



Article

# Survival Prediction in Intrahepatic Cholangiocarcinoma: A Proof of Concept Study Using Artificial Intelligence for Risk Assessment

Lukas Müller<sup>1</sup>, Aline Mähringer-Kunz<sup>1</sup>, Simon Johannes Gairing<sup>2</sup> , Friedrich Foerster<sup>2</sup>, Arndt Weinmann<sup>2</sup> , Fabian Bartsch<sup>3</sup> , Lisa-Katharina Heuft<sup>3</sup> , Janine Baumgart<sup>3</sup>, Christoph Düber<sup>1</sup>, Felix Hahn<sup>1,†</sup> and Roman Kloeckner<sup>1,\*,†</sup>

<sup>1</sup> Department of Diagnostic and Interventional Radiology, University Medical Center of the Johannes Gutenberg University Mainz, 55131 Mainz, Germany; lukas.mueller@unimedizin-mainz.de (L.M.); aline.maehringer-kunz@unimedizin-mainz.de (A.M.-K.); christoph.dueber@unimedizin-mainz.de (C.D.); felix.hahn@unimedizin-mainz.de (F.H.)

<sup>2</sup> Department of Internal Medicine, University Medical Center of the Johannes Gutenberg University Mainz, 55131 Mainz, Germany; simonjohannes.gairing@unimedizin-mainz.de (S.J.G.); friedrich.foerster@unimedizin-mainz.de (F.F.); arndt.weinmann@unimedizin-mainz.de (A.W.)

<sup>3</sup> Department of General, Visceral and Transplant Surgery, University Medical Center of the Johannes Gutenberg University Mainz, 55131 Mainz, Germany; fabian.bartsch@unimedizin-mainz.de (F.B.); lisa-katharina.heuft@unimedizin-mainz.de (L.-K.H.); janine.baumgart@unimedizin-mainz.de (J.B.)

\* Correspondence: roman.kloeckner@unimedizin-mainz.de; Tel.: +49-6131-17-2019

† Authors contributed equally to this work.



Citation: Müller, L.;

Mähringer-Kunz, A.; Gairing, S.J.; Foerster, F.; Weinmann, A.; Bartsch, F.; Heuft, L.-K.; Baumgart, J.; Düber, C.; Hahn, F.; et al. Survival Prediction in Intrahepatic Cholangiocarcinoma: A Proof of Concept Study Using Artificial Intelligence for Risk Assessment. *J. Clin. Med.* **2021**, *10*, 2071. <https://doi.org/10.3390/jcm10102071>

Academic Editors: Hauke Lang and Do Hyun Park

Received: 24 March 2021

Accepted: 10 May 2021

Published: 12 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Abstract:** Several scoring systems have been devised to objectively predict survival for patients with intrahepatic cholangiocellular carcinoma (ICC) and support treatment stratification, but they have failed external validation. The aim of the present study was to improve prognostication using an artificial intelligence-based approach. We retrospectively identified 417 patients with ICC who were referred to our tertiary care center between 1997 and 2018. Of these, 293 met the inclusion criteria. Established risk factors served as input nodes for an artificial neural network (ANN). We compared the performance of the trained model to the most widely used conventional scoring system, the Fudan score. Predicting 1-year survival, the ANN reached an area under the ROC curve (AUC) of 0.89 for the training set and 0.80 for the validation set. The AUC of the Fudan score was significantly lower in the validation set (0.77,  $p < 0.001$ ). In the training set, the Fudan score yielded a lower AUC (0.74) without reaching significance ( $p = 0.24$ ). Thus, ANNs incorporating a multitude of known risk factors can outperform conventional risk scores, which typically consist of a limited number of parameters. In the future, such artificial intelligence-based approaches have the potential to improve treatment stratification when models trained on large multicenter data are openly available.

**Keywords:** intrahepatic cholangiocarcinoma; survival prediction; risk scoring; machine learning; artificial intelligence; artificial neural network; Fudan score



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Intrahepatic cholangiocarcinoma (ICC) is the second most common type of primary liver cancer after hepatocellular carcinoma (HCC). The incidence of ICC is low in Western countries but has been rising continuously in recent decades [1–4]. Unfortunately, symptoms of ICC mostly appear in the late stages of the disease. Thus, resection, which is the only curative treatment option, is not possible in the majority of cases [5]. In addition, recurrence rates after initial resection exceed 60% [6]. Novel treatment options have become available in recent decades, and knowledge on prognostic factors is growing [7,8]. This is allowing treatment during the course of disease to be more individualized. Due to this growing heterogeneity, risk prediction is becoming more and more difficult.

Conventional scoring models for risk stratification have been proposed by several groups [9–11]. Most of them were designed primarily for patients undergoing curative resection and use histopathological factors, such as microvascular invasion or tumor grading, which are only available postoperatively [9–11]. Even though all attempts have initially shown promising results, they have failed external validation and have not entered clinical use [12,13]. The only available score for all patients regardless of subsequent treatment is the Fudan score [14]. The tumor itself plays a major role in this score, which comprises tumor diameter, number of lesions, tumor boundary, level of tumor marker carbohydrate antigen 19-9 (CA19-9), and serum alkaline phosphatase (AP) level. All of these parameters are easily assessable during the initial patient work-up. Thus, the score provides an *ab initio* method for assisting clinicians in patient stratification. However, the score has never been externally evaluated for patients with ICC regardless of the initial therapy.

All of the conventional scoring approaches are easy to calculate and may be comprehensible, but it remains questionable whether such a limited number of parameters is sufficient to achieve reliable prediction for clinical decision making.

An alternative to conventional scoring systems is the increasing integration of machine learning (ML) approaches into risk assessment. Systems based on ML have proven their feasibility and superiority compared to conventional scoring systems in survival prediction for hepatocellular and colorectal cancer [15–17]. Thus far, for ICC, a few similar approaches have been tried for the subgroup of resected patients in order to calculate the risk of recurrence, to decide upon adjuvant treatment, and to predict the median overall survival (OS) [18–20]. For these decisions, such approaches outperformed the conventional scoring systems.

We hypothesize that the main reason for the superiority of ML algorithms over conventional approaches is based on the possibility of including a wider range of parameters. In particular, artificial neural networks (ANNs) are ideal to include a wide range of different parameters and offer flexible scalability when complexity increases [15].

Thus, this study attempted to build an ANN based on a much broader range of parameters in order to improve prediction for patients with ICC prior to making decisions on treatment. In a second step, we evaluated our newly designed model against the conventional Fudan score in a head-to-head comparison.

## 2. Materials and Methods

The study was approved by the responsible ethics committee (permit number 2018—13618, date of approval: 15 October 2018). Patient records and clinical information were de-identified before analysis. Additional examinations were not performed. The TRIPOD and STROBE guidelines were followed for the construction of the manuscript (Supplementary Tables S1 and S2) [21,22].

