

# CT-Based Deep Learning Enables Postoperative Recurrence Prediction for Intrahepatic Cholangiocarcinoma

Taiichi Wakiya (✉ [wakiya1979@hirosaki-u.ac.jp](mailto:wakiya1979@hirosaki-u.ac.jp))

Hirosaki University Graduate School of Medicine

**Keinosuke Ishido**

Hirosaki University Graduate School of Medicine

**Norihisa Kimura**

Hirosaki University Graduate School of Medicine

**Hayato Nagase**

Hirosaki University Graduate School of Medicine

**Taishu Kanda**

Hirosaki University Graduate School of Medicine

**Sotaro Ichiyama**

Hirosaki University School of Medicine

**Kenji Soma**

Hirosaki University School of Medicine

**Masashi Matsuzaka**

Hirosaki University Hospital

**Yoshihiro Sasaki**

Hirosaki University Hospital

**Shunsuke Kubota**

Hirosaki University Graduate School of Medicine

**Hiroaki Fujita**

Hirosaki University Graduate School of Medicine

**Takeyuki Sawano**

Aomori Prefectural Central Hospital

**Yutaka Umehara**

Aomori Prefectural Central Hospital

**Yusuke Wakasa**

Aomori City Hospital

**Yoshikazu Toyoki**

Aomori City Hospital

**Kenichi Hakamada**

**Article**

**Keywords:**

**Posted Date:** March 15th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1444366/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Preoperatively accurate evaluation of risk for postoperative recurrence contributes to maximizing the therapeutic success of surgical resection for intrahepatic cholangiocarcinoma (iCCA). This study aimed to investigate the potential of deep learning (DL) algorithms for predicting postoperative early recurrence through the use of preoperative images. We collected the dataset, including preoperative plain computed tomography (CT) images, from 41 patients undergoing curative surgery for iCCA at multiple institutions. We built a CT patch-based predictive model using a residual convolutional neural network and used 5-fold cross-validation. The prediction accuracy of the model was analyzed. We defined early recurrence as recurrence within a year after surgical resection. Of the 41 patients, early recurrence was observed in 20 (48.8%). A total of 71,081 patches were extracted from the entire segmented tumor area of each patient. The average accuracy of the ResNet model for predicting early recurrence was 98.2% for the training dataset. In the validation dataset, the average sensitivity, specificity, and accuracy were 97.8%, 94.0%, and 96.5%, respectively. Furthermore, the area under the receiver operating characteristic curve was 0.994. Our CT-based DL model exhibited high predictive performance in projecting postoperative early recurrence, proposing a novel insight into iCCA management.

## Introduction

Primary liver cancer was the third leading cause of cancer death worldwide in 2020, with 830,180 deaths contributing to 8.3% of worldwide cancer-related deaths. In 2020, 905,677 liver cancer diagnoses were made globally, comprising 4.7% of worldwide cancer cases<sup>1</sup>. Intrahepatic cholangiocarcinoma (iCCA) is the second most common primary liver cancer after hepatocellular carcinoma and accounts for 10–15% of primary liver cancer<sup>1</sup>. Incidences of iCCA have been increasing over the last three decades<sup>2,3</sup>. However, the prognosis of iCCA, unfortunately, remains extremely poor, with a 5-year overall survival of 9%<sup>4</sup>. This cancer presents a substantial health problem worldwide, so treatments to improve survival are urgently needed.

In tackling this lethal disease, surgical resection has been the most fundamental and only treatment with the potential for cure<sup>2,5</sup>. Unfortunately, only about 20–40% of patients present with potentially operable disease<sup>3,6</sup>. In addition, recurrence following surgical resection of iCCA remains a big concern. Several studies have demonstrated that around 50% of patients recurred after curative-intent surgery<sup>5,7,8</sup>. In short, the number of patients who benefit from surgical resection alone is limited.

Preoperatively accurate evaluation of risk for postoperative recurrence contributes to maximizing the therapeutic success of surgical resection. Several lines of evidence from the clinical studies have demonstrated that factors associated with a higher risk for recurrence include large tumor size, multiple tumors, vascular invasion, lymph node metastasis, and R1 resection<sup>7–9</sup>. These reports provided us with significant insight. However, definitive preoperative diagnosis of the above factors, especially lymph node metastasis, remains challenging<sup>10</sup>. Furthermore, these reports were analyzed using multivariable logistic

regression modeling to identify the independent risk factors for postoperative recurrence. This method has been traditionally performed in clinical studies, but there have been certain limitations, such as selection of variables, confounding factors, and multicollinearity.

To resolve the above issues, in this study, we created a prediction model for postoperative recurrence using artificial intelligence (AI). AI has the potential to revolutionize disease diagnosis and management in the medical field <sup>11</sup>. Deep learning (DL) has recently gained extensive attention as a technique for realizing the full potential of AI <sup>12</sup>. Convolutional neural networks (CNNs), which are a DL approach, are especially recognized as demonstrating high performance in image recognition <sup>13,14</sup>. Indeed, there have been some successful reports in applying DL to the assessment and prediction of radiological images in clinical settings <sup>13,15,16</sup>.

In the field of iCCA, a recent study using DL has demonstrated the feasibility of applying it to liver tumor diagnoses <sup>17</sup>. Moreover, Jeong et al. showed its usefulness in the prognostic estimation and stratification of susceptible individuals for adjuvant treatment after resection <sup>18</sup>. However, there has been no report applying DL to recurrence prediction in patients with iCCA after resection. Thus, this study aimed to investigate the potential of DL algorithms for predicting postoperative recurrence through the use of preoperative images of iCCA. Here, we have successfully developed a prediction model using CNNs and propose a novel concept in iCCA management from a completely different perspective.

