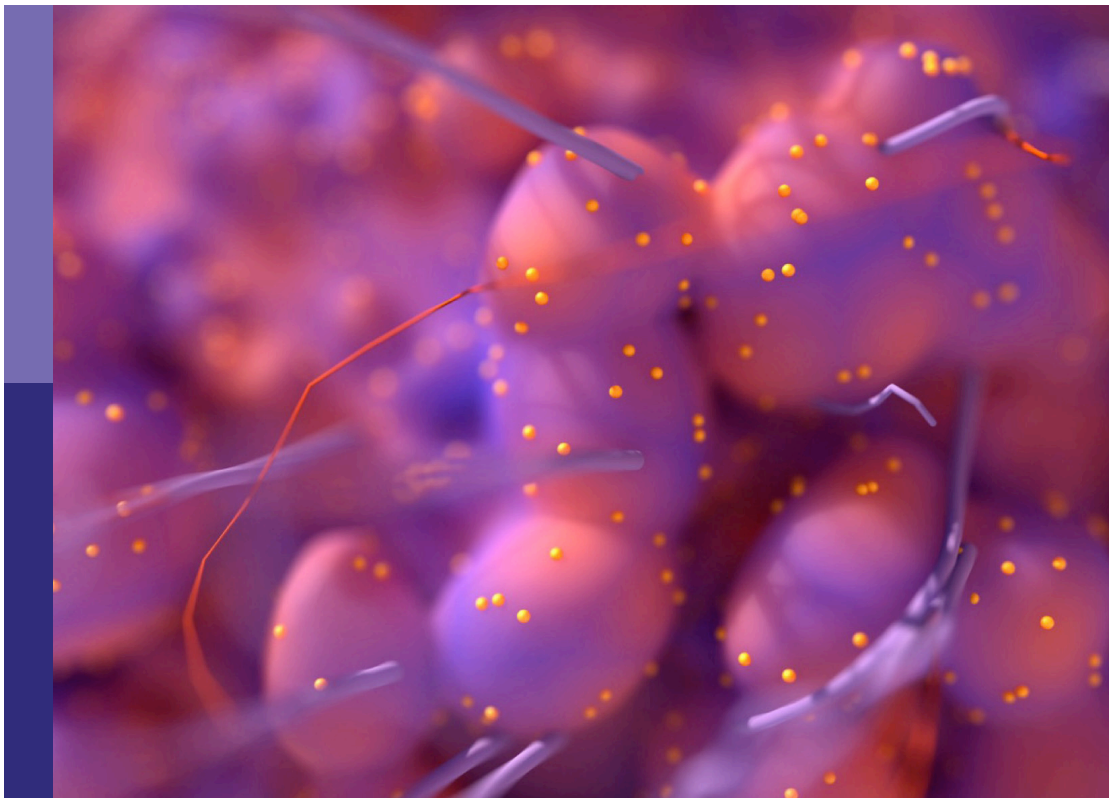
frontiers | Research Topics

Investigations into the potential benefits of artificial intelligence and deep learning to surgical oncologists

Edited by
Marco Scarpa and Hani J. Marcus

Coordinated by
Simon Williams

Published in
Frontiers in Oncology



ISSN 1664-8714
ISBN 978-2-8325-3087-0
DOI 10.3389/978-2-8325-3087-0

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Investigations into the potential benefits of artificial intelligence and deep learning to surgical oncologists

Topic editors
Marco Scarpa — University Hospital of Padua, Italy
Hani J. Marcus — University College London, United Kingdom

Topic Coordinator
Simon Williams — National Hospital for Neurology and Neurosurgery (NHNN), United Kingdom

Citation
Scarpa, M., Marcus, H. J., Williams, S., eds. (2023). *Investigations into the potential benefits of artificial intelligence and deep learning to surgical oncologists*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3087-0

Table of contents

frontiers | Frontiers in Oncology

TYPE Editorial
PUBLISHED 11 July 2023
DOI 10.3389/fonc.2023.1243305

OPEN ACCESS

EDITED AND REVIEWED BY
Aali Jan Sheen,
Manchester Royal Infirmary,
United Kingdom

*CORRESPONDENCE
Simon C. Williams
simon.williams32@nhs.net

RECEIVED 20 June 2023
ACCEPTED 28 June 2023
PUBLISHED 11 July 2023

CITATION
Williams SC and Marcus HJ (2023) Editorial: Investigations into the potential benefits of artificial intelligence and deep learning to surgical oncologists. *Front. Oncol.* 13:1243305. doi: 10.3389/fonc.2023.1243305

Editorial: Investigations into the potential benefits of artificial intelligence and deep learning to surgical oncologists

Simon C. Williams[1,2]* and Hani J. Marcus[1,3]

[1]Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London, London, United Kingdom, [2]Department of Neurosurgery, St George's University Hospital, London, United Kingdom, [3]Division of Neurosurgery, National Hospital for Neurology and Neurosurgery, London, United Kingdom

KEYWORDS

artificial intelligence, machine learning, surgery, oncology, computer vision, deep learning

Editorial on the Research Topic
Investigations into the potential benefits of artificial intelligence and deep learning to surgical oncologists

'Any sufficiently advanced technology is indistinguishable from magic."

So said science fiction writer Arthur C. Clarke, though few could have predicted the vast advances made in healthcare technology realised over the past decades. In the realm of surgical oncology, this sentiment holds true as we witness the rise of artificial intelligence (AI) and deep learning. These transformative technologies have the potential to revolutionise the field, enabling surgical oncologists to achieve remarkable advancements. This Research Topic aims to explore the potential benefits of AI and deep learning in assisting surgical oncologists and enhancing their decision-making capabilities. By harnessing the power of these technologies, we can aspire to achieve more accurate diagnoses, personalised treatment plans, optimised surgical interventions, and improved patient care (1).

This Research Topic presents nine publications spanning the surgical landscape, featuring research applied to Hepatobiliary Surgery (Jeong et al.; Li et al.; Huang et al.; Wang et al.), Orthopaedics (Yan et al.), Urology (Zhang et al.), Otolaryngology and Neurosurgery (Hill et al.; CRANIAL consortium). Each article is a demonstration of the unique opportunity AI poses to modern surgical oncologists. Here, we summarise the key insights presented in these publications, highlighting the advancements that AI presents in the field of surgical oncology.

First, the marriage of AI and deep learning with surgical oncology allows for intelligent data analysis on an unprecedented scale. These transformative technologies excel at swiftly and comprehensively processing vast amounts of complex data, as exemplified by the work of the CRANIAL Consortium et al. Their publication showcases machine learning-driven identification of predictors of cerebrospinal fluid rhinorrhoea following endonasal skull

base surgery, utilising a vast corpus of surgical data. Yet AI driven data analysis is not confined to text data - within the domain of machine learning is computer vision, a branch of AI that gives computer platforms understanding of image and video data (2). Advances in this field, particularly in radiomics, have garnered global attention due to their remarkable progress; however, their application in the operating theatre has yet to be fully realised (3). Within this Research Topic, Wang et al. describe the construction of a survival prediction model that integrates multimodal imaging data with clinical data, showcasing how AI can lead to comprehensive insights and personalised treatment planning, optimising patient outcomes.

Second, the work published in this Research Topic presents how AI and deep learning techniques have the potential to enhance surgical guidance and assistance. Zhang et al. demonstrate the utility of computer vision in pre-operative planning and intraoperative decision making through their generation of a model able to predict renal perfusion regions based on automated segmentation of renovascular imaging. Such advancement stands to increase patient safety through ever-more precise surgical planning. The application of surgical AI is not confined to the pre- and post-operative stages; by analysing intraoperative data, such as live imaging, physiological signals, and surgical instrument tracking, AI algorithms can provide surgical oncologists with valuable guidance, assisting them in navigating critical structures, optimizing surgical margins, and ensuring precise tumour resection (4). Moreover, AI-powered systems can detect and predict potential complications, alerting surgeons in advance and allowing for timely interventions, ultimately leading to improved patient safety and outcomes (5).

Perhaps most valuable, however, is the promise of AI to offer predictive models that shape treatment decisions and outcomes, as demonstrated by numerous articles within this Research Topic. By leveraging large datasets, these models identify prognostic factors, predict treatment responses, and stratify patients into risk groups (6). Jeong et al. describe the creation survival prediction platform for patients with intrahepatic cholangiocarcinoma (ICC), enabling categorisation of patients into risk groups to guide clinical interventions. A similar prognostic approach is adopted by Yan et al., who describe the creation of a deep learning model to predict overall survival in chondrosarcoma patients. Notably, their DeepSurv model outperformed traditional models of survival prediction, highlighting the unique capability of machine learning methods to identify subtle relationships between variables in large, complex datasets. Applications such as this give an insight into how AI will individualise treatment decisions based on patient-specific data, replacing crude and generic risk prediction systems. Surgical oncologists can harness these insights to tailor treatment plans to individual patients, optimize the sequencing of therapies, and explore alternative strategies.