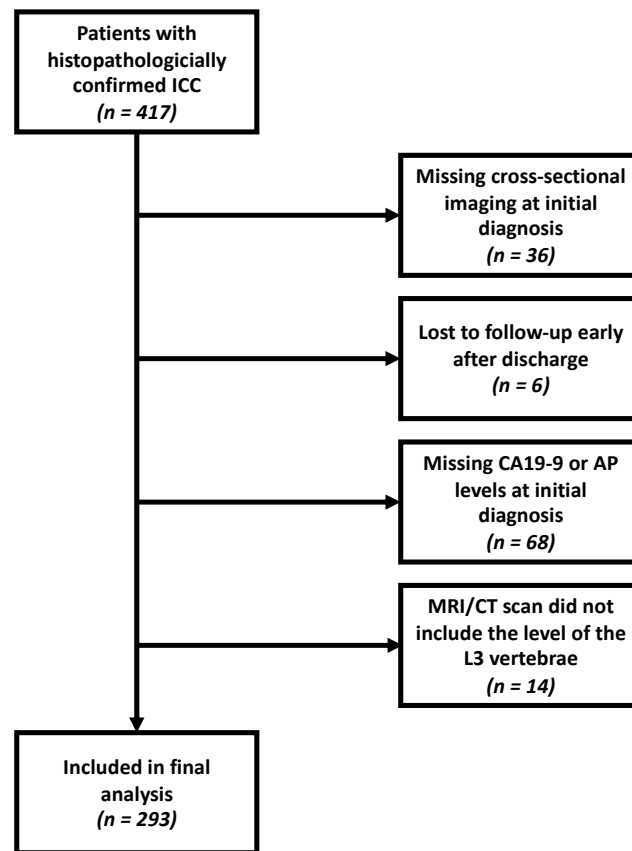
### 2.1. Patients

Between January 1997 and January 2018, 417 patients with histopathologically confirmed ICC were referred to our tertiary care center. After retrospectively identifying these patients using established clinical registry software, 124 were excluded for the reasons described in Figure 1. The final analysis was performed on the remaining 293 patients.

### 2.2. Diagnosis, Treatment and Follow-Up

Histopathological diagnosis was performed based on the European Association for the Study of the Liver guidelines for the diagnosis and management of ICC [7]. All patients underwent contrast-enhanced computed tomography (CT) or magnetic resonance imaging (MRI) for treatment planning and staging. Prior to making a treatment decision, all patients underwent an extensive discussion with an interdisciplinary tumor board consisting of visceral surgeons, hepatologists/oncologists, diagnostic and interventional

radiologists, pathologists, and, if needed, radiation therapists. Follow-up comprised clinical examination, blood sampling, and cross-sectional imaging.



**Figure 1.** Flow diagram showing the reasons for exclusion from the study. CA19-9, carbohydrate antigen 19-9. AP, alkaline phosphatase. MRI, magnetic resonance imaging. CT, computed tomography.

### 2.3. Data Acquisition

Patient data were acquired using the clinical registry unit (CRU). The CRU is an established registry that prospectively collects all patients with liver cancer treated at our tertiary care referral center [23]. The data for this study were retrospectively collected and analyzed. The CRU dataset includes all baseline characteristics, including demographic data, serological parameters, treatment-related parameters, and information on the tumor burden, including size and number of intrahepatic lesions, tumor boundary type, translobar and extrahepatic spread, and the presence of nodal and distant metastases. Standardized cut-offs for the serological and imaging parameters were derived from the original Fudan score [14]. In particular, the tumor boundary was assessed as described in the original paper [14]. Translobar spread was specified as tumor expansion per continuitatem or as intrahepatic metastasis in more than one lobe. According to the current AJCC/UICC TNM staging system, an extrahepatic spread exists if the tumor perforates the viscera of the liver and/or infiltrates adjacent organs [24]. The psoas muscle index (PMI) was defined as the total area of the psoas muscle at the level of the L3 vertebra divided by the squared body height [25,26]. For the definition of high and low PMI, we used cut-offs derived previously by our group using optimal stratification. In the resected group, “low” was defined as  $\leq 5.7 \text{ cm}^2/\text{m}^2$  for men and  $\leq 5.1 \text{ cm}^2/\text{m}^2$  for women, whereas in the non-resected subgroup, the values were  $\leq 5.5 \text{ cm}^2/\text{m}^2$  for men and  $\leq 4.8 \text{ cm}^2/\text{m}^2$  for women [25]. In the case of missing data, the information was updated using the radiology information system and the laboratory database. The primary endpoints were median OS and the 1-year survival rate. OS was defined as the time interval between the initial diagnosis and death or last

follow-up. Death dates were acquired and updated with information from the appropriate Residents' Registration Offices.

2.4. Calculation of the Fudan Score

The Fudan score was calculated as described in its original publication [14]. Figure 2 summarizes the included parameters, their weights, and the grouping used for risk stratification.

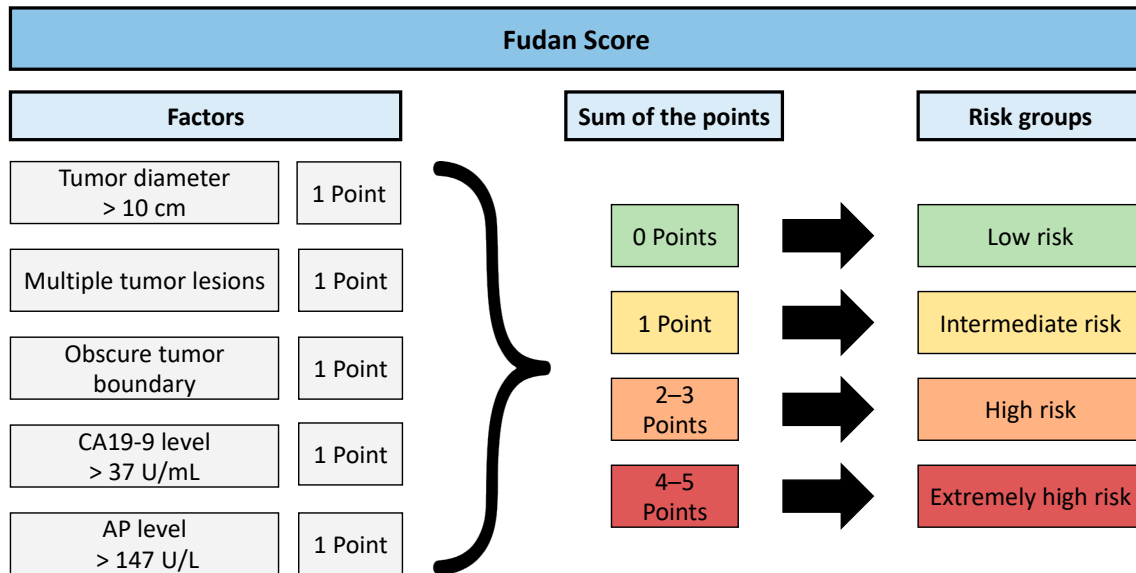


Figure 2. Calculation of the Fudan score. CA19-9, carbohydrate antigen 19-9. AP, alkaline phosphatase.

2.5. Design of the Neural Network

The neural network was built using Tensorflow (<https://www.tensorflow.org/>, version 1.13.0, Google LLC, Mountain View, USA, accessed on 31 January 2021) and Keras (<https://keras.io/>, version 2.2.0, Francois Chollet, Google LLC, Mountain View, CA, USA, accessed on 31 January 2021). It consisted of three fully connected hidden layers with 16, 12, and 8 nodes, respectively. To simplify, each of the hidden layers is a specific, complex mathematical function with different functional characteristics and designed to produce a defined output. By the conjunction of each defined output from each layer, a neural network can make a specific, overall prediction [27]. Rectified linear unit (ReLU) was used as the activation function on all hidden layers and sigmoid classification for the final output layer. To prevent overfitting, we used L2-regularization. Standardization was performed on all input parameters by subtraction of the mean and division by the standard deviation.

