Open Access Full Text Article

ORIGINAL RESEARCH

Utility of Machine Learning in the Prediction of Post-Hepatectomy Liver Failure in Liver Cancer

Hirotaka Tashiro, Takashi Onoe, Naoki Tanimine, Sho Tazuma, Yoshiyuki Shibata, Takeshi Sudo, Haruki Sada, Norimitsu Shimada, Hirofumi Tazawa, Takahisa Suzuki, Yosuke Shimizu

Department of Surgery, Kure Medical Center Chugoku Cancer Center, National Hospital Organization, Kure, Hiroshima, Japan

Correspondence: Hirotaka Tashiro, Department of Surgery, Kure Medical Center Chugoku Cancer Center, National Hospital Organization, 3-1, Aoyama, Kure, Hiroshima, 737-0023, Japan, Tel +823223111, Fax +823210478, Email tashiro.hirotaka.zn@mail.hosp.go.jp

Background: Posthepatectomy liver failure (PHLF) is a serious complication associated with high mortality rates. Machine learning (ML) has rapidly developed and may outperform traditional models in predicting PHLF in patients who have undergone hepatectomy. This study aimed to predict PHLF using ML and compare its performance with that of traditional scoring systems.

Methods: The clinicopathological data of 334 patients who underwent liver resection were retrospectively collected. The Pycaret library, a simple, open-source machine learning library, was used to compare multiple classification models for PHLF prediction. The predictive performance of 15 ML algorithms was compared using the mean area under the receiver operating characteristic curve (AUROC) and accuracy, and the best-fit model was selected among 15 ML algorithms. Next, the predictive performance of the selected ML-PHLF model was compared with that of routine scoring systems, the albumin-bilirubin score (ALBI) and the fibrosis-4 (FIB-4) index, using AUROC.

Results: The best model was extreme gradient boosting (accuracy:93.1%; AUROC:0.863) among the 15 ML algorithms. As compared with ALBI and FIB-4, the ML PHLF model had higher AUROC for predicting PHLF.

Conclusion: The novel ML model for predicting PHLF outperformed routine scoring systems.

Keywords: machine learning, posthepatectomy liver failure, liver cancer

Introduction

Hepatectomy is an effective treatment for liver cancer, including metastatic liver tumors. Although perioperative management, liver resection techniques, and advances in imaging modalities have contributed to improvements in outcomes after hepatectomy, post hepatectomy liver failure (PHLF) remains a serious complication associated with high mortality.[1,2]

The occurrence of PHLF is mainly determined by the quality and quantity of the remnant liver after resection. A variety of traditional scoring and nomogram models for predicting PHLF and permissible extent of liver resection in patients undergoing liver resection.[3] The Child-Pugh score, albumin-bilirubin (ALBI) score, and fibrosis index based on 4 factors (FIB-4) have been developed to determine the status of the liver.[4–6] However, these scoring systems cannot precisely predict PHLF. In Japan, the Makuuchi criteria, consisting of serum bilirubin levels and indocyanine green retention rate at 15 min (ICG-R15), have been used to determine the extent of hepatectomy.[7,8] However, the permissible liver resection rate cannot be calculated precisely. Recently, new models have been developed to predict PHLF using hepatobiliary scintigraphy and magnetic resonance imaging.[9,10] However, these criteria are not universally applicable.

Machine learning (ML), a field of artificial intelligence (AI), has been rapidly developed and is widely used to predict outcomes of liver diseases.[11–14] ML has also been used to predict PHLF after hepatectomy for HCC. ML may outperform traditional models as mentioned above. Mai et al developed an artificial neural network model to predict severe liver failure after hemihepatectomy in patients.[15] Wang et al also developed an ML prediction model for PHLF in HCC using a light gradient boosting machine (LightGBM) algorithm.[16] These reports demonstrated better predictive performance

1323

than traditional scoring systems. Pycaret, a simple, open-source, and low-code ML library in Python, has been utilized to predict the outcomes of several diseases.[17–19] The library compares these models and selects the best-fit model.

In this study, we used Pycaret to predict PHLF after liver resection for liver cancer and compared its performance with that of routine scoring systems.