The integration of AI and deep learning transcends clinical practice and extends to surgical training and education. Virtual reality (VR) and augmented reality (AR) platforms, combined with AI algorithms, can simulate realistic surgical scenarios, providing surgical trainees with a safe and controlled environment for practicing complex procedures, such as tumour resections. By analysing trainee performance and offering real-time feedback, AI-powered systems may accelerate the learning curve, enhancing surgical skills acquisition (7). The scope of AI to benefit trainees does not stop there, however – AI algorithms may aid in curating educational resources, extracting key insights from scientific literature, and delivering personalised learning experiences to surgical oncologists. Hill et al. demonstrate the ability of AI to aid disease classification, through their use of AI to subclassify glioblastoma, a disease with a profoundly poor prognosis, enabling a clear taxonomy and better prediction of patient outcomes.

The articles presented within this Research Topic showcase the rich field of AI research in surgical oncology, yet whilst the potential benefits of AI and deep learning in surgical oncology are vast, ethical considerations and challenges must be navigated. Issues such as data privacy, algorithm bias, transparency, and accountability must be addressed to ensure patient safety, maintain trust in the healthcare system, and mitigate potential risks (2). Close collaboration between clinicians, researchers, policymakers, and regulatory bodies is crucial to establish guidelines and frameworks that uphold ethical standards and govern the integration of AI in surgical oncology (6).

We sincerely believe that the contents of this Research Topic will be of interest to surgeons, oncologists, and members of the wider healthcare team alike. AI and deep learning have the potential to revolutionise the field of surgical oncology, the true benefits of which are yet to be fully realised. By embracing the future, we embark on a journey to redefine healthcare.

Author contributions

SW - writing - original draft. HM - writing - reviewing and editing; conceptualization. All authors contributed to the article and approved the submitted version.

Conflict of interest

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

Publisher's note

References

1. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J* (2019) 6:94–8. doi: 10.7861/futurehosp.6-2-94

2. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg* (2018) 268:70–6. doi: 10.1097/SLA.0000000000002693

3. Hashimoto DA, Rosman G, Rus D, Meireles OR. Surgical video in the age of big data. *Ann Surg* (2018) 268:e47–8. doi: 10.1097/SLA.0000000000002493

4. Starup-Hansen J, Williams SC, Funnell JP, Hanrahan JG, Islam S, Al-Mohammad A, et al. Optimising trajectory planning for stereotactic brain tumour biopsy using artificial intelligence: a systematic review of the literature. *Br J Neurosurg* (2023), 1–10. doi: 10.1080/02688697.2023.2210225

5. Stam WT, Goedknegt LK, Ingwersen EW, Schoonmade LJ, Bruns ERJ, Daams F. The prediction of surgical complications using artificial intelligence in patients undergoing major abdominal surgery: a systematic review. *Surgery* (2022) 171:1014–21. doi: 10.1016/j.surg.2021.10.002

6. Williams S, Layard Horsfall H, Funnell JP, Hanrahan JG, Khan DZ, Muirhead W, et al. Artificial intelligence in brain tumour surgery-an emerging paradigm. *Cancers (Basel)* (2021) 13:5010. doi: 10.3390/cancers13195010

7. Lam K, Chen J, Wang Z, Iqbal FM, Darzi A, Lo B, et al. Machine learning for technical skill assessment in surgery: a systematic review. *NPJ Digit Med* (2022) 5:1–16. doi: 10.1038/s41746-022-00566-0

frontiers in Oncology

ORIGINAL RESEARCH
published: 19 February 2020
doi: 10.3389/fonc.2020.00143

Latent Risk Intrahepatic Cholangiocarcinoma Susceptible to Adjuvant Treatment After Resection: A Clinical Deep Learning Approach

Seogsong Jeong[1,2†], Yang Ge[3†], Jing Chen[2†], Qiang Gao[4], Guijuan Luo[2], Bo Zheng[2], Meng Sha[1], Feng Shen[5], Qingbao Cheng[6], Chengjun Sui[7], Jingfeng Liu[8], Hongyang Wang[2*], Qiang Xia[1*] and Lei Chen[2,9*]*

[1] Department of Liver Surgery, School of Medicine, Renji Hospital, Shanghai Jiao Tong University, Shanghai, China, [2] International Cooperation Laboratory on Signal Transduction, Eastern Hepatobiliary Surgery Institute, Second Military Medical University, Shanghai, China, [3] School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [4] Department of Liver Surgery and Transplantation, Liver Cancer Institute, Zhongshan Hospital, and Key Laboratory of Carcinogenesis and Cancer Invasion (Ministry of Education), Fudan University, Shanghai, China, [5] Eastern Hepatobiliary Surgery Hospital, Second Military Medical University, Shanghai, China, [6] Biliary Tract Department I, Eastern Hepatobiliary Surgery Hospital, Second Military Medical University, Shanghai, China, [7] Department of Special Medical Care and Liver Transplantation, Eastern Hepatobiliary Surgery Hospital, Second Military Medical University, Shanghai, China, [8] The United Innovation of Mengchao Hepatobiliary Technology Key Laboratory of Fujian Province, Mengchao Hepatobiliary Hospital of Fujian Medical University, Fuzhou, China, [9] Department of Oncology, Shanghai Cancer Center and Institutes of Biomedical Sciences, Shanghai Medical College, Fudan University, Shanghai, China

Background: Artificial Intelligence (AI) frameworks have emerged as a novel approach in medicine. However, information regarding its applicability and effectiveness in a clinical prognostic factor setting remains unclear.

Methods: The AI framework was derived from a pooled dataset of intrahepatic cholangiocarcinoma (ICC) patients from three clinical centers ($n = 1,421$) by applying the TensorFlow deep learning algorithm to Cox-indicated pathologic (four), serologic (six), and etiologic (two) factors; this algorithm was validated using a dataset of ICC patients from an independent clinical center ($n = 234$). The model was compared to the commonly used staging system (American Joint Committee on Cancer; AJCC) and methodology (Cox regression) by evaluating the brier score (BS), integrated discrimination improvement (IDI), net reclassification improvement (NRI), and area under curve (AUC) values.

Results: The framework (BS, 0.17; AUC, 0.78) was found to be more accurate than the AJCC stage (BS, 0.48; AUC, 0.60; IDI, 0.29; NRI, 11.85; $P < 0.001$) and the Cox model (BS, 0.49; AUC, 0.70; IDI, 0.46; NRI, 46.11; $P < 0.001$). Furthermore, hazard ratios greater than three were identified in both overall survival (HR; 3.190; 95% confidence interval [CI], 2.150–4.733; $P < 0.001$) and disease-free survival (HR, 3.559; 95% CI, 2.500–5.067; $P < 0.001$) between latent risk and stable groups in validation. In addition, the latent risk subgroup was found to be significantly benefited from adjuvant treatment (HR, 0.459; 95% CI, 0.360–0.586; $P < 0.001$).

OPEN ACCESS

Edited by:
Michael Anthony Silva,
Oxford University Hospitals NHS Trust,
United Kingdom

Reviewed by:
Nehal Sureshkumar Shah,
Sheffield Teaching Hospitals,
United Kingdom
John Keith Roberts,
University Hospitals Birmingham NHS
Foundation Trust, United Kingdom

***Correspondence:**
Jingfeng Liu
drjingfeng@126.com
Hongyang Wang
hywangk@vip.sina.com
Qiang Xia
xiaqiang@shsmu.edu.cn
Lei Chen
chenlei@smmu.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Surgical Oncology, a section of the journal Frontiers in Oncology

Received: 10 August 2019
Accepted: 27 January 2020
Published: 19 February 2020

Citation:
Jeong S, Ge Y, Chen J, Gao Q, Luo G, Zheng B, Sha M, Shen F, Cheng Q, Sui C, Liu J, Wang H, Xia Q and Chen L (2020) Latent Risk Intrahepatic Cholangiocarcinoma Susceptible to Adjuvant Treatment After Resection: A Clinical Deep Learning Approach. Front. Oncol. 10:143. doi: 10.3389/fonc.2020.00143

Conclusions: The AI framework seems promising in the prognostic estimation and stratification of susceptible individuals for adjuvant treatment in patients with ICC after resection. Future prospective validations are needed for the framework to be applied in clinical practice.

Keywords: biliary malignancy, artificial intelligence, prognostic factor, prediction model, primary liver cancer

INTRODUCTION

Artificial Intelligence (AI) is a field of computer science in which machines mimic, recognize, and learn cognitive functions of the human mind and make empirical predictions using task-specific algorithms (1, 2). It is natural for the human mind to get confused when trying to process a lot of information simultaneously, and this necessitates an auxiliary process. This need has led to the application of AI in clinical medicine (3). AI has been applied to develop a diagnostic tool using image-based deep learning (DL), and the resulting performance was close to that of humans (4). However, no study has applied an AI framework to identify patients prone to the latent risk of recurrence even after curative treatment.

Intrahepatic cholangiocarcinoma (ICC) is a highly aggressive primary epithelial malignancy arising from the liver, and it has witnessed rising interests over the years due to rapid increase in its incidence and the resulting mortality rate (5, 6). Usually, ICC is diagnosed at an advanced stage, sporadically and without an explicit etiologic factor, thereby limiting curative approaches (7). Surgery with curative intent is the current standard of care, providing the opportunity for long-term survival (8). However, due to frequent recurrence of ICC, less than half of the post-surgery patients have been reported to survive for more than 5 years (9).