As input nodes, we included all factors of the Fudan score (tumor diameter, number of lesions, tumor boundary, CA19-9 and AP serum levels) as well as potentially meaningful parameters (tumor spread, extrahepatic tumor extension, the presence of lymph node and distant metastases). Furthermore, we included a low PMI as a parameter representing the patient's overall condition and the albumin level as a parameter representing the hepatic reserve. The final output results for the network were survival and death one year after initial diagnosis. The ANN is visualized in Figure 3.

2.6. Training and Validation of the ANN

For an 80:20 split, all patients with an initial diagnosis before 31 December 2013 ( $n = 233, 80\%$ ) were allocated to the training set. Patients with an initial diagnosis afterwards ( $n = 60, 20\%$ ) formed the holdout validation set. As suggested elsewhere, the holdout validation dataset was only used for final evaluation of the models and their comparison [15]. In the training set, a five-fold cross-validation approach was used to maximize

the training capabilities of the ANN. Figure 4 provides an overview on the process used for model training and validation.

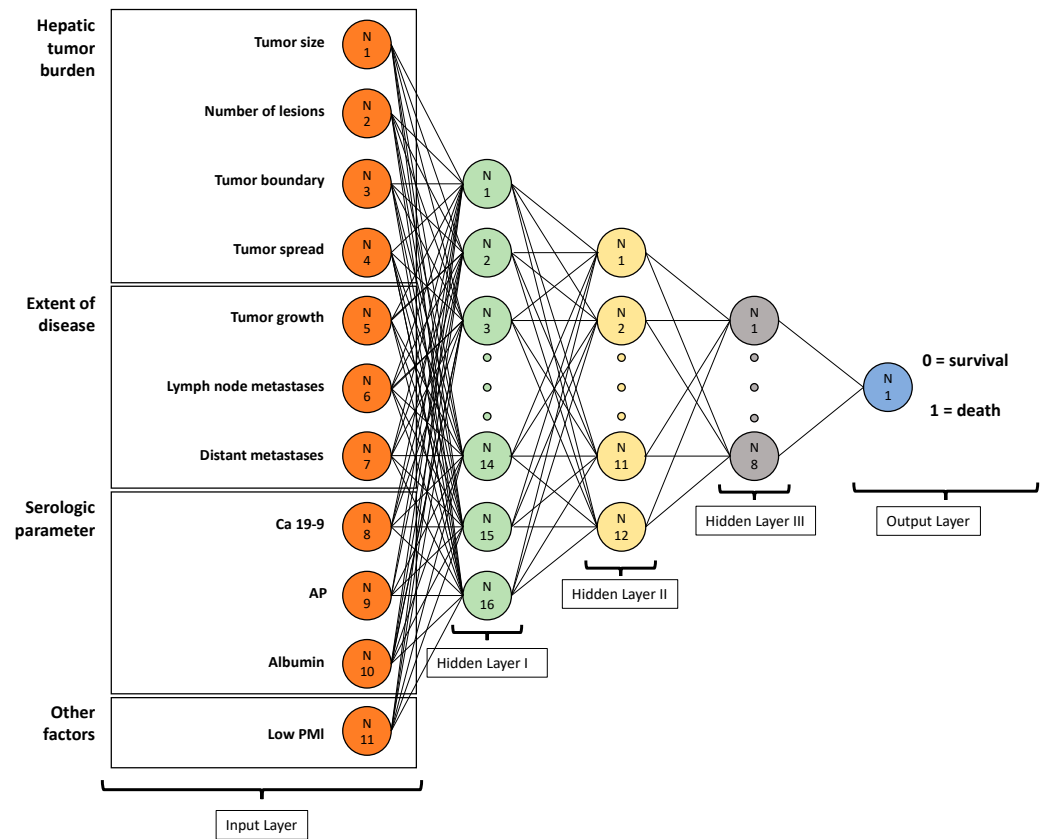


Figure 3. Visualization of the created artificial neural network.

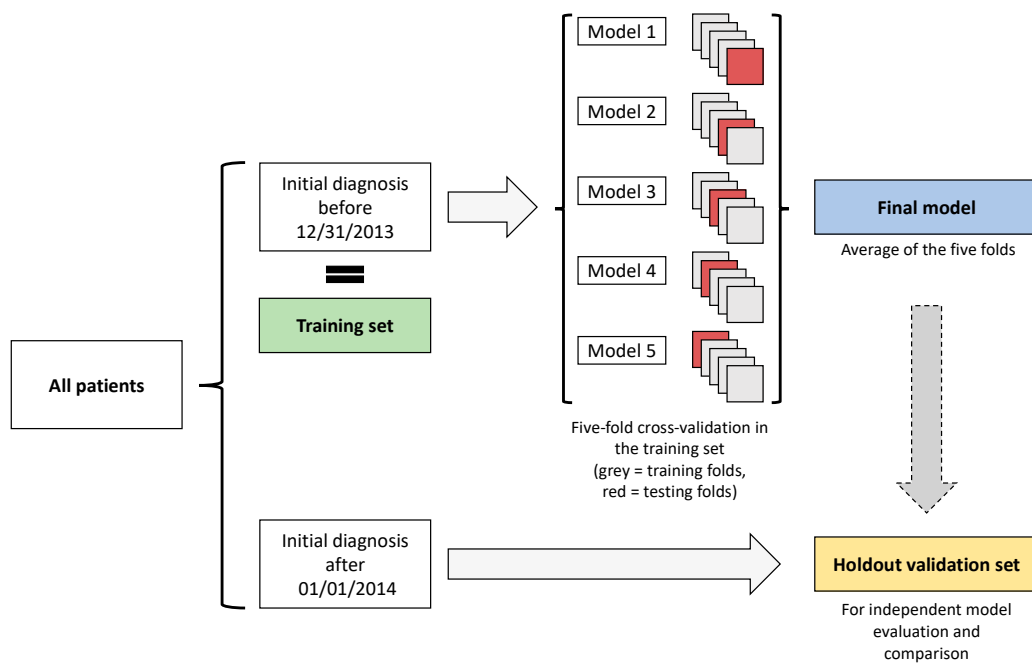


Figure 4. Visualization of the created artificial neural network.

### 2.7. Statistical Analysis

Statistical analyses and graphic design were performed in R 4.0.3 (A Language and Environment for Statistical Computing, <http://www.R-project.org>, R Foundation for Statistical Computing, Vienna, Austria, accessed on 31 January 2021). Continuous data were reported as medians and ranges. Categorical and binary baseline parameters were reported as absolute numbers and percentages. Fisher's exact tests, chi-squared tests, or Mann–Whitney *U* tests were used for *p*-value computations between the training and test sets, where appropriate. Survival analysis was performed using the packages “survminer” (<https://cran.r-project.org/package=survminer>, accessed on 31 January 2021, R Foundation for Statistical Computing, Vienna, Austria) and “survival” (<https://CRAN.R-project.org/package=survival>, accessed on 31 January 2021, R Foundation for Statistical Computing, Vienna, Austria). Strata were compared by log-rank testing. Univariate and multivariate Cox proportional hazard regression models assessing hazard ratios (HRs) and corresponding 95% confidence intervals (CIs) were performed to determine the influence of risk factors on the median OS. Performance of the Fudan score in individual survival prediction was assessed using Harrell's concordance index (C-Index) [28]. A C-Index of 0.5 indicates no predictive ability and 1.0 indicates perfect predictive power. The performance of the Fudan score and the ANN model for predicting the 1-year survival rate was measured using the area under the receiver operating characteristic curve (AUC). The AUC ranges from 0 to 1: 0.5 indicates no predictive ability, 1.0 indicates perfect prediction, and <0.5 indicates “anti-prediction”. A *p*-value of <0.05 was considered significant.