Patients and Methods

Between January 2016 and December 2022, 384 patients underwent liver resection for primary and metastatic liver cancer at the Department of Surgery, Kure Medical Center, Chugoku Cancer Center, National Organization Hospital. Of the 384 patients, 50 who underwent simultaneous resection of other organs, such as the colon and stomach, or biliary reconstruction were excluded. In total, 334 patients who did not undergo additional surgery, excluding cholecystectomy, were enrolled in this study. Patients' medical records were reviewed retrospectively and data were collected. This retrospective study was approved by the Institutional Review Board of the Kure Medical Center (approval number: 2023–16). Informed consent was waived because of the retrospective design. This study was conducted following the Declaration of Helsinki. The following 20 pre-operative variables were used to predict PHLF: patient's age, sex, body mass index (BMI), tumor size, tumor number, tumor location, surgical approach (laparoscopic or open), number of repeat hepatectomy, hemoglobin (Hb) concentration, white blood count (WBC), platelet count (PLT), Total bilirubin (T-Bil) level, albumin (ALB) level, alanine aminotransferase (ALT) level, aspartate amino-transferase (AST) level, creatinine (Cr) level, prothrombin time (PT), indocyanine green retention rate at 15 min (ICG-R15), estimated liver resection rate (Res). Liver volume was evaluated preoperatively using a computed tomography (CT) volumetric system. Dynamic-enhanced CT images were obtained preoperatively for all patients. Liver volumes were calculated using volume analysis software (SYNAPSE VINCENT; Fujifilm Corp., Tokyo, Japan). The liver resection rate (Res) was calculated by dividing the liver volume to be removed (excluding the tumor volume) by the total liver volume. Post-hepatectomy liver failure (PHLF) was defined according to the International Study Group of Liver Surgery (ISGLS) criteria[2] as increased international normalized ratio (INR) and concomitant hyperbilirubinemia on or after postoperative day 5.

ML was performed using Colaboratory (https://colab.research.google.com/). Pycaret is a machine-learning library in Python that automates ML workflows (https://pycaret.org). Hyperparameter tuning is automatically performed by a RandomGridSearch. The tuning grid for hyperparameters is already defined by Pycaret for all models in the library. Ensemble modeling,[20] which combines the predictions from multiple machine-learning models, is implemented in Pycaret. We employed Pycaret to predict PHLF. Pycaret, which has 15 built-in classification models, was trained and evaluated using the dataset. The patient dataset was divided by random splitting into two sets: a training set and a test set. The predictive performance of the 15 ML algorithms was compared using the mean area under the receiver operating characteristic curve (AUROC), accuracy, recall, precision, and F1 scores. Accuracy was calculated as follows: (true positive + true negative) / total. The recall was calculated as true positive / (true positive + false negative). Precision was calculated as true-positive / (true-positive + false-positive). F1 was calculated as 2 × (precision × recall) / (precision + recall). The model with the highest accuracy in the training set was evaluated using the test data. Feature importance, which is a process used to select features in a dataset that contribute the most to predicting the target variable, was automatically permuted. The learning curve is visualized as a plot of data accuracy.

Calculation of routine scoring systems was performed as follows: The albumin-bilirubin (ALBI) score was calculated as $0.66 \times \log10$ (17.1×total bilirubin[mg/mL]) + $-0.885 \times$(10×ALB level[g/dl]);[5] fibrosis 4 (FIB-4) score was calculated as AST(U/L)×age(years)/[platelet count (×109/L)×ALT(U/L).[6] The predictive performance of the ML-PHLF model was compared with that of ALBI and Fib-4 using calculation of AUROC.

Continuous variables are expressed as median (range or interquartile range). In the univariate analysis, continuous data were analyzed using the Mann–Whitney *U*-test. Categorical variables were compared using Fisher's exact test. Multiple logistic regression analysis was performed to estimate the predictors of PHLF. Differences were considered significant if the *P* value was less than 0.05. Statistical analyses were performed using EZR version 1.52 (Saitama Medical Center, Jichi Medical University, Saitama, Japan).[21]