Despite clinical challenges, the growing understanding of ICC, led by increased investigations, is providing new insight into heterogeneity, pathogenesis, and therapeutic strategies with regard to the disease (10–12). Locally advanced ICC is no longer a contraindication for transplantation, and adjuvant treatments are now implemented more frequently worldwide, suggesting that it is important to identify the prognostic subtype for all treatments (13, 14). However, prognostic subtypes that support the selection of therapeutic modality remain limited, especially for recurrent ICC.

Owing to the exponential increase in the number of ICC studies, prognosis of the disease is witnessing development (15). Various prognosis-predictive systems with biological, pathological, demographic, clinical, and imaging characteristics have been developed (16, 17). However, such systems could not be implemented widely due to their inaccuracy and discriminations against them. To address this issue, we have developed a DL computational framework for ICC. The framework was tested in subgroups of patients who received prophylactic adjuvant transarterial chemoembolization (PAT), post-recurrent chemotherapy (PRC), post-recurrent radiotherapy (PRR), post-recurrent transarterial chemoembolization (PRT), and post-recurrent percutaneous microwave coagulation (PRP). The tests were carried out in accordance with prognostic subtypes identified by machine learning algorithms.

METHODS

Patients

The framework was retrospectively derived using a pooled dataset from patients with ICC who received surgical resection at the Eastern Hepatobiliary Surgery Hospital, Second Military Medical University (n = 1,477), Renji Hospital, School of Medicine, Shanghai Jiao Tong University (n = 106), and Mengchao Hepatobiliary Hospital, Fujian Medical University (n = 14) between 2008 and 2015, which was independently validated by the patients from Zhongshan Hospital, Fudan University (n = 246). All four databases satisfied the following inclusion criteria: Eastern Cooperative Oncology Group (ECOG) performance status of 0–1, no neoadjuvant treatment, no mixed hepatocellular-cholangiocarcinoma and hilar/distal cholangiocarcinoma, no perioperative death (within 30 days after surgery), and no distant metastasis. In the quality assessment, 188 patients were excluded due to incomplete data, and thus a total of 1,421 and 234 patients were finally enrolled for the model training and validation, respectively. This study was carried out in accordance with the TRIPOD statement. The protocol was approved by the Ethics Committee of Renji Hospital, School of Medicine, Shanghai Jiao Tong University. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

Diagnosis

Diagnosis of ICC was performed based on results of clinical, serological, and imaging studies (contrast-enhanced computed tomography [CT] and/or magnetic resonance imaging [MRI]). Positron emission tomography (PET) was performed in patients suspicious of metastases according to clinical and radiological characteristics. After surgery, CK7, CK19, and MUC1 positivity along with CK20, HepPar1, and glypican-3 negativity was considered pathological confirmation of ICC (18).

Clinical Interventions

Resection was carried out according to the size and location of tumor, estimated post-operative liver volume, and the Couinaud segmentation as described before (16). Hepatoduodenal ligament, retropancreatic, and paraaortic lymph nodes were routinely dissected. Perihepatic lymph node metastasis identified by preoperative CT/PET was considered for surgery if considered completely removable.

PAT was considered after discussion of the pros and cons of PAT by the operating surgeons and patients. The performance of PAT mostly depended on their socioeconomic status and intention. Among the patients who actively agreed to receive PAT, those with Child-Pugh class of A to B, ECOG score of 0–1, normal kidney function, no evidence of extrahepatic metastasis, platelet count above 50×10^9/L, and white blood cell count above 3×10^9/L were enrolled. PAT was performed within 2 months after resection by injecting 3–5 ml of iodized oil emulsion (Lipiodol, Guerbet Laboratories) with 500 mg of 5-fluorouracil (FU), 10 mg of hydroxycamptothecin, and 20 mg of epirubicin (19).

For recurrent ICC, gemcitabine and/or 5-FU-based PRC was prior for metastatic diseases, whereas a dose volume histogram-based PRR (90% of dose curve covered by the plan target volume) was preferentially performed in patients with large tumors and/or vascular invasion; there was no priority between PRT and PRR, in line with the National Comprehensive Cancer Network (NCCN) guidelines. PRT was carried out using the same methodology as PAT. PRP was proceeded using an MTC-3 microwave generator (2,450 MHz, 1–100 W) at 80–100 W for 3–5 min automatically with a safety margin of 1 cm (20). Supportive information related to inclusion of patients for standardized performance of the procedures is described in the **Supplementary Material**.

Follow-up

Active follow-up of serum carbohydrate antigen (CA) 19-9, carcinoembryonic antigen (CEA), alpha fetoprotein (AFP), liver function tests, and the abdominal ultrasound was made by patients once per 2 months within 2 years after surgery and once per 3–6 months thereafter. Patients without active visits were contacted by telephone inquiries. CT/MRI was performed once per 6 months or less when recurrence was suspicious. Development of new lesions with radiological characteristics of ICC was considered as a recurrence. Follow-up was discontinued at the time of death. The terms "disease-free survival (DFS)" and "overall survival (OS)" were defined as time from surgery to the detection of recurrence/metastasis and death, respectively.

Network Architecture and Derivation Procedures

The authors who derived the framework were blinded to the validation dataset, whereas those who validated the framework were blinded to the derivation dataset. To infer an estimated probability for latent risk and latent stable as output, we conducted a 12×1 vectors, including 12 clinical indicators, in the input layer via full-connected hidden layers (12×28, 28×28, 28×14, and 14×28 nodes). For the given hidden layer i, we applied tanh for activation function between input x and output y:

$$y = f_i(x) = \tanh\left(W_i x + b_i\right)$$
$$\tanh x = \frac{sinhx}{coshx} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

x and y are two arrays of the sized p and q, whereas W_i and b_i are the weight matrix and the intercept array, respectively. For the output layer, we used the softmax as an activation function:

$$y = f_o(x) = softmax\left(W_i x + b_i\right)$$
$$softmax \quad f_i\left(\vec{x}\right) = \frac{e^{x_i}}{\sum_{j=1}^{J} e^{x_j}} \qquad for\ i = 1, \ldots, J$$

For the neural network with k layers, y is driven from:

$$y = F_{1\to k}(x) = f_k \circ f_{k-1} \ldots \circ f_1(x)$$

where $f_k \circ f_{k-1}(x) = f_k\left(f_{k-1}(x)\right)$ is the composed function of f_k with f_{k-1}. To train this AI framework to find the different weight vectors W_i and bias b_i by minimizing the error between predicted output and actual class, we chose cross entropy as the loss function, which indicates the error between predicted y_{pred} and actual ending y_{actual}.

$$Cross\ entropy \quad H\left(y_{actual}, y_{pred}\right) = -\sum y_{actual}(x) * log\ y_{pred}(x)$$

W_i and b_i were initialized with truncated normal distribution (standard deviation = 0.1; https://www.tensorflow.org/api_docs/python/tf/truncated_normal). The Adam Optimizer algorithm (initial learning rate=0.001) was used to minimize the loss function via backpropagation to update weights and biases per layer (21). In addition, we have applied a dropout layer by randomly dropping 30% weights before the output layer to improve the generalization ability, but application of the weight decay was found to decrease the performance ability of the AI framework. The model was trained for 1,500 iterations with a batch size 200 in producing a model update to support multiple updates for each iteration.

Definition of the Prognostic Subtypes

The term "latent risk (AI-framework-estimated recurrence probability > 0.5)" refers to a subset of ICCs that are under severe risk of recurrence at any time after resection; resection of the tumor is therefore not likely to be curative regardless of curative intent. "Latent stable (AI-framework-estimated recurrence probability<0.5)" refers to a relatively constant disease status that resection of the tumor provides a long-term satisfactory prognosis. To support understanding, latent risk and latent stable can be simply considered as AI-high risk and AI-low risk, respectively.

Statistical Analysis

The primary and secondary endpoints were DFS and OS, respectively. The model was evaluated by comparing with the AJCC stage and Cox multivariate hazard proportional model-derived individualized scores, which were indicated by changes in χ^2, integrated discrimination improvement (IDI) and a net reclassification improvement (NRI) with 95% confidence interval (CI), and receiver operating characteristic (ROC) curves with area under curve (AUC) values. Although an AJCC stage for ICC was not developed with intent for survival prediction, it still is the most commonly applied staging system in clinical medicine