## 3. Results

### 3.1. Baseline Characteristics

Of the 293 patients analyzed in this study, 176 (60.1%) were males and 117 (39.9%) were females. The median age at the initial TACE treatment was 66 years. Median follow-up for all patients was 12.6 months. Both the training and the validation set had no statistical differences in their baseline characteristics. Median OS of the patients in the training set was 13.1 months (95% CI 10.1–16.7 months) and 16.3 months for patients in the validation set (95% CI 11.1–22.8 months). Table 1 displays the baseline characteristics of the cohort.

**Table 1.** Baseline characteristics of the patient cohort.

		All ( <i>n</i> = 293)	Training Set ( <i>n</i> = 233)	Validation Set ( <i>n</i> = 60)	<i>p</i> -Value
Age, years	Median (IQR)	66.0 (57–73)	66.1 (57–73)	65.4 (57–73)	0.79 †
Sex, <i>n</i> (%)	Male	176 (60.1)	143 (61.4)	33 (55.0)	0.38 ‡
	Female	117 (39.9)	90 (38.6)	27 (45.0)	
Number of intrahepatic lesions, <i>n</i> (%)	1	174 (59.4)	135 (57.9)	39 (65.0)	0.07 ††
	2	30 (10.2)	28 (12.0)	2 (3.3)	
	3	14 (4.8)	14 (6.0)	0 (0.0)	
	4	14 (4.8)	10 (4.3)	4 (6.7)	
	≥5	61 (20.8)	46 (19.8)	15 (25.0)	
Tumor size, mm	Median (IQR)	89 (56–146)	88 (56–145)	98 (55–153)	0.90 †
Tumor boundary type, <i>n</i> (%)	Distinct	105 (35.8)	88 (37.8)	17 (28.3)	0.23 ‡
	Obscure	188 (64.2)	145 (62.2)	43 (71.7)	
Tumor spread, <i>n</i> (%)	Unifocal or intra-lobar metastasis	206 (70.3)	161 (69.1)	45 (75.0)	0.43 ‡
	Translobar metastasis	87 (29.7)	72 (30.1)	15 (25.0)	
UICC T stage ≥ 3, <i>n</i> (%)	Yes	64 (21.8)	51 (21.9)	13 (21.7)	0.58 ‡
	No	229 (78.2)	182 (78.1)	47 (78.3)	
Lymph node metastases, <i>n</i> (%)	Yes	88 (30.0)	70 (30.0)	18 (30.0)	1.00 ‡
	No	205 (70.0)	163 (70.0)	42 (70.0)	

Table 1. Cont.

		All (n = 293)	Training Set (n = 233)	Validation Set (n = 60)	p-Value
Distant metastases, n (%)	Yes	74 (25.3)	57 (24.5)	17 (28.3)	0.62 ‡
	No	219 (74.7)	176 (75.5)	43 (71.7)	
AP serum levels, U/L	Median (IQR)	161 (102–290)	158 (99–306)	168 (116–256)	0.50 †
Ca 19-9 serum levels, U/mL	Median (IQR)	80 (22–800)	82 (18–773)	70 (31–1046)	0.46 †
Albumin, g/dL	Median (IQR)	3.8 (3.4–4.2)	3.9 (3.4–4.2)	3.8 (3.4–4.1)	0.29 †
Initial therapy	Resection	143 (48.8)	116 (49.8)	27 (45.0)	0.19 ††
	Ablation	3 (1.0)	1 (0.4)	2 (3.3)	
	TACE *	14 (4.8)	9 (3.9)	5 (8.3)	
	SIRT *	29 (9.9)	24 (10.3)	5 (8.3)	
	Chemotherapy only	54 (18.4)	41 (17.6)	13 (21.7)	
	BSC	50 (17.1)	42 (18.0)	8 (13.3)	

\* Of the 43 patients who received transarterial treatments, 20 received additional chemotherapy (n = 12 in the training set, n = 8 in the validation set). UICC, union internationale contre le cancer. CA19-9, carbohydrate antigen 19-9. AP, alkaline phosphatase. † Mann–Whitney U test used. ‡ Fisher test used. †† Chi-squared test used.

### 3.2. Risk Factor Identification for the ANN-Based Model

To identify possible risk factors for inclusion in the ANN model, univariate Cox hazard regression was performed. Except for age > 60 years, a parameter which is included in the MEGNA score [11], all investigated risk factors reached highly significant p-values (Table 2). Therefore, all of these factors were used in the input layer of the ANN model.

Table 2. Univariate Cox hazard regression model results.

Factor	Univariate	
	HR (95% CI)	p-Value
Age > 60 years	1.2 (0.9–1.6)	0.140
Max. tumor size > 10 cm	1.9 (1.5–2.5)	<0.001
Multifocality	2.0 (1.6–2.6)	<0.001
Obscure tumor boundary	2.4 (1.8–3.2)	<0.001
Translobar spread	2.9 (2.2–3.8)	<0.001
Extrahepatic tumor growth	1.6 (1.2–2.2)	<0.001
Lymph node metastases	2.1 (1.6–2.7)	<0.001
Distant metastases	4.2 (3.1–5.7)	<0.001
Ca 19-9 > 37 U/mL	2.2 (1.7–2.9)	<0.001
AP > 147 U/L	2.0 (1.5–2.5)	<0.001
Albumin < 3.5 g/dL	2.6 (2.0–3.5)	<0.001
Low PMI	1.6 (1.2–2.0)	<0.001

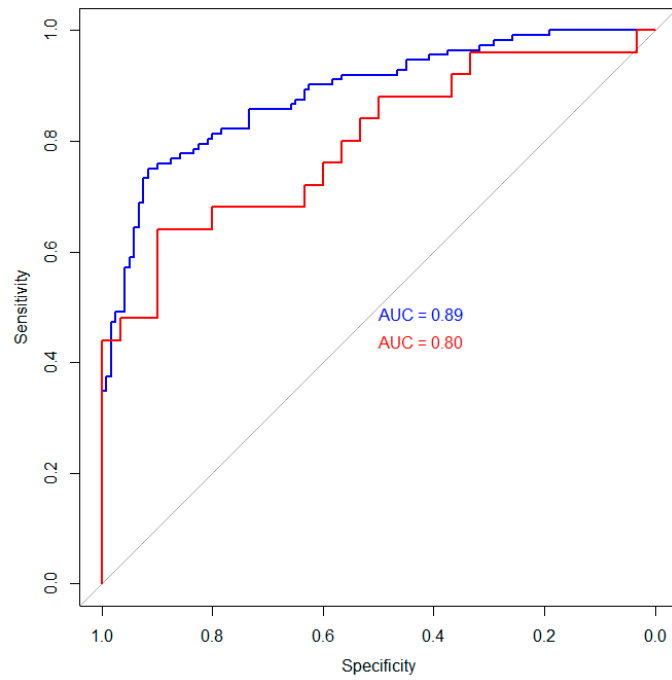
HR, hazard ratio. CI, confidence interval. CA19-9, carbohydrate antigen 19-9. AP, alkaline phosphatase. PMI, psoas muscle index.