Results

A total of 334 patients were divided into the training (n=233) and validation (n=101) cohorts. The clinical characteristics of the entire dataset, training set, and validation set are shown in Table 1. The median age of the entire dataset was 74 years and 254 patients (76%) were men. A total of 209 patients (63%) were diagnosed with HCC, and 103 (31%) underwent repeat hepatectomy. 72 patients received pre-operative chemotherapy: liver metastasis from colorectal cancer, 56; liver metastasis from gastric cancer, 2; intrahepatic cholangiocarcinoma, 5; liposarcoma, 5; others, 4. 70 patients had pathological findings of cirrhosis. The clinical characteristics of the training and validation sets did not differ significantly different (Table 1). There was no case of mortality in this study. PHLF occurred in 31 (9.3%) patients: grade A, 22; grade B, 8; grade C,1. The occurrence rate of Dindo-Clavien ≥ III complication was significantly higher in patients with PHLF than that of patients without PHLF: 22.6% vs 8.2%.

Univariate analysis showed that tumor size, prothrombin time, operation method, platelet count, CRP level, Creatine level, ALB level, Res, ICGR15, tumor location, and T-Bil, AST, and ALT levels were significantly associated with PHLF development (Table 2). Multivariate analysis showed that Res, ICGR15, tumor location, and T-Bil and ALT levels were independent risk factors for PHLF (Table 2).

Table 1 The Characteristics of Whole Dataset, Training, and Test Sets

Characteristics	All Patients (n=334)	Training Set (n=233)	Testing Set (n=101)	p value
Age, years	74(34–90)	75(34–90)	73(39–87)	0.19
Sex, male/female, n	254/80	172/61	82/19	0.184
BMI	23.2(14.36–46.63)	23.3(14.36–46.63)	23.11(15.21–33.96)	0.621
Etiology, primary/metastatic/others, n	248/95/1	166/66/1	82/29/0	0.821
Number of tumor, n	1(1–18)	1(1–18)	1(1–18)	0.14
Size of tumor, mm	19(0.25–130)	18(0.25–130)	20(2.3–83)	0.943
Location, Seg7or8/others, n	173/161	113/120	60/41	0.074
Hemoglobin, g/dl	12.8(6.4–17.8)	12.9(6.4–17.8)	12.6(8.1–16.2)	0.623
Platelet count, x10^6/L	17.25(4.3–44.3)	16.95(4.3–44.3)	18.3(6–42)	0.17
Total bilirubin, mg/dl	0.7(0.2–2.8)	0.7(0.2–2.8)	0.7(0.3–1.9)	0.879
Alanine aminotransferase, IU/L	18(4–145)	18(4–145)	18(5–65)	0.985
Aspartate aminotransferase, IU/L	24(6–102)	24(10–102)	23(6–70)	0.301
Albumin, g/dl	4(2–5.1)	4(2–5.1)	4(2–5)	0.621
Prothrombin time, %	99(57–149.8)	97.9(57–149.8)	101.2(66.5–122.8)	0.108
Creatinine, mg/dl	0.84(0.37–11.42)	0.85(0.37–9.72)	0.82(0.47–11.42)	0.453
ICGR15%	11.15(0–58.5)	11.75(0–58.5)	10.7(0.5–50.9)	0.735
CRP, mg/dl	0.13(0.01–27.1)	0.13(0.01–27.1)	0.13(0.02–2.85)	0.691
Repeat surgery, 1st/2nd/3rd/4th	231/75/21/7	164/52/13/4	67/23/8/3	0.655
Laparoscopic, yes/no	191/143	129/104	62/39	0.337
Estimated Liver resection rate, %	13.8(0.6–64)	13.1(0.6–59)	15(1–64)	0.397
ASA score	1/231/102	160/73	1/71/29	0.319
PHLF	31/303 (9.3%)	22/211 (9.4%)	9/92 (8.9%)	1

Note: Continuous values given as median (range) or absolute and relative frequencies.
Abbreviations: BMI, body mass index; ICGR15, indocyanine green retention rate; ASA, American society of anesthesiologist.