## Results

### **Comparison of the perioperative characteristics of the non-early recurrence and early recurrence groups**

The clinical characteristics of the 41 enrolled patients are shown in Table 1. Of the 41 patients, early recurrence was observed in 20 (48.8%). The early recurrence group demonstrated increased levels of tumor biomarkers such as carbohydrate antigen 19 - 9 and carcinoembryonic antigen (Table 1). However, there were no significant differences in the biomarkers between the groups. Univariate analysis indicated that Union for International Cancer Control (UICC) N category was the only significant predictor of early recurrence ( $P = 0.002$ ). The odds ratio (OR) was 11.611 (95% confidence interval (CI): 2.116–63.726).

Table 1

Comparison of the perioperative characteristics of the non-early recurrence and early recurrence groups.

	<b>All cases (n = 41)</b>	<b>Non-early recurrence (n = 21)</b>	<b>Early recurrence (n = 20)</b>	<b>P value</b>
Gender				0.879
Male, n	20 (48.8)	10 (47.6)	10 (50.0)	
Female, n	21 (51.2)	11 (52.4)	10 (50.0)	
Age, year	69 (39–81)	72 (39–81)	68 (46–81)	0.388
Body height, cm	157.0 (133.0- 178.0)	157.5 (133.0- 177.0)	156.0 (140.5– 178.0)	0.814
Body weight, kg	54.0 (34.0-81.6)	53.0 (34.0-81.6)	54.0 (38.9–73.5)	0.629
Body mass index, kg/m <sup>2</sup>	21.9 (16.6–29.3)	22.1 (17.0-29.3)	21.6 (16.6–27.4)	0.506
Child-Pugh score				0.232
A, n	39 (95.1)	21 (100)	18 (9.0)	
B, n	2 (4.9)	0	2 (10.0)	
C, n	0	0	0	
CA19-9, U/mL	152.0 (3.8- 80355.0)	128.6 (5.0- 3120.0)	314.0 (3.8- 80355.0)	0.371
CEA, ng/mL	3.5 (0.5–61.6)	2.6 (0.5–43.1)	3.7 (0.5–61.6)	0.438
Surgical procedure				0.570
Right hepatectomy, n	12 (29.3)	5 (23.8)	7 (35.0)	
Left hepatectomy, n	18 (43.9)	9 (42.9)	9 (45.0)	
Right anterior sectionectomy, n	3 (7.3)	2 (9.5)	1 (5.0)	
Right posterior sectionectomy, n	1 (2.4)	0	1 (5.0)	
Partial resection, n	7 (17.1)	5 (23.8)	2 (10.0)	
Tumor factors				
Tumor size, mm	50.0 (16.0-150.0)	50.0 (16.0-150.0)	55.0 (25.0-150.0)	0.347
Solitary tumor, n	28 (68.3)	17 (81.0)	11 (55.0)	0.074
CA19-9, carbohydrate antigen 19 – 9; CEA, carcinoembryonic antigen; UICC, Union for International Cancer Control				

	All cases (n = 41)	Non-early recurrence (n = 21)	Early recurrence (n = 20)	P value
UICC 8th edition				
T category, n				0.703
T1a	6 (14.6)	2 (9.5)	4 (20.0)	
T1b	3 (7.3)	2 (9.5)	1 (5.0)	
T2	24 (58.5)	14 (66.7)	10 (50.0)	
T3	5 (12.2)	2 (9.5)	3 (15.0)	
T4	3 (7.3)	1 (4.8)	2 (10.0)	
N category, n				0.002
N0	28 (68.3)	19 (90.5)	9 (45.0)	
N1	13 (31.7)	2 (9.5)	11 (55.0)	
M category, n				1.000
M0	41 (100)	21 (100)	20 (100)	
TNM Stage, n				0.055
IA	3 (7.3)	2 (9.5)	1 (5.0)	
IB	3 (7.3)	2 (9.5)	1 (5.0)	
II	17 (41.5)	12 (57.1)	5 (25.0)	
IIIA	3 (7.3)	2 (9.5)	1 (5.0)	
IIIB	15 (36.6)	3 (14.3)	12 (60.0)	
IV	0	0	0	
Intrahepatic metastasis, n	12 (29.3)	4 (19.0)	8 (40.0)	0.141
Positive surgical margin, n	3 (7.3)	2 (9.5)	1 (5.0)	> 0.999
CA19-9, carbohydrate antigen 19 - 9; CEA, carcinoembryonic antigen; UICC, Union for International Cancer Control				

## Binary logistic regression analysis

To predict early recurrence, we performed a binary logistic regression analysis, which is one of the traditional methods. We set early recurrence as the dependent variable. N category, which was found as a significant predictor through univariate analysis, was entered into a binary logistic regression analysis.

Furthermore, we entered the UICC T category and surgical margin status, which have been reported as factors associated with a higher risk for recurrence.

Binary logistic regression indicated that UICC N category was a significant predictor of early recurrence (Chi-Square = 11.952, and P = 0.063). The result of the Hosmer-Lemeshow test was P = 0.879. UICC N category was significant at the 5% level (Wald = 7.430, P = 0.006). The OR was 11.840 (95% CI: 2.002–70.008). The model correctly predicted 85.7% of cases without early recurrence and 60.0% of cases with early recurrence, giving an overall correct prediction rate of 73.2%. The model achieved an AUC of 0.770 (95% CI: 0.624–0.916) (Fig. 2).