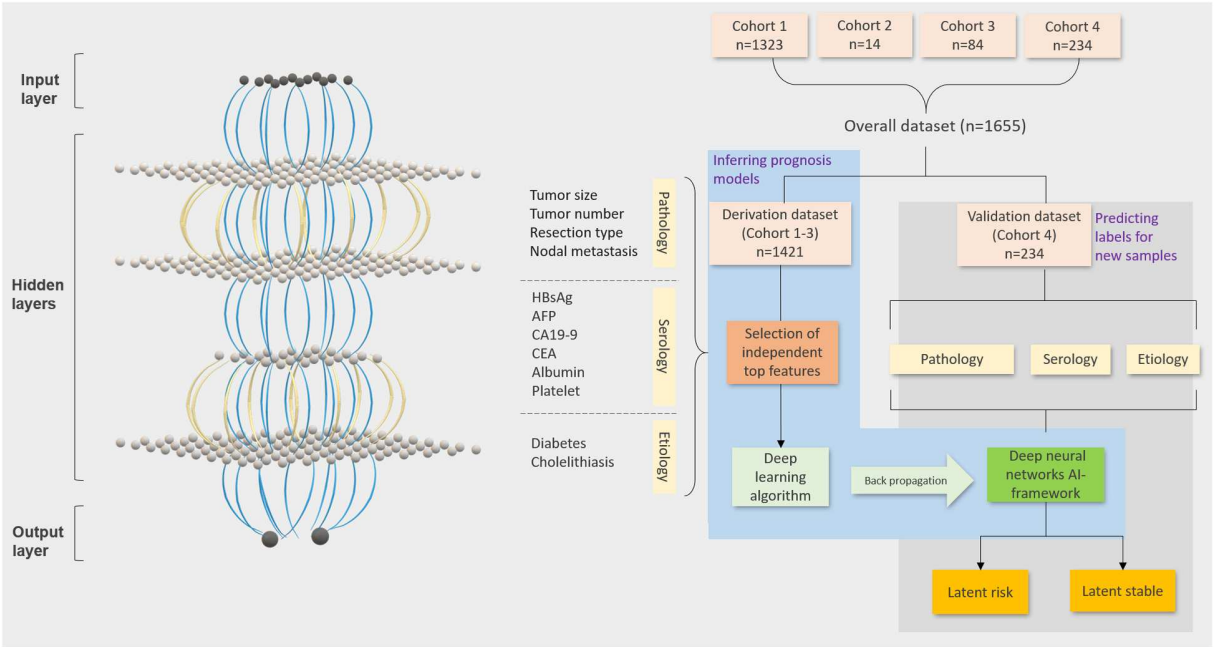


FIGURE 1 | Workflow of the ICC AI-framework. TensorFlow-based deep learning and machine learning techniques to evaluate latent risk ICC by integrating the generally obtainable pathologic, serologic, and etiologic clinical factors of the patients from four independent clinical centers. The workflow includes four steps (Step 1: randomization of derivation and validation datasets; Step 2: Selection of the significant covariates; Step 3: deep learning algorithm for evaluation of individual scores; Step 4: stratification of latent risk and stable).

supportive of survival estimation. Kaplan-Meier (KM) curves with the log-rank test for *P*-value and Mantel-Haenszel for hazard ratio (HR) were generated for evaluation and digitalization of survival outcomes. $P < 0.05$ is regarded statistically significant. The ICC AI framework was constructed using the TensorFlow (v1.2.1) on servers equipped with dual Intel (R) Core (TM) i7-4650U CPU @1.70 Ghz 2.30 GHz, 8 GB RAM, and Intel (R) HD Graphics 5000. All statistical analyses were performed using Python (v3.6.5) and R Project for Statistical Computing (v3.4.4).

RESULTS

Development of the ICC AI Framework

An AI framework to evaluate individualized probabilities for identifying categorical prognostic subtypes was developed. For this purpose, independent significant covariate features and the DL algorithm were selected using non-overlapping derivation and validation datasets (**Figure 1**). Demographic, etiological, pathological, and serological characteristics were evaluated using univariable and multivariable Cox regression models for disease-free survival. The 28 evaluated characteristics were gender; age; hepatitis B virus (HBV) and hepatitis C virus infections; HBV DNA; antiviral treatment; syphilis infection; liver cirrhosis and fluke; fatty liver; smoking and alcohol abuse; diabetes mellitus; hypertension; tumor location and differentiation; AFP, CA 19-9, 125, and 242; CEA; albumin; platelet count; vascular invasion; lymph node metastasis; tumor size and number; and surgical extent. The evaluation identified 12 of the features as most important, including tumor size and number, surgical extent, lymph node metastasis, hepatitis B surface antigen (HBsAg), AFP, CA19-9, CEA, albumin, platelet count, diabetes mellitus, and cholelithiasis (**Tables 1**, **2**). Albumin (>35 vs. ≤35 g/L), AFP (>50 vs. ≤50 ng/ml), and CA 19-9 (>37 vs. ≤37 U/ml) were categorized into normal and abnormal groups according to the standardized cut-off values for normal ranges; the platelet count was stratified into <100, 100–300, and >300 $\times$ 10^9/L; CEA was stratified into <2.5, 2.5–5.0, and >5.0 ng/ml; tumor size was stratified into <2.5, 2.5–5.0, and >5.0 cm; and tumor number was categorized into single, double, and multiple tumors. Tests were conducted to confirm if the covariates were significant prognostic factors for the OS in the derivation dataset. The multivariable analysis found all involved factors, except albumin and diabetes, to be significantly and independently predictive of the OS (**Supplementary Table 1**). Additionally, HBsAg, AFP, tumor size, and resection type were identified as insignificant independent prognostic factors in the validation dataset (**Supplementary Table 2**). Finally, a training dataset (n = 1,421) was used to derive the framework based on the 12 identified features. The framework was derived with time-to-event outcomes using a backpropagation technique, which synchronously updated each lay's weights and biases to optimize the statistical likelihood of the framework.

Validation of the ICC AI-Framework

The performance of the model was assessed by comparing the consistency of the disease status with that of the individualized stage/score from the validation set (n = 234; **Figure 2A**). Relative maldistributions were observed in the range-adjusted American Joint Committee on Cancer (AJCC) staging system (BS = 0.48) and the Cox multivariable models (BS = 0.49), whereas the

TABLE 1 | Baseline demographic and clinical characteristics of the patients.

	Derivation dataset (n = 1,421)	Validation dataset (n = 234)
Age, years	57 (49–64)	58 (50–65)
Gender, male	915 (64.4)	140 (59.8)
Albumin, g/L	40.4 (36.1–43.5)	41.0 (39.0–43.0)
Platelet count, 10^9/L	184 (142–238)	189 (147–228)
Diabetes	136 (9.6)	25 (10.7)
HBV infection, HBsAg	624 (43.9)	28 (12.0)
Cholelithiasis	132 (9.3)	18 (7.7)
AFP, ng/ml	3.0 (2.0–5.5)	2.8 (1.9–4.4)
CA19–9, U/ml	57.8 (17.8–548.1)	32.1 (11.6–239.0)
CEA, ng/ml	2.8 (1.7–5.7)	2.4 (1.5–4.8)
Tumor size, cm	6.0 (4.0–8.0)	5.0 (3.5–8.0)
Tumor number		
Single	1221 (85.9)	188 (80.3)
Multiple	200 (14.1)	46 (19.7)
Lymph node metastasis	332 (23.4)	60 (25.6)
Resection type		
Minor hepatectomy	1052 (74.0)	134 (57.3)
Hemi or extended hepatectomy	369 (26.0)	100 (42.7)
TNM stage[a]		
I–II	1089 (76.6)	174 (74.4)
III–IV	332 (23.4)	60 (25.6)

Data are n (%) or median (IQR). HBsAg, hepatitis B surface antigen; AFP, alpha fetoprotein; CA, carbohydrate antigen; CEA, carcinoembryonic antigen. [a]TNM stage: American Joint Committee on Cancer 8th edition staging for intrahepatic cholangiocarcinoma.

TABLE 2 | Selection of top covariates using the Cox multivariable regression.

	Univariable analysis		Multivariable analysis	
	HR (95% CI)	P-value	HR (95% CI)	P-value
Albumin <35 g/L	1.96 (1.66–2.31)	<0.001	1.26 (1.05–1.51)	0.015
Platelet count, $\times 10^9$/L[a]	1.68 (1.45–1.94)	<0.001	1.21 (1.04–1.40)	0.011
Diabetes	1.63 (1.34–1.99)	<0.001	1.41 (1.15–1.72)	0.001
HBsAg	0.82 (0.72–0.93)	0.002	0.79 (0.69–0.90)	0.001
Cholelithiasis	1.57 (1.28–1.92)	<0.001	1.40 (1.13–1.73)	0.002
AFP >50 ng/ml	1.49 (1.19–1.86)	0.001	1.60 (1.26–2.02)	<0.001
CA19–9 > 37 U/ml	1.49 (1.32–1.69)	<0.001	1.18 (1.03–1.37)	0.020
CEA, ng/ml[b]	1.37 (1.27–1.47)	<0.001	1.12 (1.03–1.22)	0.011
Tumor size, cm[c]	1.69 (1.56–1.84)	<0.001	1.59 (1.46–1.73)	<0.001
Tumor number[d]	1.51 (1.37–1.67)	<0.001	1.28 (1.15–1.42)	<0.001
Lymph node metastasis	1.93 (1.68–2.22)	<0.001	1.40 (1.21–1.63)	<0.001
Resection type[e]	1.57 (1.42–1.74)	<0.001	1.17 (1.05–1.31)	0.005

HR, hazard ratio; CI, confidence interval; HBsAg, hepatitis B surface antigen; AFP, alpha fetoprotein; CA, carbohydrate antigen; CEA, carcinoembryonic antigen. [a]was stratified into <100, 100–300, and >300. [b]was stratified into <2.5, 2.5–5.0, and >5.0. [c]was stratified into ≤2.0, 2.1–3.0, 3.1–5.0, and >5.0. [d]was stratified into single, double, and multiple. [e]was stratified into minor hepatectomy, hemihepatectomy, and extended hepatectomy.

framework (BS = 0.17) demonstrated well-propagated DL scores. Furthermore, visualization of the score-dependent disease status revealed predominance in the AI framework, demonstrating gradual increase of recurrence in proportion to the DL score (**Figure 2B**). Additionally, the AI framework, covariates, AJCC staging system, and Cox multivariable regression were evaluated using ROC curves and AUC values, and the Cox score and AJCC stage were evaluated by the validation dataset (**Figure 2C**). The framework was derived (AUC = 0.84) and validated (AUC = 0.78) to be excellent, whereas the AJCC stage (AUC = 0.60) and Cox score (AUC = 0.70) showed less sensitivity. Calibration plot also showed good association between actual proportion and predicted probability for the AI framework (**Figure 2D**). In terms of integrated discrimination index (IDI) and net reclassification index (NRI), performance of the framework was significantly better compared with the AJCC (derivation: χ^2 = 54.93, $P < 0.001$, IDI = 0.30, NRI = 19.62; validation: $\chi^2 = 7.22$, $P < 0.001$, IDI = 0.29, NRI = 11.85) and Cox (derivation: χ^2 = 849.09, $P < 0.001$, IDI = 0.51, NRI = 63.46; validation: χ^2 = 146.44, $P < 0.001$, IDI = 0.46, NRI = 46.11) models (**Table 3**).