Table 2 Univariate and Multivariate Analyses for PHLF

	Univariate Analysis			Multivariate Analysis		
	PHLF		**p value**	**OR**	**95% CI**	**p value**
Characteristics	**Yes**	**No**				
Age, years	73(67.5–78.5)	74(68–79)	0.329			
Sex, male/female, n	28/3	226/77	0.07			
BMI	24.74(21.09–25.75)	23.26(21.46–25.55)	0.519			
Etiology, primary/metastatic/others, n						
Number of tumor, n	1(1–2)	1(1–2)	0.152			
Size of tumor, cm	35(17–65)	16(11–28)	<0.001			
Location of tumor, Seg7or8/others	23/8	150/153	0.013	5.13	1.31–20.1	0.019
Hemoglobin, g/dl	12.8(11.6–14.4)	12.8(11.6–14.05)	0.864			
Platelet count, ×10^6/L	16.5(11.3–19.75)	17.4(13.4–22.1)	0.101			
Total bilirubin, mg/dl	1.0(0.8–1.2)	0.7(0.5–0.9)	<0.001	12.1	2.6–56.50	<0.001
Alanine aminotransferase, IU/L	25(18–41)	17(13–26)	<0.001			
Asparatate aminotransferase, IU/L	33(26–40)	24(19–32)	<0.001	1.07	1.02–1.13	0.006
Albumin, g/dl	3.8(3.15–4.15)	4.1(3.8–4.3)	<0.001			
Prothrombin time, %	95.9(79.2–101.35)	99.4(92.5–107.1)	0.029			
Creatinine, mg/dl	0.82(0.665–0.875)	0.84(0.720–0.980)	0.25			
ICGR15,%	15.9(9.4–30.25)	10.7(6.5–15.75)	<0.001	1.2	1.09–1.31	<0.001
CRP, mg/dl	0.32(0.13–0.875)	0.12(0.06–0.280)	<0.001			
Repeat surgery, n(%)	1(1–1)	1(1–2)	0.442			
Laparoscopic, yes/no	8/23	183/120	<0.001			
Estimated Liver resection rate, %	29(13.5–43)	13(4–24)	<0.001	1.14	1.07–1.21	<0.001
ASA score	2(2–3)	2(2–3)	0.517			

Note: Continuous values given as median (interquartile range).
Abbreviations: BMI, body mass index; ICGR15, indocyanine green retention rate; ASA, American society of anesthesiologist.

Using 20 variables, the ML model was trained and tuned to the training set, and the performance of 15 classification algorithms was compared. ML analysis showed the accuracy, AUC, recall, precision, and F1 in the training set (Supplementary Table 1). Linear Discriminant analysis (LDA)[22] was the best model (accuracy:91.8%, AUROC:0.879). The learning curve, confusion matrix, classification report, and AUROC (Supplementary Figure 1) in the testing set showed that the prediction of PHLF negativity outperformed the precision, recall, and F1 scores by 0.957, 0.967, and 0.962, respectively. However, the prediction of PHLF positivity did not outperform that of PHLF negativity; the precision, recall, and F1 scores were 0.625, 0.556, and 0.588, respectively (Supplementary Figure 1). Next, five risk factors identified by multivariate analysis were used as input variables: Res, ICGR15, tumor location, T-Bil level, and ALT level. However, the performance of ML using 20 variables was almost the same as that of ML using five independent risk factors: Res, ICGR15, tumor location, T-Bil level, and ALT level (Supplementary Figure 2 and Supplementary Table 2). Thus, the ML was trained and tuned using 12 variables identified by univariate analysis. The performance of ML using the 12 variables improved slightly. Extreme Gradient Boosting (XGBoost) was the best model

(accuracy:93.2%, AUC:0.862) (Table 3). The learning curve showed that the training scores were higher than the cross-validation scores, and the cross-validation score reached 0.9 in 100 samples (Figure 1A). The feature importance plot of ML in the training set is shown in Figure 1B. The five highest features were Res, which was the most important feature among the clinical characteristics, ALB level, PT, ICGR-15, and laparoscopic approach. The confusion matrix (Figure 1C) and classification report (Figure 1D) for the testing set are shown in Figure 1. The ML classification report demonstrated that the predictive performance of no PHLF with precision, recall, and F1 scores were 0.958, 0.989, and 0.973, respectively. The predictive performance of the PHLF model showed precision, recall, and F1 scores of 0.833, 0.556, and 0.677, respectively (Figure 1D). Last, we performed ensemble model using light gradient boosting machine (lightGBM) and XGBoost (Supplementary Table 3).