### 3.3. Predictive Performance of the ANN

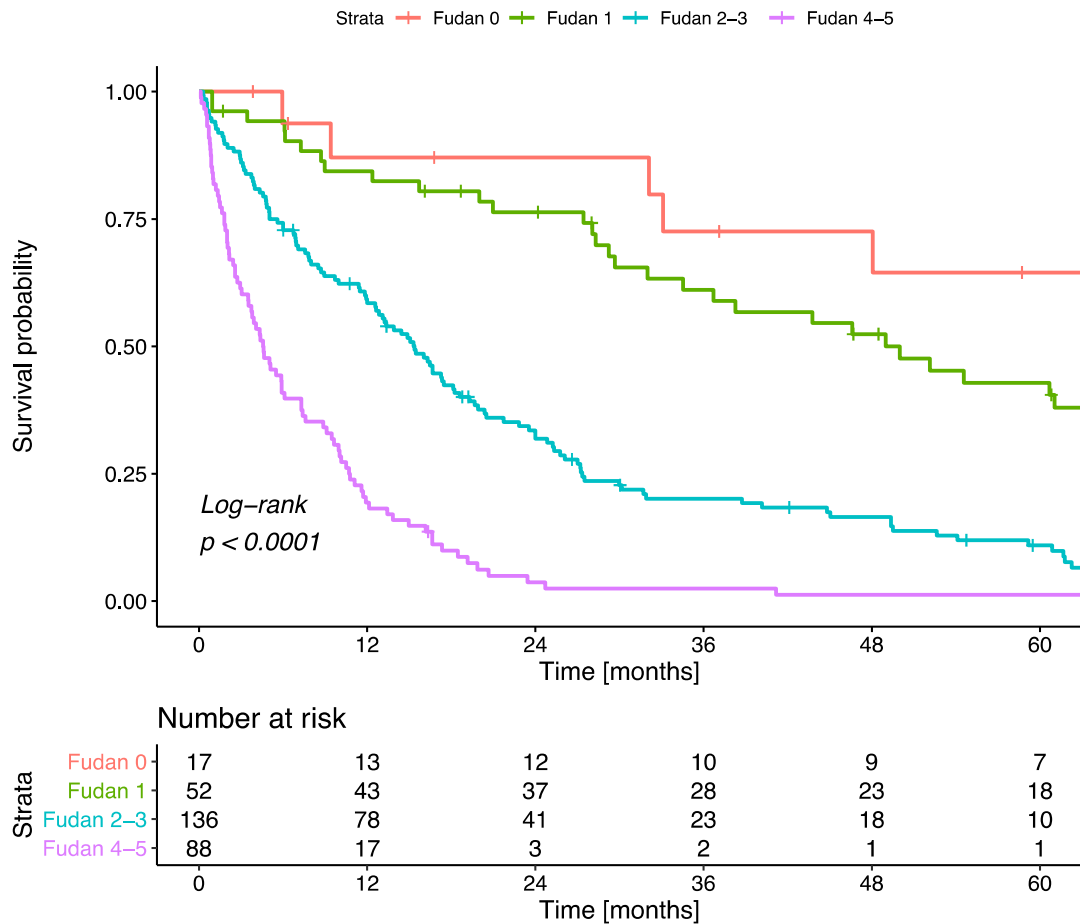
For the ANN, the AUC was 0.89 (95% CI 0.84–0.93) for the training set and 0.80 (95% CI 0.68–0.92) for the holdout validation set (Figure 5).

### 3.4. Predictive Performance of the Fudan Score

In a second step, we performed a head-to-head comparison of our newly developed ANN and the conventional Fudan score. Of the 293 patients, 17 (5.8%) had a low, 52 (17.8%) an intermediate, 136 (46.4%) a high, and 88 (30.0%) an extremely high Fudan score. The median OS was 69 months, 50 months, 15 months, and 5 months in the low-, intermediate-, high-, and extremely high risk groups, respectively (log-rank p-value < 0.001, Figure 6).



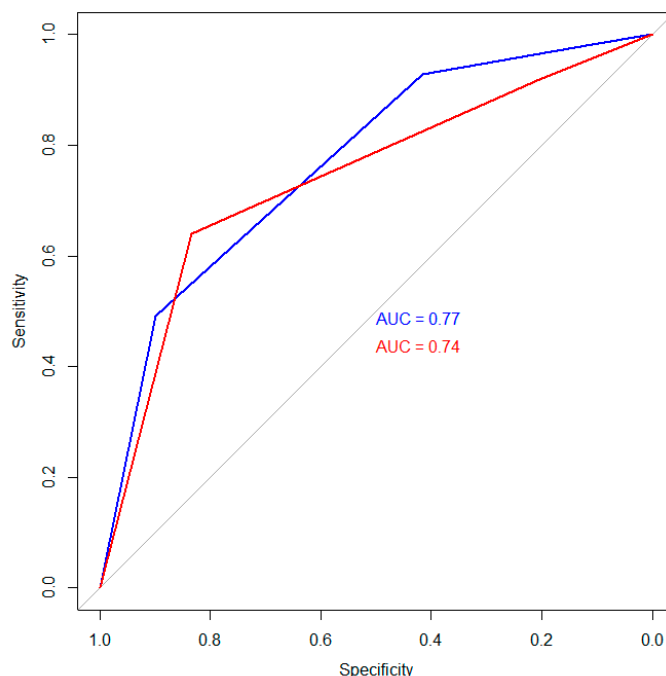
**Figure 5.** Visualization of the created artificial neural network. Receiver operating characteristic curves for the training (blue) and validation (red) sets.



**Figure 6.** Kaplan–Meier curves of overall survival stratified according to Fudan score.



Regarding individual risk prediction, the Fudan score yielded a Harrell's C-Index of 0.69 and an AUC for predicting 1-year survival probability of 0.77 (95% CI 0.71–0.82) for the training set and 0.74 (95% CI 0.61–0.87) for the holdout validation set (Figure 7).



**Figure 7.** Receiver operating characteristic curves for the training (blue) and validation (red) sets using the Fudan score.

Comparing both models, the AUC differed significantly for the training cohort (0.89 vs. 0.77,  $p < 0.001$ ), but the difference between both AUCs for the validation set did not reach significance (0.80 vs. 0.74,  $p = 0.24$ ).

#### 4. Discussion

In this study, we evaluated the feasibility of an ANN for ab initio risk prediction in patients with ICC. In a second step, we evaluated the Fudan score and performed a head-to-head comparison. In summary, the ANN reached an AUC of 0.89 in the training set and therefore outperformed the Fudan score (0.77) significantly ( $p < 0.001$ ). In the validation set, the ANN was also superior compared to the Fudan score (0.80 vs. 0.74). However, this difference did not reach significance ( $p = 0.24$ ), which might be attributable to the smaller sample size of the validation set. However, ANN models have excellent scalability; therefore, novel risk factors can easily be added to the developed model. Hence, these approaches will further improve risk prediction in patients with ICC.

Thus far, several scoring systems have been developed, especially for patients who have undergone tumor resection. The Hyder nomogram depends on tumor size, nodal status, vascular invasion, multifocality, presence/absence of cirrhosis, and age [9]. The Wang nomogram includes carcinoembryonic antigen and CA19-9 levels, vascular invasion, nodal status, and direct invasion or local metastasis, as well as tumor size [10]. The MEGNA score stratifies risk groups using the parameters multifocality, extrahepatic tumor extension, tumor grading, lymph node metastasis, and age [11]. Despite promising initial results, they all failed in external validation; though the Hyder nomogram had a C-Index of 0.69 in the derivation cohort, in an external validation by Doussot et al., the C-Index only reached 0.63. In the same study, the Wang nomogram reached superior values in estimating prognosis (C-Index 0.72). In two recent evaluations, the MEGNA score was found to be a useful stratification tool but failed in individual risk prediction [13,29]. Thus, none of the scores were implemented in the daily clinical routine.

