could achieve significant decrease in ER probability through receiving the recommended treatment rather than their actual treatment (Fig. 3d). When the hybrid model was not adopted, for the overall 115 patients in comprehensive test set, the ER rate was 38.3% and ER probability of actual treatment was 38.2%. However, if the hybrid model was used to assist treatment decision, patients could follow the hybrid model recommendations to choose the treatment with less ER probability. In an ideal state, all 115 patients could receive the recommended treatment, so we predicted ER probability according to the recommended treatment for each patient. Finally, the ER probability for all 115 patients would drop from 38.3% to 25.6% (Fig. 4).

---

**Fig. 3.** Predictive performance of combined model in test set. a The distinguishing ability of combined model in actual recurrence. b Combined model’s predicted recurrence rate. c Comparison of actual recurrence rate and predict recurrence rate in 4 groups. d The effect of changing treatment in aLH-rMWA group and aMWA-rLH group.
**Discussion**

Tumor recurrence forces HCC patients to have to receive repeated and endless treatments, which inevitably cause huge damage to patient’s body condition and place enormous pressure on their economic situation. Further, in the end, with the exception of a few lucky individuals, the majority of patients lose in their long war against cancer [16]. Previous studies have demonstrated that ER is an important factor affecting overall survival for HCC patients receiving LH or MWA [3, 17]. Since previous studies have proved that both LH and MWA can achieve comparable and satisfactory therapeutic effects for small HCC (≤3 cm), we focused on those with a diameter of 3–5 cm [15, 18], which is still an area of concern when using LH and MWA. Due to the complexity of 3–5-cm HCC, multidisciplinary panel is often organized to make decision recommendation. However, even in multidisciplinary panel, guidelines and clinical experience are still the main foundation for treatment decisions. But guidelines focus on the universality, not individualization, and clinical experience inevitably contains subjectivity and tendencies. These factors make multidisciplinary panel not necessarily able to recommend the optimal and individualized treatment decision. Therefore, it is imperative to make an effective tool for treatment decisions to prevent ER and maximize treatment benefits.

There have been some studies on the predicted recurrence of HCC after resection or thermal ablation [10, 17, 19–23], which mainly divided patients into high-risk and low-risk groups using predictive models, and then observe differences in tumor recurrence. These studies often recommend the high-risk patients to undergo transcatheter arterial chemoembolization, chemotherapy, or target therapy after the first treatment to enhance or consolidate treatment benefits [22, 23]. However, because these studies only focused on one treatment modality, they could not solve the problem of how to make optimal treatment decisions. On this basis, some studies have focused on the treatment decision-making problem of the two different treatment modalities. Liu et al. [10] built two models based on semantic information and ultrasound image information to predict ER for ≤5-cm HCC patients receiving RFA or LH. Based on two models, patients in the LH group and RFA group would be divided into high-risk group and low-risk group for each treatment. After estab-
lishing the risk group confusion matrix, they observed changes in ER probability based on the assumption that patients in the high-risk group would exchange treatment. This type of study is an enlightening development upon previous research designs, which makes full use of semantic and image information and could partly answer the question of treatment decision-making. Regrettably, this research design ignores the real gap in ER rate among different treatments. In the study by Wang et al., the ER rate of LH was less than that of RFA. Inevitably, there is a possibility that the recurrence probability of the low-risk RFA group may still be higher than that of the high-risk LH group. When doctors make treatment decision for patients, they concern about the recurrence probability of each treatment, not just risk classification. Therefore, we once again optimized the design by making treatment decision based on the ER possibilities of different treatments.

In the process of variable selection, we chose traditional univariate feature selection as the practical method for its advantage of excellent robustness and interpretability. We did not use other variable selection methods, such as recursive feature elimination, selection from model, or feature selection as part of a pipeline, which are more suitable for super-large data analysis. Liver dysfunction and tumor size were prognostic factors of ER for LH and thermal ablation, which were consistent with previous predictive studies for HCC [24–26].