We further compared the predictive performance of the ML PHLF model with that of routine clinical models, including ALBI and Fib-4. The ML-PHLF model had the highest AUROC for the prediction of PHLF among the non-invasive models in the training and testing sets (Table 4 and Supplementary Figure 3). The ML PHLF model using Pycaret was more accurate than routine scoring models in predicting PHLF.

Discussion

In the current study, we developed and validated an ML model for predicting PHLF after hepatectomy for primary and metastatic liver cancers using PyCaret. The accuracy of the PHLF model was superior to that of conventional models. To the best of our knowledge, this is the first study using Pycaret to predict PHLF after hepatectomy for liver cancers.

ML has been increasingly applied to the field of liver surgery to predict the outcomes of patients who have undergone liver resection. Two studies on ML models developed to predict PHLF have been reported. Mai et al reported that an artificial neural network model predicted the risk of severe PHLF after hemi-hepatectomy in patients with HCC.[15] Wang et al developed an ML model based on light gradient boosting machines (Light GBM), which accurately predicted the risk of PHLF compared to conventional models such as ALBI, FIB-4, and CTP.[16] In this study, we adopted the Pycaret

Table 3 Comparison of 15 Classification Algorithms by Pycaret

	Model	Accuracy	AUC	Recall	Precision	F1
xgboost	Extreme Gradient Boosting	0.9312	0.8628	0.4333	0.6583	0.4924
qda	Quadratic Discriminant Analysis	0.9274	0.7588	0.4167	0.5167	0.4433
lr	Logistic Regression	0.9232	0.8763	0.35	0.45	0.3833
svm	SVM - Linear Kernel	0.9232	0	0.2333	0.4	0.2833
lda	Linear Discriminant Analysis	0.9232	0.8784	0.4	0.5667	0.4567
lightgbm	Light Gradient Boosting Machine	0.9225	0.8345	0.4167	0.5833	0.47
rf	Random Forest Classifier	0.9185	0.7831	0.1833	0.4	0.25
gbc	Gradient Boosting Classifier	0.9185	0.841	0.35	0.5	0.4033
et	Extra Trees Classifier	0.9145	0.8207	0.2	0.35	0.25
knn	K Neighbors Classifier	0.9143	0.6837	0.1	0.2	0.1333
Ridge	Ridge Classifier	0.9143	0	0.1333	0.3	0.1833
ada	Ada Boost Classifier	0.9101	0.7545	0.35	0.5333	0.3971
dt	Decision Tree Classifier	0.9058	0.6846	0.4167	0.4433	0.4055
Dummy	Dummy Classifier	0.9058	0.5	0	0	0
nb	Naive Bayes	0.8844	0.8829	0.5833	0.4	0.4676