The only scoring system available for patients regardless of histopathological factors is the Fudan score. This score consists of five common parameters assessed during standard work-up at the time of initial diagnosis and is not based on histopathological factors [14]. In a previous study by our group, all the included factors correlated with an impaired survival in our patient cohort [25]. Thus, the high discriminative ability ( $p < 0.001$ ) of the score in this study is not surprising. However, regarding individual survival prediction, the corresponding C-Index was only moderate (0.69), and 1-year survival prediction reached values of 0.77 for the training set and 0.74 for the validation set, which can be classified as a “fair prediction” [30,31]. One reason for the only moderate predictive ability of the Fudan score in our patient cohort might be the fact that we calculated the score regardless of the initial treatment. In the original publication, the authors developed the score on a population of resected cases and evaluated its performance on a small set of unresected patients.

All of the above-mentioned stratification systems rely on well-known clinical, histopathological, serological, and imaging-derived factors. However, they may not cover the clinical complexity because they are all based only on a few, mainly tumor burden-associated factors. Knowledge about novel risk factors, such as the tumor microenvironment, the influence of inflammation and immune reactions, body composition assessment, tumor standardized uptake in hybrid positron emission tomography/computed tomography imaging, and image-based texture analysis has continuously been increasing [25,32–37]. Therefore, the integration of these factors into scoring systems has great potential. For a successful translation into daily patient care, ML-based approaches offer a solution for the conjunction of well-known risk factors and this emerging knowledge. In addition, automated parameter processing using ML-based approaches becomes more applicable due to the continuous growth of digitization in the clinical infrastructure and electronic availability of patient data. In the future, dedicated software pipelines based on these approaches will enable automatic risk prediction.

However, ML-based studies on survival prediction in patients with ICC are scarce. Thus far, three attempts have been made: Focusing on tumor burden and the relationship between tumor size and number, Bagante et al. used a classification and regression tree model (CART) to identify prognostic groups of patients after curative-intent resection [18]. With their CART model, the group was able to visualize the hierarchical association between tumor burden and other clinical and histopathological factors. Li et al. applied different decision tree- and random forest-based ML algorithms to identify the most important risk factors for patients with ICC after resection [19]. In a second step, they created a novel scoring system based on the T and N categories of the ICC staging framework in the AJCC 8th edition, namely, carcinoembryonic antigen, CA19-9, alpha-fetoprotein, and prealbumin. Although their so-called EHBH-ICC score outperformed the AJCC 8th and LCSGJ staging systems, the final model’s C-Index was only moderate (0.69 for training and 0.67 for internal validation). The latest attempt by Jeong et al. achieved better values: in contrast to the two attempts before, but similar to our study, they used a Tensorflow deep learning algorithm to create a scoring system based on the wide range of four postoperative histopathological, six serological, and two etiological factors [20]. This system yielded an AUC of 0.78 in the original study and was more accurate than the AJCC staging system (0.60). In combination with our results, this supports our hypothesis that the inclusion of more risk factors enhances individual survival prediction.

Compared to other ML approaches and conventional scoring systems, the main advantage of ANNs may be that a multitude of different variables can be included quickly and the networks are easily scalable when novel parameters are integrated and complexity increases [15]. ANNs have the disadvantage of being kind of “black boxes” with complex interactions between included parameters and subsequent layers [15]. Furthermore, ANNs cannot deal with missing values. Thus, datasets have to be as complete as possible. In the future, this bias may be attenuated, as the digitization of medical records is continuously progressing and more and more parameters are automatically assessed. However, our results should only be interpreted as a proof of feasibility due to the single-center design

and missing external validation. Hence, large-scale validation studies are mandatory in the future.

One point that further stresses the potential of artificial intelligence-based approaches for survival prediction is the following fact: even though there was considerable heterogeneity regarding initial treatment, our approach reached a strong prognostic ability—even when applied at the very beginning of the patient’s clinical history.

Our study has several limitations: First, the dataset was acquired in a retrospective manner and the final sample size was only moderate ( $n = 293$ ) due to the monocentric nature of the study. However, the number of included patients was comparable to other studies examining the role of risk prediction and stratification for patients with ICC [9–12,14]. Second, as incidence is low in Western countries, the recruitment period was relatively long. In the meantime, significant improvements have been made in treatment, especially for patients with an unresectable tumor burden, and indication criteria have changed tremendously [8,38]. To reduce this bias, we actively decided to choose patients with an initial diagnosis in 2014 or later for the validation set. Third, we included only patients with complete datasets and actively decided against imputing missing values. Thus, we were not able to include important prognostic factors such as the Eastern Cooperative Oncology Group Performance Status or inflammation parameters such as the neutrophil to lymphocyte ratio or the platelet to lymphocyte ratio as the determination of these factors has not been a standard for patients treated before 2010. Therefore, the integration of these parameters would have considerably reduced the number of patients included into final analysis. However, especially the growing knowledge on inflammation indices offers great potential for survival prediction in patients with intrahepatic cholangiocarcinoma as they are easily available pre-operative serum markers. Fourth, for the sake of a clear methodology, we decided to use an 80:20 split based on the time of the initial diagnosis. However, as mentioned above, significant improvements have been made in treatment and in indication criteria. Therefore, the allocation according to the initial diagnosis date could have introduced a bias. However, even though treatment options evolved during the study period, our approach outperformed the Fudan score clearly for the validation set and reached a good predictive ability. Fifth, scoring systems derived from a single-center cohort of patients face the problem of “overfitting”. “Overfitting” describes “a phenomenon occurring when a model maximizes its performance on some set of data, but its predictive performance is not confirmed elsewhere due to random fluctuations of patients’ characteristics in different clinical and demographical backgrounds” [39]. Multicenter studies and the inclusion of patients with different ethnic backgrounds will attenuate this bias. Such studies would also enable us to approach the full capability of an ANN-based model.

## 5. Conclusions

ML-based approaches and especially ANNs offer the possibility of integrating a broad range of different patient parameters into risk prediction. This study proved the feasibility of this approach for patients with ICC prior to treatment. The ANN outperformed conventional risk scoring, leading to the conclusion that especially the inclusion of more risk factors offers a great potential for survival prediction. To reach the full capability of such approaches, large multicenter clinical databases are needed. Afterwards, such “big data”-based ANNs could easily be implemented into, for example, web-based risk calculations and integrated into the clinical routine workflow in order to support clinicians in daily decision making.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/jcm10102071/s1>, Table S1: TRIPOD Checklist: Prediction Model Development and Validation, Table S2: STROBE Statement—Checklist of items that should be included in reports of cohort studies.

**Author Contributions:** L.M., A.M.-K., S.J.G., F.F., C.D., A.W., F.B., L.K.-H., J.B., F.H. and R.K. devised the study, assisted in data collection, participated in the interpretation of the data, and helped draft

the manuscript. L.M., S.J.G., F.F., A.W., F.B. and F.H. carried out the data collection. A.M.-K., C.D., L.-K.H., J.B. and R.K. supported the data collection efforts. L.M., F.H. and R.K. created all of the figures and participated in the interpretation of data. L.M., F.H. and R.K. performed the statistical analysis. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the responsible Ethics Committee of the Medical Association of Rhineland Palatinate, Mainz, Germany (permit number 2018–13618, date of approval: 15 October 2018) for the retrospective analysis of clinical data. Additional examinations were not performed.