In our research, a traditional linear model (LR) and four classic machine learning models (RF, NNET, SGB, and XGB) were included and used to make predictions. Compared with the traditional linear model, although the machine learning model based on the decision tree (RF, SGB, and XGB) and the NNET model based on the neural network increase the amount of calculation, the model may get rid of the constraints of the traditional linear model caused by the fixed coefficients of each variable and could get the ability to use semantic information more flexibly and to excavate potential laws more fully. Some machine learning algorithms could partially overcome data insufficiency through the built-in interpolation function, and effectively improve the promotion ability. In LH and MWA verification sets, LH-XGB model and MWA-SGB model achieved the highest AUC, respectively. Therefore, these two models are connected and form the hybrid model. Both XGB and SGB belong to the GBM algorithm. SGB introduces randomness into the process of gradient boosting, while XGB strengthens and pushes this process to the eXtreme. XGB has certain advantages in the process of machine learning with an extremely large sample size, but it does not necessarily have absolute advantages in the scope of clinical medical data sample size. Therefore, the algorithm selected by the LH group and the MWA group is not the same, which is in line with our expectation.

In the process of verifying the distinguishing ability of the hybrid model, we found that whether in the LH group or MWA group, the ER rate of patients whose treatment were consistent with model recommended was significantly lower than the ER rate of inconsistent patients. This result can prove that our model has the effective ability to distinguish patients who could benefit more by LH from patients who could benefit more by MWA in overall patient population. In the process of verifying the predictive ability, the hybrid model has been used to predict the ER probability for each patient according to their actual treatment. Then we compare the average predicted ER probability with the corresponding actual ER rate in each quadrant of the treatment decision confusion matrix. There is no statistically different between actual ER rate and predicted ER probability in all four quadrants ($p > 0.1$). This result can prove that our model has the satisfactory predictive ability for all patients. In the process of verifying the clinical use value of the hybrid model, the patients whose treatment are inconsistent with model recommended could effectively reduce the ER probability by following the recommendations of model and changing the treatment method. And the reduction of ER probability in these patients will result in a huge decrease of ER probability in overall patient population by about one-third. At the same time, we also observed that there were more people in the aMWA-rLH group than the aLH-rMWA group. This is because the model completely recommends treatment based on ER probability. Due to the objective ER advantage of LH, in the ideal condition, more patients would be recommended for LH as the first treatment by the model, which is in line with our expectation.

However, treatment decision is also affected by many other factors, which would lead to the decrease in patients’ compliance for model recommendation. HCC patients usually have multiple liver-related comorbidities such as insufficient liver function reserve, portal hypertension, and coagulation dysfunction. In addition, with increasing age and comorbidity emergence, a reduction in tolerance ability also becomes a nonnegligible restrictive factor in treatment decision-making. Although our analysis shows that these factors are not significantly related to ER, they are closely related to treatment safety and decision-making. Thereby, expecting that all patients...
could blatantly ignore the restrictions of treatment itself and fully follow model recommendation is a utopian situation. In clinical practice, the restrictions for MWA mainly include whether tumor is visible on the image; and whether tumor is in a risky location. However, these restrictions could be overcome by multimodal image fusion and affiliated techniques [27–29]. The restrictions for LH mainly include whether the patient’s physical condition, such as comorbidity and coagulation function, can tolerate general anesthesia, and resection [6]; and whether tumor is located in a position that is difficult for manual LH to manage, such as posterosuperior liver segment [15]. These restrictions also could be surmounted by adopting new technology and medication. Robot-assisted laparoscopy represented by the Da Vinci system could easily resect tumors in positions that are difficult to reach with manual LH [30]. New medications, such as Avatrombopag Maleate, could improve blood clotting function and minimize the range of contraindications to surgery [31, 32]. Therefore, we believe that with the application of effective preoperative and intraoperative medical measures, the universality and realizability of model will continuously be promoted and the expected reduction in ER will gradually become a real-world situation.

Our research has some limitations. (1) This study was a retrospective multicenter study; so, data on some effective clinical variables may have been lost due to the different medical habits, and cases may have been excluded because of incomplete information. This may cause insufficient sample size and cause deviation. (2) This study only used multiple aspects of clinical semantic information, and whether adding image information could further improve the predictive ability for the ER is still worthy of further study.

**Conclusion**

Our study built a hybrid model by connecting two machine learning models in parallel to accurately predict the ER probability of LH and MWA, which could effectively reduce ER for single 3–5-cm HCC patients by providing reliable evidence to make optimal treatment decision.