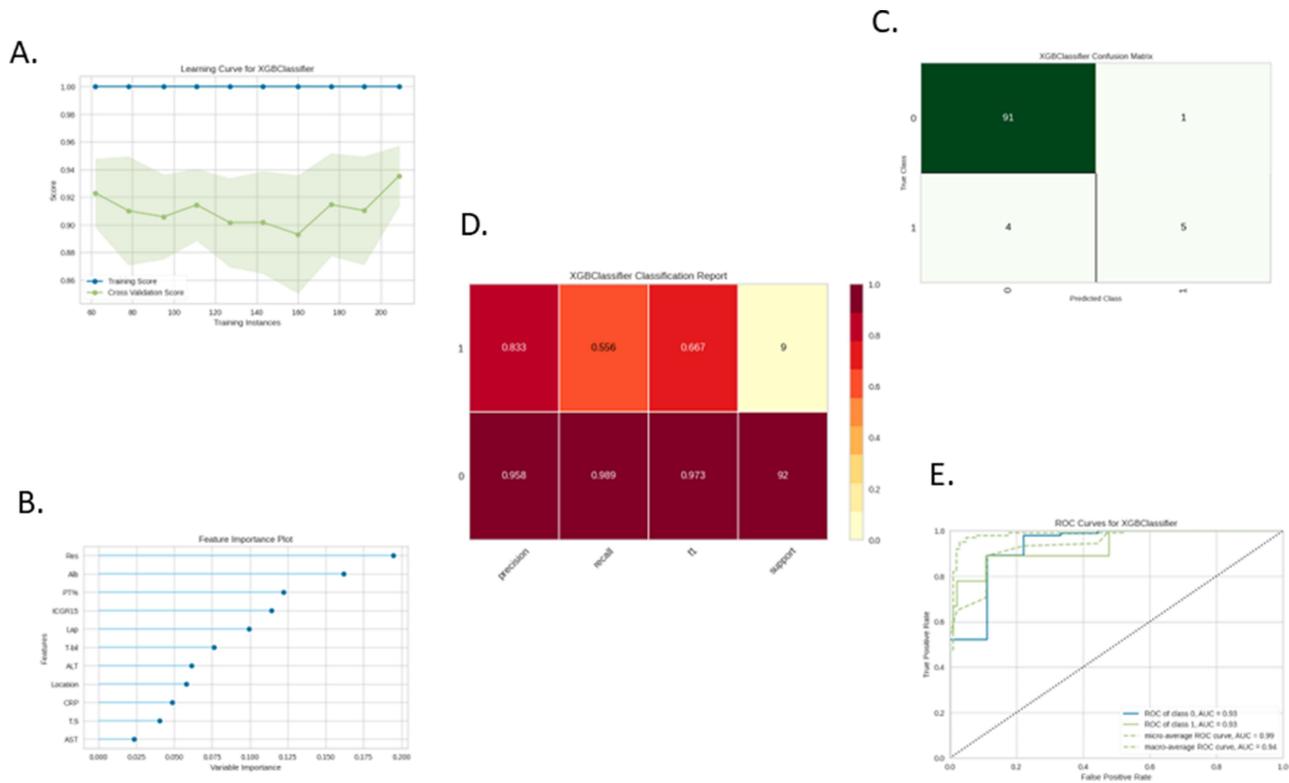**Abbreviation**: AUC, area under curve.

Figure 1 Learning curve (**A**) feature importance plot (**B**) confusion matrix (**C**) classification report (**D**) receiver operating characteristics (ROC) curve (**E**) of XGBClassifier using 12 variables identified by univariate analysis.

library, which contains 15 built-in classification models. Pycaret contains decision tree classifiers, such as the Random Forest classifier and Light GBM, discriminant analysis, logistic regression, K Neighbor Classifier, and support vector machine (SVM). In Pycaret, the predictive performance of the 15 ML algorithms was compared based on the accuracy, mean AUROC, recall, precision, and F1 scores. The best model was evaluated in terms of accuracy and AUROC for the test set. A confusion matrix, learning curve, and feature importance were provided. First, we performed ML using 20 variables and the model's accuracy was 91.8%. Using the five variables identified by multivariate analysis, the accuracy was 92.3%. Finally, using the 12 variables identified by univariate analysis, the accuracy was 93.1%. Consequently, the performance of ML was trained, tuned, and improved. In the training set, XGBoost, which is a scalable, distributed gradient-boosted decision tree (GBDT) ML library,[23] showed the best performance among the 15 algorithms, with an accuracy, AUROC, recall, and precision of 0.9183, 0.8628, 0.433, and 0.6583, respectively. Our study has an advantage

Table 4 Comparison of Performances of ALBI, Fib-4, and ML for Predicting PHLF

	Area Under the ROC	95% CI
ALBI	0.754	0.645–0.863
FIB-4	0.643	0.512–0.774
ML	0.862	0.763–0.912

Abbreviations: ROC, receiver operating characteristics; ALBI, the albumin-bilirubin score; FIB-4, the fibrosis-4 index; ML, machine learning.

over previous studies in terms of comparing the performances of 15 ML algorithms and predicting the occurrence of PHLF using the best model.

Among the 101 patients in the test set, 91 of 92 cases of no PHLF (specificity:0.967) and 5 of 9 cases of PHLF (recall:55.6%) were precisely predicted. ML performed well in predicting negative PHLF with precision, recall, and F1 scores of 0.957, 0.967, and 0.962, respectively. However, the predictive performance for positive PHLF was low, with precision, recall, and F1 scores of 0.833, 0.556, and 0.667, respectively. This discrepancy is believed to be due to the imbalance in the rates of positivity and negativity of PHL in our study: the incidence of PHLF was 9.3%. The performance of ML in predicting PHLF should improve in cases limited to major hepatectomy because the incidence rate of PHLF is high in cases of major hepatectomy. Large-scale studies are required to improve the performance of ML in predicting PHLF.