## Performance of the ResNet50 model

A total of 25,765 patches were obtained from the 21 patients without early recurrence. Furthermore, a total of 45,316 patches were obtained from the 20 patients with early recurrence after surgery. Finally, a total of 71,081 patches were extracted from the 41 patients in the current study.

The average accuracy of the ResNet model for predicting early recurrence was 98.2% for the training dataset. The average sensitivity, specificity, and positive and negative predictive values were 98.9%, 97.0%, 98.3%, and 98.0%, respectively (Table 2). The model achieved an AUC of 0.9983 (95% CI: 0.9982–0.9985) in the training dataset.

Table 2  
Performance of the DL model in the training data set.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average	SD
Sensitivity, %	99.0	98.5	99.0	99.6	98.4	98.9	0.5
Specificity, %	94.4	97.9	96.6	99.1	97.1	97.0	1.7
False negative rate, %	1.0	1.5	1.0	0.4	1.6	1.1	0.5
False positive rate, %	5.6	2.1	3.4	0.9	2.9	3.0	1.7
Positive predictive value, %	96.9	98.8	98.1	99.5	98.4	98.3	1.0
Negative predictive value, %	98.1	97.5	98.2	99.3	97.1	98.0	0.9
Accuracy, %	97.3	98.3	98.1	99.4	97.9	98.2	0.8
AUC	0.997	0.999	0.998	1.000	0.998	0.998	0.1
AUC, the area under the receiver operating characteristic curve; SD, standard deviation							

Likewise, the model showed high predicting performance in the validation dataset. The average sensitivity, specificity, and positive and negative predictive values were 97.8%, 94.0%, 96.7%, and 96.1%, respectively. In the validation dataset, the ResNet model achieved an accuracy of 96.5% (Table 3). The model achieved an AUC of 0.994 (95% CI: 0.993–0.995) in the validation dataset (Fig. 2).

Table 3  
Performance of the DL model in the validation data set.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average	SD
Sensitivity, %	97.9	97.6	97.8	98.4	97.5	97.8	0.3
Specificity, %	90.6	95.9	92.5	96.6	94.6	94.0	2.5
False negative rate, %	2.1	2.4	2.2	1.6	2.5	2.2	0.3
False positive rate, %	9.4	4.1	7.5	3.4	5.4	6.0	2.5
Positive predictive value, %	94.8	97.7	95.8	98.1	96.9	96.7	1.3
Negative predictive value, %	96.2	95.8	95.9	97.1	95.5	96.1	0.6
Accuracy, %	95.3	97.0	95.9	97.7	96.4	96.5	1.0
AUC	0.990	0.996	0.993	0.998	0.994	0.994	0.3
AUC, the area under the receiver operating characteristic curve; SD, standard deviation							

## Highlighting areas with the risk of early recurrence by heatmap

A representative heatmap of the tumor area on a preoperative plain CT based on probability calculated using our prediction model is shown in Fig. 3. The heatmap can be superimposed on the input image to highlight the areas the model considers important in making its diagnosis. In short, the heatmap can contribute to assisting physicians before surgery by highlighting areas with the risk of early recurrence.

## Factors that can influence misprediction

We further investigated factors that can influence misprediction. Using our own model, we calculated the prediction accuracy of each case based on the patches in each one. In this study, we defined cases with prediction accuracy of the first quartile or less (96.0%), as occurrences of misprediction. Table 4 reveals a comparison of the perioperative characteristics of the cases with an accuracy of 96.0% or less and the cases with over 96.0% accuracy. Univariate analysis showed that smaller tumor size was a significant factor in misprediction ( $P = 0.025$ ). Likewise, looking at tumor size only, also showed a significant difference in the comparison of the two groups when divided by the median value of prediction accuracy (97.8%). Collectively, these data suggested that our prediction model can perform excellently, particularly with larger iCCA.

Table 4  
Comparison of perioperative characteristics depending on prediction accuracy.

	<b>Under 25%</b> <b>(n = 11)</b>	<b>Over 25%</b> <b>(n = 30)</b>	<b>P value</b>
Gender			0.159
Male, n	3 (27.3)	17 (56.7)	
Female, n	8 (72.7)	13 (43.3)	
Age, year	68 (60–77)	70 (39–81)	0.755
Body height, cm	154.0 (133.0-173.5)	158.5 (140.5–178.0)	0.333
Body weight, kg	51.0 (34.0-73.5)	55.3 (38.0-81.6)	0.547
Body mass index, kg/m <sup>2</sup>	22.1 (19.2–26.0)	21.7 (16.6–29.3)	0.937
Child-Pugh score			0.380
A, n	11 (100)	28 (93.3)	
B, n	0	2 (6.7)	
C, n	0	0	
CA19-9, U/mL	39.1 (5.0-3120.0)	179.0 (3.8-80355.0)	0.526
CEA, ng/mL	4.1 (0.5–43.1)	3.0 (0.5–61.6)	0.706
Surgical procedure			0.661
Right hepatectomy, n	3 (27.3)	9 (30.0)	
Left hepatectomy, n	5 (45.5)	13 (43.3)	
Right anterior sectionectomy, n	0	3 (10.0)	
Right posterior sectionectomy, n	0	1 (3.3)	
Partial resection, n	3 (27.3)	4 (13.3)	
Tumor factors			
Tumor size, mm	42.0 (16.0–60.0)	57.5 (17.0-150.0)	0.025
Solitary tumor, n	8 (72.7)	20 (66.7)	> 0.999
UICC 8th edition			
T category, n			0.482