Survival Outcomes of the Latent Risk and Stable Subtypes

Taking into consideration the entire dataset, the ratio of the latent risk group to the stable group was found to be ∼8:2 using probabilistic stratification of the AI framework. KM curves were generated to evaluate the prognostic subtypes. The differences between latent risk and stable groups in disease-free survival (DFS) (HR, 4.920; 95% CI, 4.272–5.666; $P < 0.001$; **Figure 3A**) and overall survival (OS) (HR, 3.526; 95% CI, 3.026–4.108; $P < 0.001$; **Figure 3B**) in the training dataset were significant. On the contrary, in the validation dataset, similar results were observed in both DFS (HR, 3.559; 95% CI, 2.500–5.067; $P < 0.001$; **Figure 3C**) and OS (HR, 3.190; 95% CI, 2.150–4.733; $P < 0.001$; **Figure 3D**). The censored subjects-excluded 1-, 3-, and 5-year OS were 95.0, 79.4, and 38.9% vs. 73.2, 36.1, and 2.3%, respectively, in the latent stable group compared to latent risk group, and the DFS were 87.5, 60.0, and 36.4% vs. 54.1, 21.1, and 1.3%, respectively, in the validation dataset.

Potential Applicability of the AI-Prognostic Subtypes

In this paper, an attempt has been made to study whether an AI framework is able to provide guidance for clinical interventions as recommended in NCCN as seen in **Figure 4** (22). While evaluating the effectiveness, PAT can result into significant survival benefit (median survival benefit, 19 months; HR, 0.459; 95% CI, 0.360–0.586; $P < 0.001$) in the latent risk group. However, no significant difference was observed in the latent stable group (HR, 0.800; 95% CI, 0.374–1.713; $P = 0.719$). In case of the local intrahepatic recurrent patients, the AI-framework-derived prognostic subtypes could be effectively utilized to stratify patients who have been significantly benefited from PRT (HR, 4.684; 95% CI, 2.997–7.320; $P < 0.001$) and PRP (HR, 4.625; 95% CI, 2.458–8.704; $P < 0.001$), respectively. On the contrary, the patients who underwent radiotherapy did not show any significant difference as seen in case of the latent risk and stable groups (HR, 1.839; 95% CI, 0.670–5.046; $P = 0.364$). Moreover, chemotherapy did not indicate any significant results of survival amongst the prognostic subtypes (HR, 1.421; 95% CI, 0.574–3.521; $P = 0.482$).

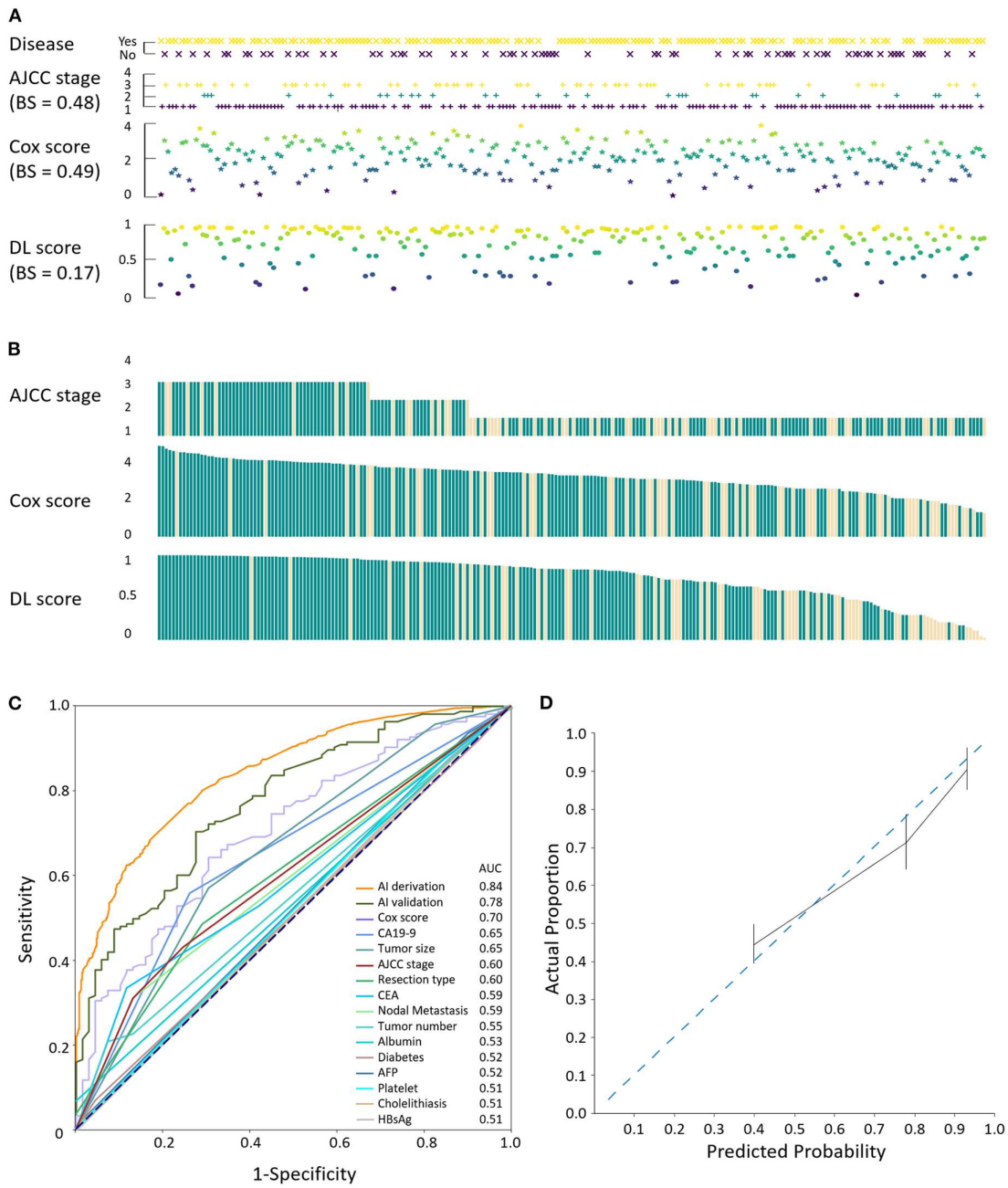


FIGURE 2 | Validation of the ICC AI-framework. **(A)** Evaluation of the consistency between disease status and the AJCC stage, Cox score, and DL, respectively. BS, brier score. **(B)** Coherence comparison among staging/scoring systems. Light yellow, events. **(C)** ROC curves with AUC values of the AI derivation and validation, Cox score, AJCC stage, and involved covariates. **(D)** Calibration plot for evaluation of the actual proportion and predicted proportion of the events using the validation dataset.

DISCUSSION

We adopted a DL approach to learn prognostic prediction using significant clinical factors and created dimidiate prognostic subtypes with distinctive prognosis and efficacy of clinical interventions. This model was compared for accuracy with the most widely used, pre-existing AJCC staging system and the Cox methodology, which was systematically evaluated in context to current clinical standard for recurrent ICC. In comparison to the previous studies on prediction of OS, the current framework specifically caters to cancer-specific survival, excluding mortality due to unknown causes. Moreover, this approach increases the accuracy of equal covariates-generated Cox multivariable hazard proportional model and the stratified prognostic subtypes depicting significant differences amongst various recurring treatments. Collectively, the DL approach

TABLE 3 | Discriminative and risk: reclassification ability of the ICC AI-framework.

	Model performance		IDI (95% CI)	Risk reclassification				
	Change in χ^2	*P*-value		Events		Non-events		NRI (95%CI)
				Risk up	Risk down	Risk up	Risk down	
AI vs. Cox								
Derivation	849.09	<0.001	0.51 (0.50–0.52)	0.90	0.06	0.54	0.34	63.46 (61.68–65.24)
Validation	146.44	<0.001	0.46 (0.44–0.47)	0.88	0.08	0.61	0.29	46.11 (40.56–51.66)
AI vs. AJCC								
Derivation	54.929	<0.001	0.30 (0.29–0.30)	0.64	<0.01	0.48	0.03	19.62 (18.8–20.45)
Validation	7.2197	0.007	0.29 (0.28–0.31)	0.61	<0.01	0.54	0.04	11.85 (9.39–14.32)

IDI, integrated discrimination improvement; CI, confidence interval; NRI, net reclassification improvement; AJCC, American Joint Committee on Cancer.