The top five features ranked in our ML model were Res, ALB level, PT, ICGR-15, and laparoscopic approach. These Results are partially incompatible with the results of a multivariate analysis that showed that Res, ICGR15, tumor location, T-Bil level, and ALT level were independent risk factors for PHLF. Res and ICGR-15 were common important variables in both the ML model and the multivariate analysis. Res, which is the most commonly identified significant predictor of PHLF,[3] was the most important feature of the ML model. ICGR15, which is included in the Makuuchi criteria[7] and is recognized as one of the most useful indicators of liver function in Japan, was a top five important feature in the ML model. ALB and prothrombin, which are important proteins produced in the liver, were the second- and third-most important features in the ML model. ALB level and prothrombin time are important parameters used in MELD and Child-Pugh scores.[4–6] The occurrence rate of PHLF was significantly lower in the laparoscopic approach than in the open approach. The laparoscopic approach is less invasive than the open approach. Differences in the hepatectomy approach are unlikely to affect postoperative liver function. These discrepancies between the ML and multivariate analyses may be attributed to the study methodologies.

The ML PHLF model using Pycaret was more accurate in predicting PHLF than the traditional scoring models, Fib-4 and ALBI. However, these traditional models can more easily calculate scores and predict PHLF. These statistics-based traditional scoring models and the ML model are complementary.[24] The ML model for predicting PHLF combined with traditional scoring models should contribute to safer and more accurate liver resection.

This study has some limitations. First, the severity of PHLF was classified into 3 grades. Several studies have focused on severe PHLF (grade B or C). In the current study, the number of cases of PHLF grades A, B and C was 22, 8 and 1, respectively, and this was too small to allow focus on severe PHLF. Second, this study was non-specific for primary liver cancer because it consisted of patients diagnosed with primary and secondary liver cancers, including metastasis, who had undergone hepatectomy. Differences in etiology may affect liver function. There is a need for a large-scale study using a multicenter database focusing on severe PHLF and specific etiology.

In conclusion, ML using Pycaret in combination with traditional scoring systems may contribute to safer and more accurate liver resection.

Acknowledgment

Author Contributions

All authors contributed to data analysis, drafting or revising the article, have agreed on the journal to which the article will be submitted, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

Disclosure

The authors have no commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements) that might pose a conflict of interest related to the submitted manuscript.