CA19-9, carbohydrate antigen 19 – 9; CEA, carcinoembryonic antigen; UICC, Union for International Cancer Control

	<b>Under 25%</b> <b>(n = 11)</b>	<b>Over 25%</b> <b>(n = 30)</b>	<b>P value</b>
T1a	1 (9.1)	5 (16.7)	
T1b	0	3 (10.0)	
T2	8 (72.7)	16 (53.3)	
T3	2 (18.2)	3 (10.0)	
T4	0	3 (10.0)	
N category, n			> 0.999
N0	8 (72.7)	20 (66.7)	
N1	3 (27.3)	10 (33.3)	
M category, n			1.000
M0	11 (100)	30 (100)	
TNM Stage, n			0.406
ⅠA	1 (9.1)	2 (6.7)	
ⅠB	0	3 (10.0)	
Ⅱ	5 (45.5)	12 (40.0)	
ⅢA	2 (18.2)	1 (3.3)	
ⅢB	3 (27.3)	12 (40.0)	
Ⅳ	0	0	
Intrahepatic metastasis, n	3 (27.3)	9 (30.0)	> 0.999
Positive surgical margin, n	1 (9.1)	2 (6.7)	> 0.999
Early recurrence, n	4 (36.4)	16 (53.3)	0.484
Prediction accuracy, %	95.4 (85.6–96.0)	98.4 (96.2–100)	< 0.001
CA19-9, carbohydrate antigen 19 - 9; CEA, carcinoembryonic antigen; UICC, Union for International Cancer Control			

## Discussion

We applied the DL model to predict postoperative early recurrence of iCCA. We have successfully demonstrated high performance in the prediction of postoperative recurrence using plain preoperative CT images. The accuracy of the DL model far exceeded that of the binary logistic regression analysis (AUC,

0.994 vs. 0.770). This report represents the first study in which a DL model based on CT images is used to predict postoperative recurrence in iCCA. Our results may yield a novel insight into personalized treatment strategies, including neoadjuvant and adjuvant chemotherapy, in iCCA management.

Adjuvant chemotherapy is certainly expected to increase the survivorship of patients with iCCA<sup>19-21</sup>. Isolating results has been a challenge as past prospective randomized trials have included not only iCCA but also other bile duct cancers<sup>21-23</sup>. Furthermore, the indication for adjuvant chemotherapy in those studies was heterogeneity. In short, the selection criteria of susceptible individuals for adjuvant chemotherapy is not well established. To address this issue, Jeong et al. showed the usefulness of an AI framework in the prognostic estimation and stratification of susceptible individuals for adjuvant treatment after resection in iCCA patients. In contrast, we intended to predict postoperative recurrence directly. Our model, which can directly predict early recurrence, would be used to predict who should receive adjuvant chemotherapy based on their risk of recurrence.

Liang et al. conducted a single-center retrospective study and built a radiomics nomogram to predict early recurrence of iCCA after surgical resection<sup>24</sup>. Their nomogram, using preoperative arterial-phase contrast-enhanced magnetic resonance imaging (MRI), achieved an AUC of 0.82 and 0.77 in the training and validation cohorts, respectively. They used manual engineered features and selected the earliest recurrence-related features using a least absolute shrinkage and selection operator logistic regression analysis. Zhao et al. used radiomics from MRI to predict early recurrence. Their radiomics model showed a preferable predictive performance (AUC 0.889)<sup>25</sup>. Compared with the previous radiomics model using MRI, our model, which is based on DL features, achieved higher predictive performance (AUC 0.994).

Based on our results, which perform in such a highly predictive manner with the model addressing postoperative early recurrence, we propose a new concept in iCCA management. Though we need to discuss further which population, patients, those with or without early recurrence, is fit for adjuvant chemotherapy, achieving quite high levels of predictive accuracy, compared to conventional methods, can provide valuable information for determining adjuvant therapy and developing surgical plans, thereby facilitating pretreatment decisions.

Thanks to the advantages reaped from DL, we physicians, can easily apply computer-aided diagnosis<sup>16,26</sup>. Deep learning algorithms, such as CNN, have been widely used in the field of image diagnosis and prediction owing to their being fast, accurate, and reproducible<sup>26,27</sup>. CNN can uncover details in medical images that human experts cannot find, and automatically render a quantitative assessment<sup>28</sup>. Generally, even expert radiologists and surgeons cannot always access meaningful findings that would enable physicians to decide on a treatment strategy from plain CT images. In fact, there have been no reports or guidelines that recommend using plain CT images for risk assessment of postoperative recurrence in iCCA. Several lines of evidence, including our study, can lead to a paradigm shift in the recognition of AI in the field of iCCA treatment.

The present study has several limitations. This is a retrospective study. In addition, although this is a pilot study, the patient population was small. However, our model achieved high predictive performance. If we had access to additional training data from a large cohort, we could achieve even higher prediction accuracy and generality. To establish clinical applications, sufficient datasets are fundamental requirements. A novel AI approach based on analyzing a huge database, such as national or regional datasets, would be attractive to both clinicians treating iCCA and their patients. An accurate and robust prediction model can ultimately contribute to a better prognosis in iCCA patients.