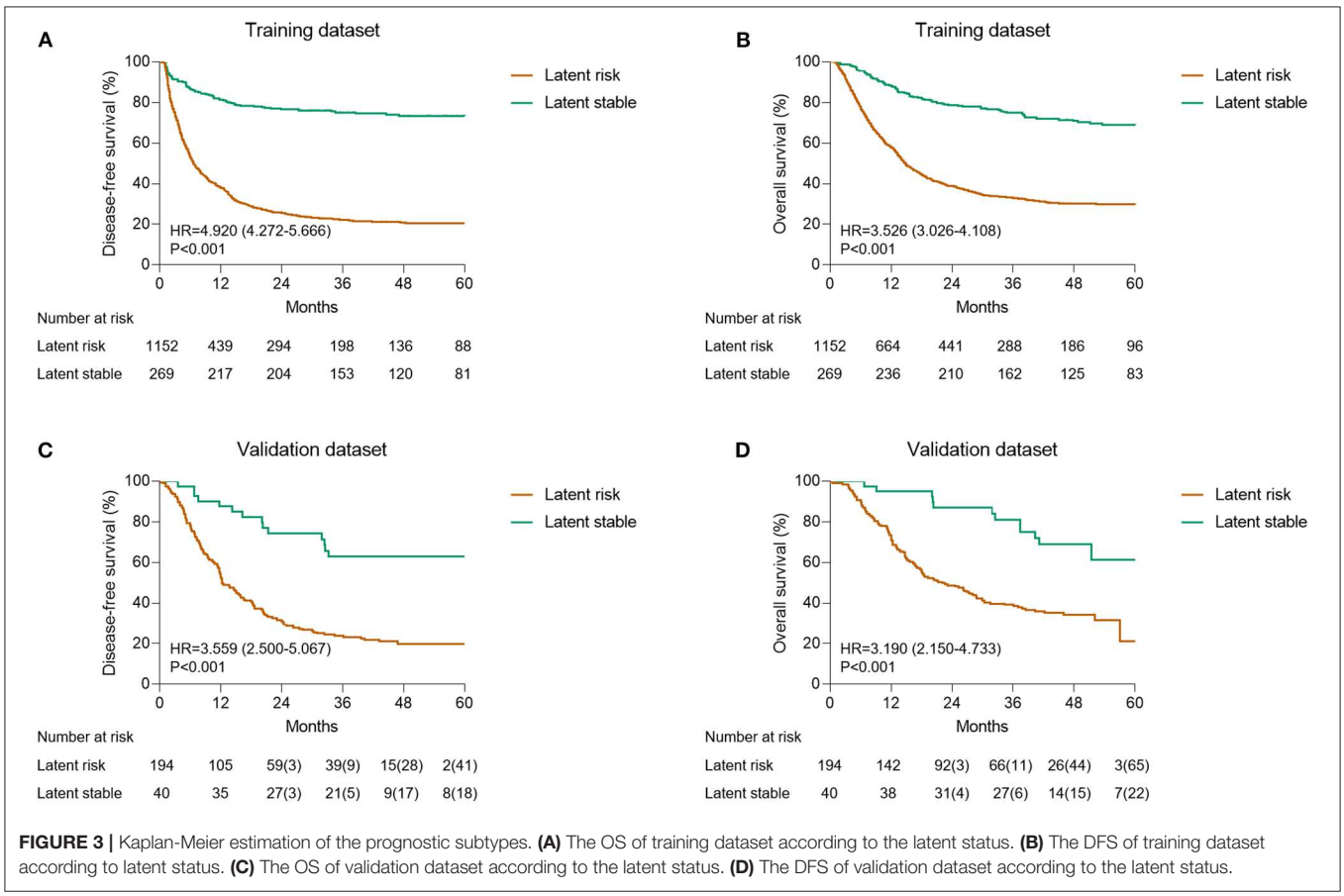


FIGURE 3 | Kaplan-Meier estimation of the prognostic subtypes. **(A)** The OS of training dataset according to the latent status. **(B)** The DFS of training dataset according to latent status. **(C)** The OS of validation dataset according to the latent status. **(D)** The DFS of validation dataset according to the latent status.

was found to be effective in estimation of survival and to devise a categorical strategy to deal with heterogeneity of ICCs by classifying them into latent risk and stable groups for clinical interventions.

We have attempted to maximize the chances for identification of prognostic factors for ICC since it is a disease with diverse outcomes and the issues in identification of the prognostic factors arises due to its exclusivity (23, 24). Therefore, we adopted an 8:2 ratio in randomization of the derivation and validation datasets for meticulous detection of prognostic factors, which enabled us to detect 12 independent prognostic factors. These factors are pre-specified by the Cox hazards regression model, as it is difficult to apply different factors owing to complexity.

Recent studies have demonstrated that a post-operative prophylactic adjuvant therapeutic approach can account for significant survival benefits by preventing events or by prolonging the time-to-recurrence (25–27). Latent risk ICC might be the reason for survival benefits, because the latent

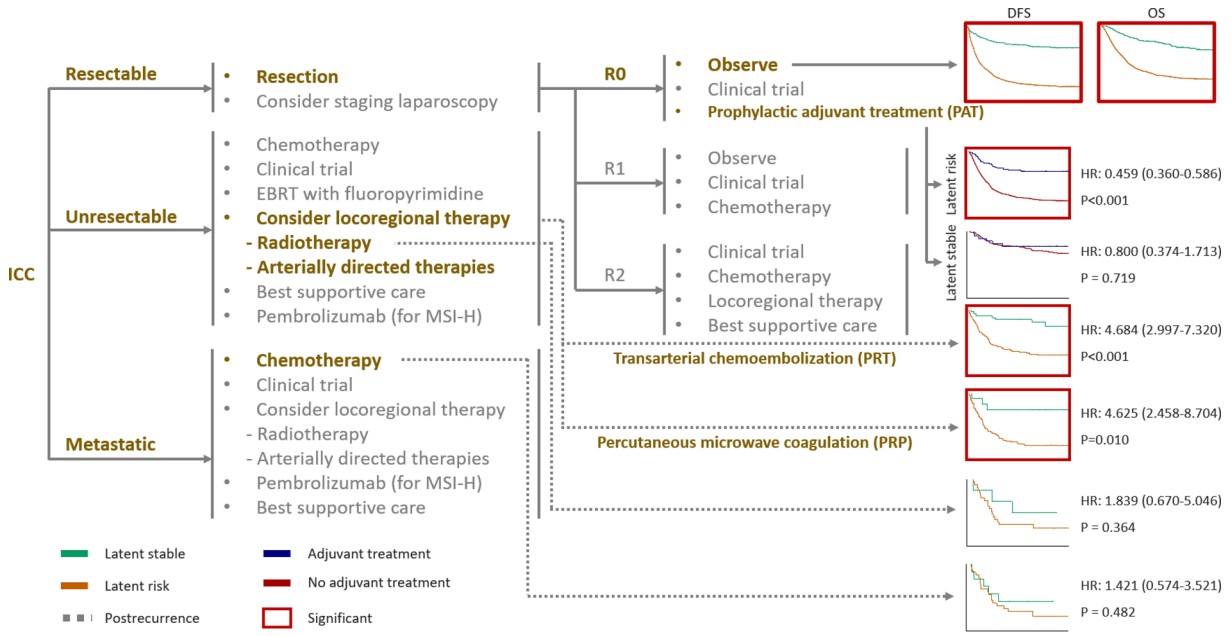


FIGURE 4 | Outcomes of the NCCN guidelines clinical interventions according to the latent status. Kaplan-Meier curves were generated for each clinical intervention, including prophylactic adjuvant treatment and recurrence treatment, according to the latent status. For generation of the survival curves, post-recurrence survival was applied for transarterial chemoembolization, percutaneous microwave coagulation, radiotherapy, and chemotherapy, whereas overall survival was applied for prophylactic adjuvant treatment.

stable group was associated with significantly favorable prognosis without the application of adjuvant treatments.

According to the National Cancer Database of the American College of Surgeons and the American Cancer Society that takes into account ∼70% of the US population, the median OS for chemotherapy (n = 2,176) and chemoradiotherapy (n = 666) are 10.5 (95% CI, 10.0–11.5) and 13.6 (95% CI, 12.3–15.7) months, respectively, for unresectable ICC (28, 29). In our study, estimated latent risk for recurrent ICC treated with chemotherapy and chemoradiotherapy showed no significance compared with the latent stable. Recent studies have suggested, with reasonable evidence, that the application concurrent chemoradiotherapy has better efficacy is better than solely applying adjuvant chemotherapy or radiotherapy (28, 29). Therefore, evaluation of the latent risk combined with chemoradiotherapy may lead to significant survival benefits, though this awaits further validation by future trials.

Although our study provides insights into the use of DL for ICC in a clinical factor setting, some clinical interventions, such as immunotherapy and liver transplantation, are not involved, and the framework is therefore not comprehensive for all circumstances. Furthermore, while we have used our techniques for ICC—the application of the AI-based clinical factors-derived estimative approaches for other tumors might provide auxiliary perspicacious insights. The evaluation of the resection candidate also needs to be considered. There were few patients with regional disease, which is considered not a candidate for surgical resection in some surgery centers. Lastly, the proportion of etiologic subtypes needs to be considered when interpreting the results. There were relatively large proportion of ICCs arose from HBV infection, which is not prevalent in Western countries. Therefore, validation by Western population is essential for the framework to be applied in clinical practice.