**Statement of Ethics**

The studies involving human participants were reviewed and approved by the Ethics Committee of PLA General Hospital and the Ethical approval number is S2017-046-03. Our studies complied with the guidelines for human studies and were conducted ethically in accordance with the World Medical Association Declaration of Helsinki. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

**Conflict of Interest Statement**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Funding Sources**

This work was supported by Grant Nos. 81971625, 82030047, 92159305 and 91859201 from the National Scientific Foundation Committee of China, Health special scientific research topics: 20BJZ42.

**Author Contributions**

W.D.: protocol and project development, methodology, data collection, manuscript writing, and manuscript editing. Z.W.: data collection, manuscript writing, and manuscript editing. F.-Y.L., Z.-G.C., Z.H., and X.Y.: protocol and project development, and methodology. H.Z: protocol and project development. J.Y. and P.L.: protocol and project development, and manuscript review. All authors contributed to the article and approved the submitted version.

**Data Availability Statement**

We declare that all data supporting the findings of this study are included in the paper and online supplementary materials. Partial data may be available from the authors upon reasonable request and with permission of all centers of this study.

**References**

1 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):723–50.

2 Chen ZH, Zhang XP, Lu YG, Li LQ, Chen MS, Wen TF, et al. Actual long-term survival in HCC patients with portal vein tumor thrombus after liver resection: a nationwide study. *Hepatol Int*. 2020;14(5):754–64.

3 Poon RT, Fan ST, Lo CM, Liu CL, Wong J. Intrahepatic recurrence after curative resection of hepatocellular carcinoma: long-term results of treatment and prognostic factors. *Ann Surg*. 1999;229:216–22.
LH versus MWA: A Hybrid Machine Learning Model

4 Shimada M, Takenaka K, Gion T, Fujiwara Y, Kajiyama K, Maeda T, et al. Prognosis of recurrent hepatocellular carcinoma: a 10-year surgical experience in Japan. Gastroenterology. 1996;111(3111):720–116.

5 Buell JF, Thomas MT, Rudich S, Marvin M, Nagabandi R, Ravinda KV, et al. Experience with more than 500 minimally invasive hepatic procedures. Ann Surg. 2008;248(3):475–86.

6 Berardi G, Van Cleven S, Fretland ÅÅ, Halls M, Cipriani F, et al. Evolution of laparoscopic liver surgery from innovation to implementation to mastery: perioperative and oncologic outcomes of 2,238 patients from 4 European specialized centers. J Am Coll Surg. 2017;225(5):639–49.

7 Vietti Violi N, Duran R, Guiz B, Cercueil JP, Aubé C, Digiklia A, et al. Efficacy of microwave ablation versus radiofrequency ablation for the treatment of hepatocellular carcinoma in patients with chronic liver disease: a randomised controlled phase 2 trial. Lancet Gastroenterol Hepatol. 2018;3(5):317–25.

8 Zhu F, Rhim H. Thermal ablation for hepatocellular carcinoma: what’s new in 2019. Chin Clin Oncol. 2019;8(6):58.

9 Vogel A, Cervantes A, Chau I, Daniele B, Iorio JM, Meyer T, et al. Hepatocellular carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2018;29(in):v238–55.

10 Liu F, Liu D, Wang X, Xie H, Li B, Wang M, et al. Deep learning radiomics based on contrast-enhanced ultrasound might optimize curative treatments for very-early or early-stage hepatocellular carcinoma patients. Liver Cancer. 2020;9(4):397–413.

11 Minoda Y, Ishida E, Komori K, Ogino H, Otsuka Y, Chinen T, et al. Efficacy of endoscopic ultrasound with artificial intelligence for the diagnosis of gastrointestinal stromal tumors. J Gastroenterol. 2020;55(12):1119–26.

12 Marya NB, Powers PD, Chari ST, Gleeson FC, Leggett CL, Abu Dayyeh BK, et al. Utilisation of artificial intelligence for the development of an EUS-convolutional neural network model trained to enhance the diagnosis of autoimmune pancreatitis. Gut. 2021;70(7):1335–44.

13 Charlson M, Sztawrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. J Clin Epidemiol. 1994;47(11):1245–51.