There was the question of possible lack of homogeneity in CT techniques over the past 20 years that has been a point of contention (Supplemental Table 1). Nevertheless, our model achieved high predictive performance. These results suggested that relative heterogeneity of CT techniques may not be a big issue because of the handling of huge information from CT images through DL. Certainly, homogeneity of CT techniques would be preferable. However, it would not be practical in a real clinical setting for all patients to undergo CT exams using the same scanner and technique. In short, the use of a diverse set of CT acquisitions was not a limitation, it was a benefit to the study.

In conclusion, our DL model, using plain preoperative CT images of iCCA, exhibited high predictive performance in projecting postoperative early recurrence. The present multicenter study has provided a novel approach to predict recurrence after surgery. This model may help clinicians in the selection of patients for neoadjuvant and/or adjuvant therapy. This approach can contribute to personalized strategies in iCCA treatment. To establish a clinical application, conducting a study using a huge dataset, such as national dataset, is the hope for the future.

## Methods

### Patients and study design

This multi-institutional, retrospective, observational study was approved by the Committee of Medical Ethics of Hirosaki University Graduate School of Medicine (Aomori, Japan; reference no. 2020-230-1). Informed consent was obtained in the form of opt-out on our website (<https://www.med.hirosaki-u.ac.jp/hospital/outline/resarch/resarch.html>), with the approval of the Committee of Medical Ethics of Hirosaki University Graduate School of Medicine. Our study did not include minors. This study was designed and carried out in accordance with the Declaration of Helsinki.

Between January 2001 and December 2019, 41 patients undergoing liver surgery for iCCA at three institutions were included in this study. All patients had a confirmed pathologic diagnosis. In this study, the exclusion criteria were as follows: patients who had not undergone CT examination before surgery or patients with missing postoperative course data. Recurrence after surgery was diagnosed using enhanced CT and/or MRI. In this study, early recurrence was defined as recurrence within a year after liver surgery.

### CT acquisition and tumor segmentation

Our workflow is shown in Fig. 1. Preoperative axial plain CT images for each case were obtained from multiple institutions and were used for this study. Radiological assessment was performed by board-certified radiologists who were blind to the outcome of the patients. Board-certified surgeons and medical students performed CT acquisition and tumor segmentation based on the radiological assessment. Using a commercial viewer (ShadeQuest/ViewR, Fujifilm, Japan), the CT image showing the largest tumor area was selected. The entire tumor region was manually segmented with Adobe Illustrator and saved.

## Preparation of dataset

We trimmed a patch with a size of 128 x 128 pixels with 32 pixel stride from the entire segmented tumor area. Finally, 71,081 patches were obtained from 41 patients in the current study.

## Architecture of the CNN

ResNet50<sup>29</sup> and Pytorch (a python library) were utilized (Available at:<https://github.com/pytorch/pytorch>). We did not use a pretrained model. The original acquired images of 128 x 128 pixels were converted into images of 224 x 224 pixels. We tuned the hyperparameters as follows: number of training epochs, 50; batch size, 128; learning rate, 0.00025 via trial and error; and number of outer layers, 2 classes.

## Evaluation methods

We used cross-validation to obtain more accurate results with less bias in the machine learning studies<sup>30</sup>. In this study, the dataset is split into five folds, one fold of which is for validation and the other folds are for training. The proportion of patients with versus without early recurrence was equal in each fold. The training and validation processes were repeated five times using different folds each time. The final results were then averaged and the standard deviation was calculated. The accuracy, sensitivity, specificity, positive predictive values, and negative predictive values were evaluated. The model was also evaluated using the area under the receiver operating characteristic (ROC) curve (AUC).

## Heatmap

The probability of early recurrence of the patches calculated by the trained model over the entire tumor area was modulated in gray scale from 0 to 255, which was assigned to pseudo-coloring; blue for a low-risk patch and red for a high-risk one on CT imaging. Heatmaps were generated by applyColorMap in OpenCV to visualize the most indicative region for the risk of early recurrence. Prediction of the probability of early recurrence was visualized by pseudo-coloring; COLORMAP\_JET was applied to the grayscale CT images.

## Statistical analyses

Continuous variables were expressed as medians (ranges) and analyzed using nonparametric methods for non-normally distributed data (Mann–Whitney U-test). Categorical variables were reported as numbers (percentages) and analyzed using the chi-squared test or Fisher's exact test, as appropriate. Variables

with a significant relationship to early recurrence in univariate analysis were used in a binary logistic regression model. A difference was considered to be significant for values of  $P < 0.05$ . The statistical analyses were performed using IBM SPSS Statistics for Windows, Version 26.0 (IBM Corp, Armonk, NY, USA).

## **Declarations**

### **Availability of data and materials**

The data generated or analysed during this study are included in this published article and its supplementary information files. Some datasets generated and/or analysed during the current study are not publicly available due to privacy but are available from the corresponding author on reasonable request.

### **Acknowledgments**

We sincerely thank Shari Joy Berman for professionally editing the English draft of this manuscript.

### **Author contributions statement**

TW contributed to the study conception and design. TW, KI, NK, HN, TS, YU, YW and YT performed surgical resection and collected the clinical data. TW, TK, SI, KS, SK and HF performed CT acquisition and tumor segmentation. YS established the prediction model. TW wrote the first draft of the manuscript paper. MM, YS and KH contributed to the review, and/or critical revision of the manuscript. All authors have approved the final article.

### **Competing interests**

The authors declare no competing interests.

The corresponding author is responsible for submitting a competing interests statement on behalf of all authors of the paper.

### **Funding**

There was no funding for this study.