In conclusion, the AI approach revealed precision prognostic estimation compared to the AJCC stage for ICC and Cox multivariable regression model in terms of survival prediction and prognostic subtype stratification in patients with ICC after resection. Future validation studies are required to confirm its applicability in patients with ICC from other regions and in other cancers.

DATA AVAILABILITY STATEMENT

Datasets for this study are available from the involved authors under reasonable request.

ETHICS STATEMENT

Ethical approval was not provided for this study on human participants because it was waived by the institutional review board according to retrospective nature of the study. The patients/participants provided their written informed consent.

AUTHOR CONTRIBUTIONS

SJ, YG, JC, QG, JL, HW, QX, and LC designed the study. SJ, YG, JC, QG, GL, BZ, MS, FS, QC, CS, JL, HW, QX, and LC collected

data. SJ, YG, JC, QG, JL, HW, QX, and LC performed analyses. SJ, YG, JC, QG, GL, BZ, MS, FS, QC, CS, JL, HW, QX, and LC were involved in data interpretation. All authors drafted, reviewed, and approved the manuscript.

FUNDING

This study was supported by the National Key R&D Program of China (2017YFA0505803), National Key Research on Precision Medicine of China (2017YFC0908102), and Top Priority Clinical Medical Center of Shanghai Health, and Family Planning Commission (2017ZZ01018).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00143/full#supplementary-material

REFERENCES

1. Hutter M. *Unniversal Artifical Intelligence*. Berlin: Springer (2005).
2. Luger G, Stubblefield W. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. 5th ed. Boston, MA: Addison-Wesley; Benjamin Cummings (2004).
3. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface*. (2018) 15:20170387. doi: 10.1098/rsif.2017.0387
4. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. (2018) 172:1122–31. doi: 10.1016/j.cell.2018.02.010
5. Jusakul A, Cutcutache I, Yong CH, Lim JQ, Huang MN, Padmanabhan N, et al. Whole-genome and epigenomic landscapes of etiologically distinct subtypes of cholangiocarcinoma. *Cancer Discov*. (2017) 7:1116–35. doi: 10.1158/2159-8290.CD-17-0368
6. Saha SK, Zhu AX, Fuchs CS, Brooks GA. Forty-year trends in cholangiocarcinoma incidence in the U.S.: intrahepatic disease on the risk. *Oncologist*. (2016) 21:594–9. doi: 10.1634/theoncologist.2015-0446
7. Sirica AE, Gores GJ, Groopman JD, Selaru FM, Strazzabosco M, Wei Wang X, et al. Intrahepatic cholangiocarcinoma: continuing challenges and translational advances. *Hepatology*. (2019) 69:1803–15. doi: 10.1002/hep.30289
8. Mavros MN, Economopoulos KP, Alexiou VG, Pawlik TM. Treatment and prognosis for patients with intrahepatic cholangiocarcinoma: systematic review and meta-analysis. *JAMA Surg*. (2014) 149:565–74. doi: 10.1001/jamasurg.2013.5137
9. Wirth TC, Vogel A. Surveillance in cholangiocellular carcinoma. *Best Pract Res Clin Gastroenterol*. (2016) 30:987–99. doi: 10.1016/j.bpg.2016.11.001
10. Sirica AE, Gores GJ. Desmoplastic stroma and cholangiocarcinoma: clinical implications and therapeutic targeting. *Hepatology*. (2014) 59:2397–402. doi: 10.1002/hep.26762
11. Razumilava N, Gores GJ. Cholangiocarcinoma. *Lancet*. (2014) 383:2168–79. doi: 10.1016/S0140-6736(13)61903-0
12. de Jong MC, Nathan H, Sotiropoulos GC, Paul A, Alexandrescu S, Marques H, et al. Intrahepatic cholangiocarcinoma: an international multi-institutional analysis of prognostic factors and lymph node assessment. *J Clin Oncol*. (2011) 29:3140–5. doi: 10.1200/JCO.2011.35.6519
13. Lunsford KE, Javle M, Heyne K, Shroff RT, Abdel-Wahab R, Gupta N, et al. Liver transplantation for locally advanced intrahepatic cholangiocarcinoma treated with neoadjuvant therapy: a prospective case-series. *Lancet Gastroenterol Hepatol*. (2018) 3:337–48. doi: 10.1016/S2468-1253(18)30045-1
14. Jeong S, Wang H, Xia Q, Chen L. Liver transplantation for locally advanced intrahepatic cholangiocarcinoma. *Lancet Gastroenterol Hepatol*. (2018) 3:529. doi: 10.1016/S2468-1253(18)30158-4
15. Doussot A, Groot-Koerkamp B, Wiggers JK, Chou J, Gonen M, DeMatteo RP, et al. Outcomes after resection of intrahepatic cholangiocarcinoma: external validation and comparison of prognostic models. *J Am Coll Surg*. (2015) 221:452–61. doi: 10.1016/j.jamcollsurg.2015.04.009
16. Wang Y, Li J, Xia Y, Gong R, Wang K, Yan Z, et al. Prognostic nomogram for intrahepatic cholangiocarcinoma after partial hepatectomy. *J Clin Oncol*. (2013) 31:1188–95. doi: 10.1200/JCO.2012.41.5984
17. Jeong S, Cheng Q, Huang L, Wang J, Sha M, Tong Y, et al. Risk stratification system to predict recurrence of intrahepatic cholangiocarcinoma after hepatic resection. *BMC Cancer*. (2017) 17:464. doi: 10.1186/s12885-017-3464-5
18. Sempoux C, Jibara G, Ward SC, Fan C, Qin L, Roayaie S, et al. Intrahepatic cholangiocarcinoma: new insights in pathology. *Semin Liver Dis*. (2011) 31:49–60. doi: 10.1055/s-0031-1272839
19. Li J, Wang Q, Lei Z, Wu D, Si A, Wang K, et al. Adjuvant transarterial chemoembolization following liver resection for intrahepatic cholangiocarcinoma based on survival risk stratification. *Oncologist*. (2015) 20:640–7. doi: 10.1634/theoncologist.2014-0470
20. Xu Y, Shen Q, Wang N, Wu PP, Huang B, Kuang M, et al. Microwave ablation is as effective as radiofrequency for very-early-stage hepatocellular carcinoma. *Chin J Cancer*. (2017) 36:14. doi: 10.1186/s40880-017-0183-x
21. Kingma, DP, Ba J. Adam: a method for stochastic optimization. In: *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, San Diego, CA (2015).
22. *NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) Hepatobiliary Cancers Version 3.2018 — August 29, 2018: NCCN Guidelines Version 3.2018 Intrahepatic Cholangiocarcinoma*. Available online at: http://www.nccn.org/
23. Beal EW, Tumin D, Moris D, Zhang XF, Chakedis J, Dilhoff M, et al. Cohort contributions to trends in the incidence and mortality of intrahepatic cholangiocarcinoma. *Hepatobiliary Surg Nutr*. (2018) 7:270–6. doi: 10.21037/hbsn.2018.03.16
24. Massarweh NN, El-Serag HB. Epidemiology of hepatocellular carcinoma and intrahepatic cholangiocarcinoma. *Cancer Control*. (2017) 24:1073274817729245. doi: 10.1177/1073274817729245
25. Schweitzer N, Weber T, Kirstein MM, Fischer M, Kratzel AM, Reineke-Plaaß T, et al. The effect of adjuvant chemotherapy in patients with intrahepatic cholangiocarcinoma: a matched pair analysis. *J Cancer Res Clin Oncol*. (2017) 143:1347–55. doi: 10.1007/s00432-017-2392-8
26. Zheng X, Chen B, Wu JX, Jia AY, Rong WQ, Wang LM, et al. Benefit of adjuvant radiotherapy following narrow-margin hepatectomy in patients with intrahepatic cholangiocarcinoma that adhere to major vessels. *Cancer Manag Res*. (2018) 10:3973–81. doi: 10.2147/CMAR.S172940
27. Lin YK, Hsieh MC, Wang WW, Lin YC, Chang WW, Chang CL, et al. Outcomes of adjuvant treatments for resectable intrahepatic cholangiocarcinoma: Chemotherapy alone, sequential chemoradiotherapy, or concurrent chemoradiotherapy. *Radiother Oncol*. (2018) 128:575–83. doi: 10.1016/j.radonc.2018.05.011
28. Verma V, Kusi Appiah A, Lautenschlaeger T, Adeberg S, Simone CB, Lin C. Chemoradiotherapy versus chemotherapy alone for unresected intrahepatic cholangiocarcinoma: practice patterns and outcomes from the national cancer data base. *J Gastrointest Oncol*. (2018) 9:527–35. doi: 10.21037/jgo.2018.01.15
29. Bilimoria KY, Stewart AK, Winchester DP, Ko CY. The National Cancer Data Base: a powerful initiative to improve cancer care in the United States. *Ann Surg Oncol*. (2008) 15:683–90. doi: 10.1245/s10434-007-9747-3