References

1. Belghiti J, Hiramatsu K, Benoist S, Massault P, Sauvanet A, Farges O. Seven hundred forty-seven hepatectomies in the 1990s: an update to evaluate the actual risk of liver resection. *J Am Coll Surg*. 2000;191(1):38–46. doi:10.1016/S1072-7515(00)00261-1
2. Rahbari N, Garden J, Padbury R, et al. Posthepatectomy liver failure: a definition and grading by the International Study Group of Liver Surgery (ISGLS). *Surgery*. 2011;149(5):713–724. doi:10.1016/j.surg.2010.10.001
3. Yoshino K, Yoh T, Taura K, Seo S, Ciria R, Briceno-Delgado J. A systematic review of prediction models for post-hepatectomy liver failure in patients undergoing liver surgery. *HPB*. 2021;23(9):1311–1320. doi:10.1016/j.hpb.2021.05.002
4. Pugh RN, Murray-Lyon IM, Dawson JL, Pietroni MC, Williams R. Transection of the oesophagus for bleeding oesophageal varices. *Br J Surg*. 1973;60(8):646–649. doi:10.1002/bjs.1800600817
5. Johnson PJ, Berhane S, Kagebayashi C, et al. Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach-the ALBI grade. *J Clin Oncol*. 2015;33(6):550–558. doi:10.1200/JCO.2014.57.9151
6. Sterling RK, Lissen E, Clumeck N, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology*. 2006;43(6):1317–1325. doi:10.1002/hep.21178
7. Makuuchi M, Kosuge T, Takayama T, et al. Surgery for small liver cancers. *Semin Surg Oncol*. 1993;9(4):298–304. doi:10.1002/ssu.2980090404
8. Imamura H, Seyama Y, Kokudo N, et al. One thousand fifty-six hepatectomies without mortality in 8 years. *Arch Surg*. 2003;138(11):1198–1206. doi:10.1001/archsurg.138.11.1198
9. Olthof PB, Arntz P, Truant S, et al. Hepatobiliary scintigraphy to predict postoperative liver failure after major liver resection; a multicenter cohort study in 547 patients. *HPB*. 2023;25(4):417–424. doi:10.1016/j.hpb.2022.12.005
10. Li C, Wang Q, Zou M, et al. A radiomics model based on preoperative gadoxetic acid-enhanced magnetic resonance imaging for predicting post-hepatectomy liver failure in patients with hepatocellular carcinoma. *Front Oncol*. 2023;13:1164739. doi:10.3389/fonc.2023.1164739
11. Famularo S, Donadon M, Cipriani F, et al. Machine learning predictive model to guide treatment allocation for recurrent hepatocellular carcinoma after surgery. *JAMA Surgery*. 2023;158(2):192–202. doi:10.1001/jamasurg.2022.6697
12. Ruzzenente A, Bagante F, Poletto E, et al. A machine learning analysis of difficulty scoring systems for laparoscopic liver surgery. *Surg Endo*. 2022;36(12):8869–8880. doi:10.1007/s00464-022-09322-7
13. Theysohn J, Demircioglu A, Kleditzsch M, et al. Prediction of left lobe hypertrophy after right lobe radioembolization of the liver using a clinical data model with external validation. *Sci Rep*. 2022;12(1):20718. doi:10.1038/s41598-022-25077-6
14. Amygdalos I, Muller-Franzes G, Bednarsch J, et al. Novel machine learning algorithm can identify patients at risk of poor overall survival following curative resection for colorectal liver metastases. *J Hepatobiliary Pancreat Sci*. 2023;30(5):602–614. doi:10.1002/jhbp.1249
15. Mai RY, Bai T, et al. Artificial neural network model for preoperative prediction of severe liver failure after hemihepatectomy in patients with hepatocellular carcinoma. *Surgery*. 2020;168(4):643–652. doi:10.1016/j.surg.2020.06.031
16. Wang J, Zheng T, Liao Y, et al. Machine learning prediction model for post-hepatectomy liver failure in hepatocellular carcinoma: a multicenter study. *Front Oncol*. 2022;12:986867. doi:10.3389/fonc.2022.986867
17. Evrimler S, Gedik MA, Serel TA, et al. Bladder urothelial carcinoma: machine learning-based computed tomography radiomics for prediction of histological variant. *Acad Radiol*. 2022;29(11):1682–1688. doi:10.1016/j.acra.2022.02.007
18. Yang E, Ding Q, Fan X, et al. Machine learning modeling and prognostic value analysis of invasion-related genes in cutaneous melanoma. *Comput Biol Med*. 2023;162:107089. doi:10.1016/j.compbiomed.2023.107089
19. Lundervold AJ, Hillestad EMR, Lied GA, et al. Assessment of self-reported executive function in patients with irritable bowel syndrome using a machine-learning framework. *J Clin Med* 2023;31:3771. doi: 10.3390/jcm12113771
20. Zhang Z, Chen L, Xu P, Hong Y. Predictive analytics with ensemble modeling in laparoscopic surgery: a technical note. *Laparosc Endosc Rob Surg*. 2022;5(1):25–34. doi:10.1016/j.lers.2021.12.003
21. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. *Bone Marrow Transplant*. 2013;48(3):452–458. doi:10.1038/bmt.2012.244
22. Fisher RA. The use of multiple measurements in taxonomic problems. *Ann Eugenics*. 1936;7(2):179–188. doi:10.1111/j.1469-1809.1936.tb02137.x
23. Chen T, He T, Benesty M, et al. Extreme Gradient Boosting. Package Version-0.4-1.4; 2015. Available from: https://xgboost.ai/. Accessed May, 15, 2023.
24. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods*. 2018;15(4):233–234. doi:10.1038/nmeth.4642