### **Ethics approval and consent to participate**

This study was approved by the Committee of Medical Ethics of Hirosaki University Graduate School of Medicine (Aomori, Japan; reference no. 2020-230-1). Informed consent was obtained in the form of opt-out on our website (<https://www.med.hirosaki-u.ac.jp/hospital/outline/resarch/resarch.html>), with the approval of the Committee of Medical Ethics of Hirosaki University Graduate School of Medicine. This study was designed and carried out in accordance with the Declaration of Helsinki.

## Consent for publication

Informed consent was obtained in the form of opt-out on our website (<https://www.med.hirosaki-u.ac.jp/hospital/outline/resarch/resarch.html>), with the approval of the Committee of Medical Ethics of Hirosaki University Graduate School of Medicine.

## References

1. Sung, H. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* **n/a**, doi:<https://doi.org/10.3322/caac.21660>.
2. Aljiffry, M. *et al.* Evidence-based approach to cholangiocarcinoma: a systematic review of the current literature. *J Am Coll Surg* **208**, 134–147, doi:10.1016/j.jamcollsurg.2008.09.007 (2009).
3. Bridgewater, J. *et al.* Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. *J Hepatol* **60**, 1268–1289, doi:10.1016/j.jhep.2014.01.021 (2014).
4. Society, A. C. *Bile Duct Cancer Survival Rates | Cholangiocarcinoma Survival Rates*, <<https://www.cancer.org/cancer/bile-duct-cancer/detection-diagnosis-staging/survival-by-stage.html>> (2021).
5. Endo, I. *et al.* Intrahepatic cholangiocarcinoma: rising frequency, improved survival, and determinants of outcome after resection. *Ann Surg* **248**, 84–96, doi:10.1097/SLA.0b013e318176c4d3 (2008).
6. Tan, J. C., Coburn, N. G., Baxter, N. N., Kiss, A. & Law, C. H. Surgical management of intrahepatic cholangiocarcinoma—a population-based study. *Ann Surg Oncol* **15**, 600–608, doi:10.1245/s10434-007-9627-x (2008).
7. Choi, S. B. *et al.* The prognosis and survival outcome of intrahepatic cholangiocarcinoma following surgical resection: association of lymph node metastasis and lymph node dissection with survival. *Ann Surg Oncol* **16**, 3048–3056, doi:10.1245/s10434-009-0631-1 (2009).
8. Yamamoto, M., Takasaki, K., Otsubo, T., Katsuragawa, H. & Katagiri, S. Recurrence after surgical resection of intrahepatic cholangiocarcinoma. *J Hepatobiliary Pancreat Surg* **8**, 154–157, doi:10.1007/s005340170039 (2001).
9. Tsilimigras, D. I. *et al.* Very Early Recurrence After Liver Resection for Intrahepatic Cholangiocarcinoma: Considering Alternative Treatment Approaches. *JAMA surgery* **155**, 823–831, doi:10.1001/jamasurg.2020.1973 (2020).
10. Ma, K. W. *et al.* Diagnostic and Prognostic Role of 18-FDG PET/CT in the Management of Resectable Biliary Tract Cancer. *World J Surg* **42**, 823–834, doi:10.1007/s00268-017-4192-3 (2018).
11. Kochanny, S. E. & Pearson, A. T. Academics as leaders in the cancer artificial intelligence revolution. *Cancer* **127**, 664–671, doi:<https://doi.org/10.1002/cncr.33284> (2021).
12. Kermany, D. S. *et al.* Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* **172**, 1122–1131.e1129, doi:10.1016/j.cell.2018.02.010 (2018).

13. Yasaka, K., Akai, H., Kunimatsu, A., Abe, O. & Kiryu, S. Liver Fibrosis: Deep Convolutional Neural Network for Staging by Using Gadoteric Acid-enhanced Hepatobiliary Phase MR Images. *Radiology* **287**, 146–155, doi:10.1148/radiol.2017171928 (2018).
14. Tonzuka, R. *et al.* Deep learning analysis for the detection of pancreatic cancer on endosonographic images: a pilot study. *Journal of Hepato-Biliary-Pancreatic Sciences* **28**, 95–104, doi:https://doi.org/10.1002/jhbp.825 (2021).
15. Ji, G. W. *et al.* Radiomic Features at Contrast-enhanced CT Predict Recurrence in Early Stage Hepatocellular Carcinoma: A Multi-Institutional Study. *Radiology* **294**, 568–579, doi:10.1148/radiol.2020191470 (2020).
16. Igarashi, S., Sasaki, Y., Mikami, T., Sakuraba, H. & Fukuda, S. Anatomical classification of upper gastrointestinal organs under various image capture conditions using AlexNet. *Comput Biol Med* **124**, 103950, doi:10.1016/j.combiomed.2020.103950 (2020).
17. Hamm, C. A. *et al.* Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI. *European Radiology* **29**, 3338–3347, doi:10.1007/s00330-019-06205-9 (2019).
18. Jeong, S. *et al.* Latent Risk Intrahepatic Cholangiocarcinoma Susceptible to Adjuvant Treatment After Resection: A Clinical Deep Learning Approach. *Frontiers in oncology* **10**, 143, doi:10.3389/fonc.2020.00143 (2020).
19. Ma, K. W. *et al.* Adjuvant chemotherapy improves oncological outcomes of resectable intrahepatic cholangiocarcinoma: A meta-analysis. *Medicine (Baltimore)* **98**, e14013, doi:10.1097/md.00000000000014013 (2019).
20. Shroff, R. T. *et al.* Adjuvant Therapy for Resected Biliary Tract Cancer: ASCO Clinical Practice Guideline. *J Clin Oncol* **37**, 1015–1027, doi:10.1200/jco.18.02178 (2019).
21. Primrose, J. N. *et al.* Capecitabine compared with observation in resected biliary tract cancer (BILCAP): a randomised, controlled, multicentre, phase 3 study. *Lancet Oncol* **20**, 663–673, doi:10.1016/s1470-2045(18)30915-x (2019).
22. Ebata, T. *et al.* Randomized clinical trial of adjuvant gemcitabine chemotherapy versus observation in resected bile duct cancer. *Br J Surg* **105**, 192–202, doi:10.1002/bjs.10776 (2018).
23. Edeline, J. *et al.* Gemcitabine and Oxaliplatin Chemotherapy or Surveillance in Resected Biliary Tract Cancer (PRODIGE 12-ACCORD 18-UNICANCER GI): A Randomized Phase III Study. *J Clin Oncol* **37**, 658–667, doi:10.1200/jco.18.00050 (2019).
24. Liang, W. *et al.* Novel Nomogram for Preoperative Prediction of Early Recurrence in Intrahepatic Cholangiocarcinoma. *Frontiers in oncology* **8**, 360, doi:10.3389/fonc.2018.00360 (2018).
25. Zhao, L. *et al.* Prediction for early recurrence of intrahepatic mass-forming cholangiocarcinoma: quantitative magnetic resonance imaging combined with prognostic immunohistochemical markers. *Cancer imaging: the official publication of the International Cancer Imaging Society* **19**, 49, doi:10.1186/s40644-019-0234-4 (2019).

26. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444, doi:10.1038/nature14539 (2015).
27. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H. & Aerts, H. Artificial intelligence in radiology. *Nat Rev Cancer* **18**, 500–510, doi:10.1038/s41568-018-0016-5 (2018).
28. Napel, S., Mu, W., Jardim-Perassi, B. V., Aerts, H. J. W. L. & Gillies, R. J. Quantitative imaging of cancer in the postgenomic era: Radio(geno)mics, deep learning, and habitats. *Cancer* **124**, 4633–4649, doi:https://doi.org/10.1002/cncr.31630 (2018).
29. He, K., Zhang, X., Ren, S. & Sun, J. in 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778.
30. Ojala, M. & Garriga, G. C. Permutation Tests for Studying Classifier Performance. *J. Mach. Learn. Res.* **11**, 1833–1863 (2010).

## Figures

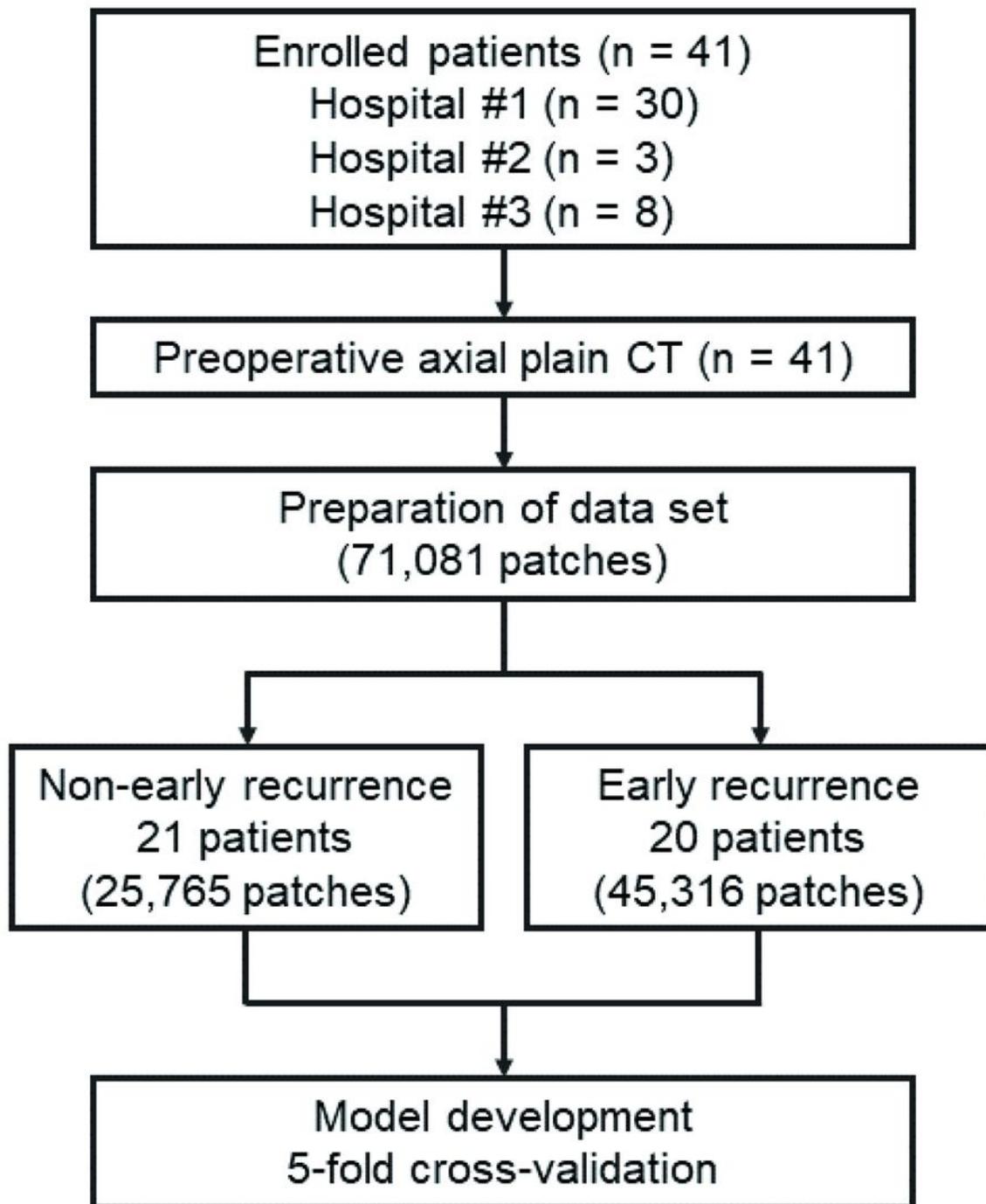


Figure 1

The study workflow and methodological process.

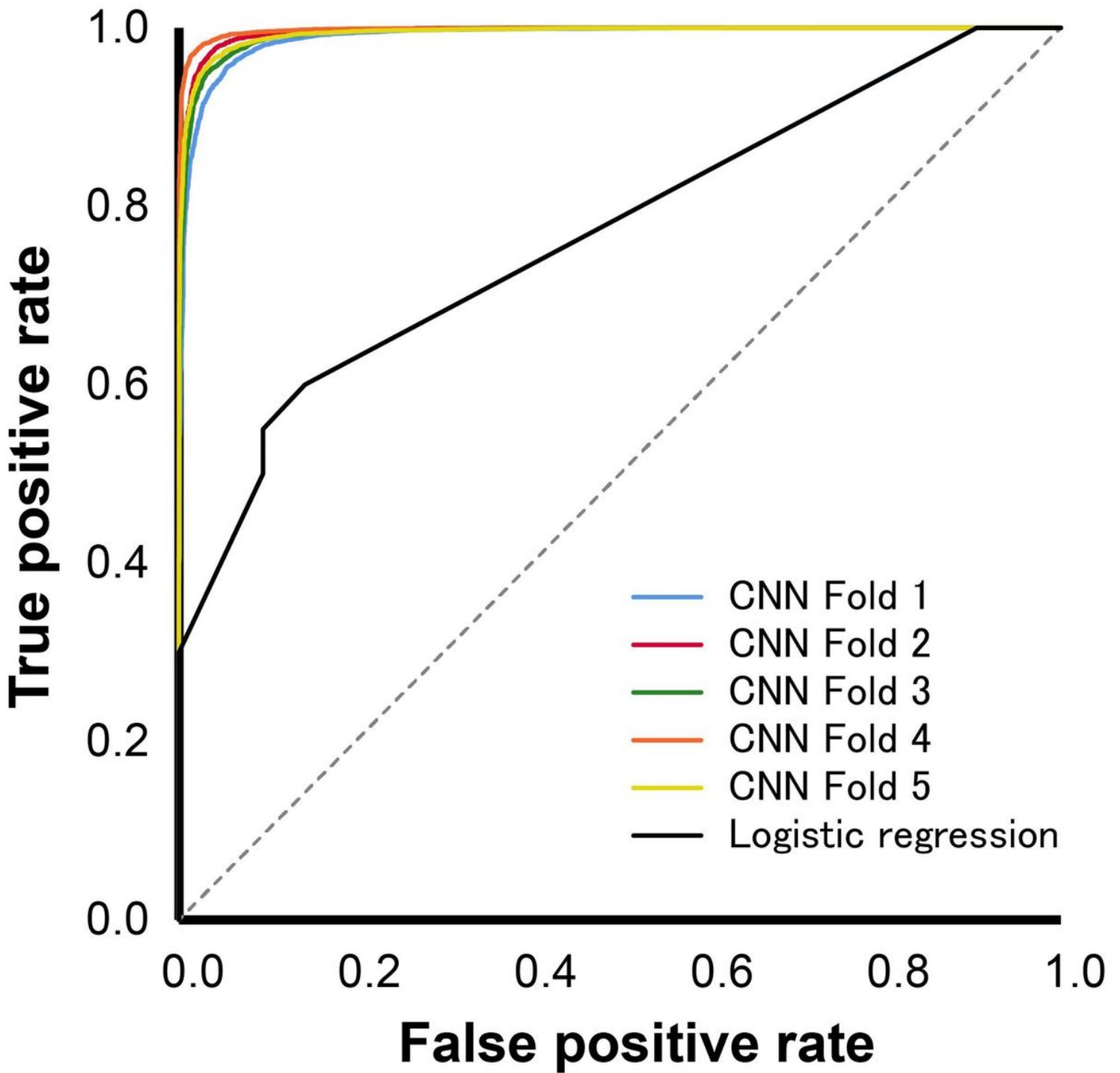


Figure 2

The receiver operating characteristics curves of logistic regression analysis and the DL model.

ROC curves show the performance of logistic regression analysis and the ResNet model in the validation dataset in detecting early recurrence. The AUC of logistic regression analysis is 0.770, and the average AUC of the convolutional neural network (CNN) model is 0.994.

### Figure 3

A heatmap of iCCA on a preoperative plain CT using our prediction model.

The color bar illustrates the degree of probability the model paid to it. Red areas represent a high risk of early recurrence; blue areas represent a low risk of early recurrence. (A), original image of a non-early recurrence case; (B), original image of an early recurrence case; (C), heatmap of a non-early recurrence case; (D), heatmap of an early recurrence case.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarytable1.docx](#)
- [S1File.txt](#)