**Informed Consent Statement:** According to the responsible Ethics Committee of the Medical Association of Rhineland Palatinate, Mainz, Germany, informed consents were not needed given the retrospective study design. Patient records and clinical information were de-identified prior to analysis.

**Data Availability Statement:** Data cannot be shared publicly because of institutional and national data policy restrictions imposed by the Ethics Committee of the Medical Association of Rhineland Palatinate, Mainz, Germany, since the data contain potentially identifying patient information. Data are available upon request for researchers who meet the criteria for access to confidential data.

**Acknowledgments:** L.M. and S.J.G. are supported by the Clinician Scientist Fellowship “Else Kröner Research College: 2018\_Kolleg.05”.

**Conflicts of Interest:** A.W. has received speaker fees and travel grants from Bayer. R.K. has received consultancy fees from Boston Scientific, Bristol-Myers Squibb, Guerbet, Roche, and SIRTEX and lectures fees from BTG, Eisai, Guerbet, Ipsen, Roche, Siemens, SIRTEX, and MSD Sharp & Dohme. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Von Hahn, T.; Ciesek, S.; Wegener, G.; Plentz, R.R.; Weismüller, T.J.; Wedemeyer, H.; Manns, M.P.; Greten, T.F.; Malek, N.P. Epidemiological trends in incidence and mortality of hepatobiliary cancers in Germany. *Scand. J. Gastroenterol.* **2011**, *46*, 1092–1098. [[CrossRef](#)]
2. Yang, J.D.; Kim, B.; Sanderson, S.O.; Sauver, J.L.S.; Yawn, B.P.; Pedersen, R.A.; Larson, J.J.; Therneau, T.M.; Roberts, L.R.; Kim, W.R. Hepatocellular carcinoma in Olmsted County, Minnesota, 1976–2008. In *Proceedings of the Mayo Clinic Proceedings*; Elsevier: Amsterdam, The Netherlands, 2012; Volume 87, pp. 9–16.
3. Shaib, Y.H.; Davila, J.A.; McGlynn, K.; El-Serag, H.B. Rising incidence of intrahepatic cholangiocarcinoma in the United States: A true increase? *J. Hepatol.* **2004**, *40*, 472–477. [[CrossRef](#)]
4. Petrick, J.L.; Braunlin, M.; Laversanne, M.; Valery, P.C.; Bray, F.; McGlynn, K.A. International trends in liver cancer incidence, overall and by histologic subtype, 1978–2007. *Int. J. Cancer* **2016**, *139*, 1534–1545. [[CrossRef](#)]
5. Guro, H.; Kim, J.W.; Choi, Y.; Cho, J.Y.; Yoon, Y.-S.; Han, H.-S. Multidisciplinary management of intrahepatic cholangiocarcinoma: Current approaches. *Surg. Oncol.* **2017**, *26*, 146–152. [[CrossRef](#)] [[PubMed](#)]
6. Park, H.M.; Yun, S.P.; Lee, E.C.; Lee, S.D.; Han, S.-S.; Kim, S.H.; Park, S.-J. Outcomes for patients with recurrent intrahepatic cholangiocarcinoma after surgery. *Ann. Surg. Oncol.* **2016**, *23*, 4392–4400. [[CrossRef](#)]
7. Bridgewater, J.; Galle, P.R.; Khan, S.A.; Llovet, J.M.; Park, J.-W.; Patel, T.; Pawlik, T.M.; Gores, G.J. Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. *J. Hepatol.* **2014**, *60*, 1268–1289. [[CrossRef](#)] [[PubMed](#)]
8. Rizzo, A.; Brandi, G. First-line Chemotherapy in Advanced Biliary Tract Cancer Ten Years After the ABC-02 Trial: “And Yet It Moves!”. *Cancer Treat. Res. Commun.* **2021**, *27*, 100335. [[CrossRef](#)]
9. Hyder, O.; Marques, H.; Pulitano, C.; Marsh, J.W.; Alexandrescu, S.; Bauer, T.W.; Gamblin, T.C.; Sotiropoulos, G.C.; Paul, A.; Barroso, E. A nomogram to predict long-term survival after resection for intrahepatic cholangiocarcinoma: An Eastern and Western experience. *JAMA Surg.* **2014**, *149*, 432–438. [[CrossRef](#)] [[PubMed](#)]
10. Wang, Y.; Li, J.; Xia, Y.; Gong, R.; Wang, K.; Yan, Z.; Wan, X.; Liu, G.; Wu, D.; Shi, L. Prognostic nomogram for intrahepatic cholangiocarcinoma after partial hepatectomy. *J. Clin. Oncol.* **2013**, *31*, 1188–1195. [[CrossRef](#)]
11. Raoof, M.; Dumitra, S.; Ituarte, P.H.G.; Melstrom, L.; Warner, S.G.; Fong, Y.; Singh, G. Development and validation of a prognostic score for intrahepatic cholangiocarcinoma. *JAMA Surg.* **2017**, *152*, e170117. [[CrossRef](#)]
12. Doussot, A.; Groot-Koerkamp, B.; Wiggers, J.K.; Chou, J.; Gonen, M.; DeMatteo, R.P.; Allen, P.J.; Kingham, T.P.; D’Angelica, M.I.; Jarnagin, W.R. Outcomes after resection of intrahepatic cholangiocarcinoma: External validation and comparison of prognostic models. *J. Am. Coll. Surg.* **2015**, *221*, 452–461. [[CrossRef](#)]

13. Hahn, F.; Müller, L.; Mähringer-Kunz, A.; Schotten, S.; Düber, C.; Hinrichs, J.B.; Maschke, S.K.; Galle, P.R.; Bartsch, F.; Lang, H.; et al. Risk prediction in intrahepatic cholangiocarcinoma: Direct comparison of the MEGNA score and the 8th edition of the UICC/AJCC Cancer staging system. *PLoS ONE* **2020**, *15*, e0228501. [[CrossRef](#)]
14. Jiang, W.; Zeng, Z.-C.; Tang, Z.-Y.; Fan, J.; Sun, H.-C.; Zhou, J.; Zeng, M.-S.; Zhang, B.-H.; Ji, Y.; Chen, Y.-X. A prognostic scoring system based on clinical features of intrahepatic cholangiocarcinoma: The Fudan score. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **2011**, *22*, 1644–1652. [[CrossRef](#)]
15. Mähringer-Kunz, A.; Wagner, F.; Hahn, F.; Weinmann, A.; Brodehl, S.; Schotten, S.; Hinrichs, J.B.; Düber, C.; Galle, P.R.; Pinto dos Santos, D.; et al. Predicting survival after transarterial chemoembolization for hepatocellular carcinoma using a neural network: A Pilot Study. *Liver Int.* **2020**, *40*, 694–703. [[CrossRef](#)]
16. Cucchetti, A.; Piscaglia, F.; Grigioni, A.D.; Ravaioli, M.; Cescon, M.; Zanello, M.; Grazi, G.L.; Golfieri, R.; Grigioni, W.F.; Pinna, A.D. Preoperative prediction of hepatocellular carcinoma tumour grade and micro-vascular invasion by means of artificial neural network: A pilot study. *J. Hepatol.* **2010**, *52*, 880–888. [[CrossRef](#)]
17. Yamashita, R.; Long, J.; Longacre, T.; Peng, L.; Berry, G.; Martin, B.; Higgins, J.; Rubin, D.L.; Shen, J. Deep learning model for the prediction of microsatellite instability in colorectal cancer: A diagnostic study. *Lancet Oncol.* **2021**, *22*, 132–141. [[CrossRef](#)]
18. Bagante, F.; Spolverato, G.; Merath, K.; Weiss, M.; Alexandrescu, S.; Marques, H.P.; Aldrighetti, L.; Maithel, S.K.; Pulitano, C.; Bauer, T.W. Intrahepatic cholangiocarcinoma tumor burden: A classification and regression tree model to define prognostic groups after resection. *Surgery* **2019**, *166*, 983–990. [[CrossRef](#)]
19. Li, Z.; Yuan, L.; Zhang, C.; Sun, J.; Wang, Z.; Wang, Y.; Hao, X.; Gao, F.; Jiang, X. A Novel Prognostic Scoring System of Intrahepatic Cholangiocarcinoma With Machine Learning Basing on Real-World Data. *Front. Oncol.* **2021**, *10*, 3146.
20. Jeong, S.; Ge, Y.; Chen, J.; Gao, Q.; Luo, G.; Zheng, B.; Sha, M.; Shen, F.; Cheng, Q.; Sui, C. Latent Risk Intrahepatic Cholangiocarcinoma Susceptible to Adjuvant Treatment After Resection: A Clinical Deep Learning Approach. *Front. Oncol.* **2020**, *10*, 143. [[CrossRef](#)] [[PubMed](#)]
21. Collins, G.S.; Reitsma, J.B.; Altman, D.G.; Moons, K.G.M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement. *Circulation* **2015**, *131*, 211–219. [[CrossRef](#)] [[PubMed](#)]
22. Vandembroucke, J.P.; Von Elm, E.; Altman, D.G.; Gøtzsche, P.C.; Mulrow, C.D.; Pocock, S.J. The strengthening the reporting of observational studies in epidemiology (strobe) statement: Guidelines for reporting. *Ann. Intern. Med.* **2007**, *147*, 573–578. [[CrossRef](#)] [[PubMed](#)]
23. Weinmann, A.; Koch, S.; Niederle, I.M.; Schulze-Bergkamen, H.; König, J.; Hoppe-Lotichius, M.; Hansen, T.; Pitton, M.B.; Düber, C.; Otto, G. Trends in epidemiology, treatment, and survival of hepatocellular carcinoma patients between 1998 and 2009: An analysis of 1066 cases of a German HCC Registry. *J. Clin. Gastroenterol.* **2014**, *48*, 279–289. [[CrossRef](#)]
24. Meng, Z.-W.; Pan, W.; Hong, H.-J.; Chen, J.-Z.; Chen, Y.-L. Macroscopic types of intrahepatic cholangiocarcinoma and the eighth edition of AJCC/UICC TNM staging system. *Oncotarget* **2017**, *8*, 101165. [[CrossRef](#)]
25. Hahn, F.; Müller, L.; Stöhr, F.; Mähringer-Kunz, A.; Schotten, S.; Düber, C.; Bartsch, F.; Lang, H.; Galle, P.R.; Weinmann, A.; et al. The role of sarcopenia in patients with intrahepatic cholangiocarcinoma: Prognostic marker or hyped parameter? *Liver Int.* **2019**, *39*, 1307–1314. [[CrossRef](#)] [[PubMed](#)]
26. Chakedis, J.; Spolverato, G.; Beal, E.W.; Woelfel, I.; Bagante, F.; Merath, K.; Sun, S.H.; Chafitz, A.; Galo, J.; Dillhoff, M. Pre-operative sarcopenia identifies patients at risk for poor survival after resection of biliary tract cancers. *J. Gastrointest. Surg.* **2018**, *22*, 1697–1708. [[CrossRef](#)]
27. Hidden Layer. Available online: <https://deepai.org/machine-learning-glossary-and-terms/hidden-layer-machine-learning> (accessed on 23 April 2021).
28. Uno, H.; Cai, T.; Pencina, M.J.; D’Agostino, R.B.; Wei, L.-J. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat. Med.* **2011**, *30*, 1105–1117. [[CrossRef](#)]
29. Schnitzbauer, A.A.; Eberhard, J.; Bartsch, F.; Brunner, S.M.; Ceyhan, G.O.; Walter, D.; Fries, H.; Hannes, S.; Hecker, A.; Li, J. The MEGNA score and preoperative anemia are major prognostic factors after resection in the German intrahepatic cholangiocarcinoma cohort. *Ann. Surg. Oncol.* **2019**, *27*, 1147–1155. [[CrossRef](#)]
30. Swets, J.A. Measuring the accuracy of diagnostic systems. *Science* **1988**, *240*, 1285–1293. [[CrossRef](#)]
31. Duncan, I.G. *Healthcare Risk Adjustment and Predictive Modeling*; Actex Publications: Winsted, TX, USA, 2011; ISBN 1566987695.
32. Xue, B.; Wu, S.; Zheng, M.; Jiang, H.; Chen, J.; Jiang, Z.; Tian, T.; Tu, Y.; Zhao, H.; Shen, X. Development and Validation of a Radiomic-Based Model for Prediction of Intrahepatic Cholangiocarcinoma in Patients With Intrahepatic Lithiasis Complicated by Imagologically Diagnosed Mass. *Front. Oncol.* **2021**, *10*, 2807. [[CrossRef](#)] [[PubMed](#)]
33. Chu, H.; Liu, Z.; Liang, W.; Zhou, Q.; Zhang, Y.; Lei, K.; Tang, M.; Cao, Y.; Chen, S.; Peng, S. Radiomics using CT images for preoperative prediction of futile resection in intrahepatic cholangiocarcinoma. *Eur. Radiol.* **2020**, *31*, 2368–2376. [[CrossRef](#)] [[PubMed](#)]
34. Ji, G.-W.; Zhu, F.-P.; Zhang, Y.-D.; Liu, X.-S.; Wu, F.-Y.; Wang, K.; Xia, Y.-X.; Zhang, Y.-D.; Jiang, W.-J.; Li, X.-C. A radiomics approach to predict lymph node metastasis and clinical outcome of intrahepatic cholangiocarcinoma. *Eur. Radiol.* **2019**, *29*, 3725–3735. [[CrossRef](#)]
35. Coussens, L.M.; Werb, Z. Inflammation and cancer. *Nature* **2002**, *420*, 860–867. [[CrossRef](#)] [[PubMed](#)]
36. Sun, K.; Chen, S.; Xu, J.; Li, G.; He, Y. The prognostic significance of the prognostic nutritional index in cancer: A systematic review and meta-analysis. *J. Cancer Res. Clin. Oncol.* **2014**, *140*, 1537–1549. [[CrossRef](#)] [[PubMed](#)]

37. Lamarca, A.; Barriuso, J.; Chander, A.; McNamara, M.G.; Hubner, R.A.; O'Reilly, D.; Manoharan, P.; Valle, J.W. 18F-fluorodeoxyglucose positron emission tomography (18FDG-PET) for patients with biliary tract cancer: Systematic review and meta-analysis. *J. Hepatol.* **2019**, *71*, 115–129. [[CrossRef](#)] [[PubMed](#)]
38. Seidensticker, M.; Schütte, K.; Seidensticker, R.; Mühlmann, M.; Schulz, C. Multi-modal and sequential treatment of liver cancer and its impact on the gastrointestinal tract. *Best Pract. Res. Clin. Gastroenterol.* **2020**, *48–49*, 101709. [[CrossRef](#)]
39. Facciorusso, A.; Bhoori, S.; Sposito, C.; Mazzaferro, V. Repeated transarterial chemoembolization: An overfitting effort? *J. Hepatol.* **2015**, *62*, 1440–1442. [[CrossRef](#)]