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

frontiers in Oncology

ORIGINAL RESEARCH
published: 20 January 2021
doi: 10.3389/fonc.2020.576901

A Novel Prognostic Scoring System of Intrahepatic Cholangiocarcinoma With Machine Learning Basing on Real-World Data

Zhizhen Li[1†], Lei Yuan[1†], Chen Zhang[2], Jiaxing Sun[3], Zeyuan Wang[4], Yu Wang[5], Xin Hao[5], Fei Gao[3] and Xiaoqing Jiang[1*]*

[1] Department of Biliary Tract Surgery I, Eastern Hepatobiliary Surgery Hospital, Shanghai, China, [2] Winchester School of Art, University of Southampton, Southampton, United Kingdom, [3] Department of Medicine, Beijing Medicinovo Technology Co., Ltd., Beijing, China, [4] School of Computer Science, University of Sydney, Sydney, NSW, Australia, [5] Department of Medicine, Dalian Medicinovo Technology Co., Ltd., Dalian, China

OPEN ACCESS

Edited by:
Jiankun Hu,
Sichuan University, China

Reviewed by:
Qing Wang,
Tsinghua University, China
Yingbin Liu,
Shanghai Jiaotong University, China
Guohao Wu,
Fudan University, China

***Correspondence:**
Fei Gao
gaofei9000@163.com
Xiaoqing Jiang
jiangxiaoqingpro@sina.com

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Surgical Oncology, a section of the journal Frontiers in Oncology

Received: 27 June 2020
Accepted: 07 December 2020
Published: 20 January 2021

Citation:
Li Z, Yuan L, Zhang C, Sun J, Wang Z, Wang Y, Hao X, Gao F and Jiang X (2021) A Novel Prognostic Scoring System of Intrahepatic Cholangiocarcinoma With Machine Learning Basing on Real-World Data. Front. Oncol. 10:576901. doi: 10.3389/fonc.2020.576901

Background and Objectives: Currently, the prognostic performance of the staging systems proposed by the 8th edition of the American Joint Committee on Cancer (AJCC 8th) and the Liver Cancer Study Group of Japan (LCSGJ) in resectable intrahepatic cholangiocarcinoma (ICC) remains controversial. The aim of this study was to use machine learning techniques to modify existing ICC staging strategies based on clinical data and to demonstrate the accuracy and discrimination capacity in prognostic prediction.

Patients and Methods: This is a retrospective study based on 1,390 patients who underwent surgical resection for ICC at Eastern Hepatobiliary Surgery Hospital from 2007 to 2015. External validation was performed for patients from 2015 to 2017. The ensemble of three machine learning algorithms was used to select the most important prognostic factors and stepwise Cox regression was employed to derive a modified scoring system. The discriminative ability and predictive accuracy were assessed using the Concordance Index (C-index) and Brier Score (BS). The results were externally validated through a cohort of 42 patients operated on from the same institution.

Results: Six independent prognosis factors were selected and incorporated in the modified scoring system, including carcinoembryonic antigen, carbohydrate antigen 19-9, alpha-fetoprotein, prealbumin, T and N of ICC staging category in 8th edition of AJCC. The proposed scoring system showed a more favorable discriminatory ability and model performance than the AJCC 8th and LCSGJ staging systems, with a higher C-index of 0.693 (95% CI, 0.663–0.723) in the internal validation cohort and 0.671 (95% CI, 0.602–0.740) in the external validation cohort, which was then confirmed with lower BS (0.103 in internal validation cohort and 0.169 in external validation cohort). Meanwhile, machine learning techniques for variable selection together with stepwise Cox regression for survival analysis shows a better prognostic accuracy than using stepwise Cox regression method only.

Conclusions: This study put forward a modified ICC scoring system based on prognosis factors selection incorporated with machine learning, for individualized prognosis evaluation in patients with ICC.

Keywords: intrahepatic cholangiocarcinoma, prognosis, staging system, machine learning, overall survival

INTRODUCTION

Intrahepatic cholangiocarcinoma (ICC) is a malignant neoplasm originating from the epithelial cells of bile ducts located above the secondary bile duct branch (1). It is the second most common primary malignancy of liver and its incidence has been increasing in recent years (2–4). Surgical resection is the main potentially curative for ICC, the 5-year overall survival (OS) rates after hepatectomy and lymphadenectomy is 15 to 35% (5–9). Appropriate staging for ICC patients can be used to describe the severity and range of involvement of malignant tumors, thus prompting clinicians to understand the prognosis of the disease.

Now the eighth edition of American Joint Committee on Cancer (AJCC 8th) staging system and the Liver Cancer Study Group of Japan (LCSGJ) staging system are widely used in clinical practice (10–13). Although studies have demonstrated that the modified AJCC staging system improves stratifying ability, it remains controversial (14, 15). The LCSGJ staging system focuses on the hepatocellular carcinoma (HCC) which has distinct differences in biological behaviors and postoperative outcomes (16). Some new stratification strategies begin to incorporate readily available clinical parameters, such as carbohydrate antigen 19-9 (CA19-9), alkaline phosphatase (ALP) and alpha-fetoprotein (AFP) (17–19). To more effectively utilize these clinical parameters, not just on surgical-pathological factors, we combined the robust machine learning methods to analyze the high-dimension data in clinical practice.

Meanwhile, the selection of variables which involved in the outcome imputation was significant for staging performance. In similar studies, multivariate analysis using Cox regression to identify the independent prognostic factors for survival was a common method, such as the ICC prognostic staging systems performed by Zhou et al. (19), the modified staging system for mass-forming ICC (16), the Fudan score (17), and in nomogram predicting strategies (18). In present study, we attempted to improve the conventional survival analysis by combining with machine learning algorithms for variable selection, since in the real-world studies, variables are not always independent to each other and they are closely related in the non-linear way. The normal used multivariate analysis methods or linear models cannot capture the complex relationships of variables, which are machine learning methods skilled in, especially we used decision tree-based ensemble methods, i.e., eXtreme Gradient Boosting (XGBoost), random forest (RF), and gradient boosted decision tree (GBDT). The three methods are able to divide and re-aggregate the variables to achieve the minimum prediction error when growing sub-trees. Through this way, the non-linear relationship between variables can be well captured. In addition, they are all with the ability of learning from data with missing values directly, that can better adapt to the data situation in the real world. To confirm their effectiveness, we performed the three variable selection methods for comparison and our proposed method outperforms others by a significant margin. Moreover, our study also incorporated the prognostic factors for TNM staging as an improvement of traditional strategy.

The objective of the current study is to integrate pathological factors and clinical parameters to construct a useful and personalized scoring system with machine learning methods, which can accurately predict the survival outcomes of ICC patients under surgical resection.

MATERIALS AND METHODS

Patients Cohort

The cohort comprised 1,390 pathologically confirmed ICC patients who underwent hepatectomy between January 2007 and October 2015 at the Eastern Hepatobiliary Surgery Hospital (EHBH) in Shanghai, China, which is a high-volume medical center. The data collection was cut-off on November, 2018. Patients diagnosed with Perihilar (Klatskin) tumors and mixed with hepatocellular carcinoma tumors were excluded. All deaths were confirmed to have occurred after ICC recurrence to avoid the interference of competing mortality. The data collection and tumor staging processes were supervised and examined by two pathologists. The patients in external validation cohort (n=42, January 2016 to June 2017) were screened with the same criteria of the internal cohort. The data collection was cut-off on June, 2020. Variable characteristic statistics of the training cohort and external validation cohort were summarized in **Supplemental Table** and **Supplementary Data of Entire Cohort**. The protocol of this study has been approved by the Ethics Committee of the EHBH, and the informed consent has been exempted in the Ethical approval documents.

We collected data of 27 clinical independent variables including provided basic clinical information (age, gender, jaundice, history of stone, history of tumor, and smoking), laboratory results [blood type, hepatitis B virus (HBV), CA19-9, γ-glutamyltranspeptidase (γ-GT), albumin (Alb), alanine aminotransferase (ALT), ALP, prealbumin (PA), aspartate aminotransferase (AST), carcinoembryonic antigen (CEA), AFP, direct bilirubin (DBIL), and total bilirubin (TBIL)], and perioperative data (T/N/M or TNM stage in AJCC 8th, T or TNM stage in LCSGJ, resection type, and tumor size). All laboratory examinations were performed within 1 week before resection or intervention. To be applicable to machine

learning, all relevant variables were cleansed and converted into numerical codes.

Study Design

The aim of this research was to construct a more accurate and simple ICC scoring system for predicting the prognosis after resection based on the clinical factors and stages. Overall Survival for 3 years after resection was the end point in our study. We enriched many types of variables in the initial cohort, and variable selection was implemented *via* three machine learning methods, i.e., XGBoost, RF, and GBDT. The algorithms calculated the contribution of each independent variable to the target variable and obtained the importance score (IS). We combined the intersection variables with the highest IS for further analysis.

Cox proportional hazard models with backward stepwise regression were used to evaluate the impacts of intersection variables on survival, and the prognostic scoring equation was obtained. Overall, the predictive accuracy and discrimination ability between models were compared. In addition, for validating the advantages of the research methods, we compared survival predictions with/without machine learning screening. Since the data collection and research were implemented in the Eastern Hepatobiliary Surgery Hospital (Shanghai, China), this scoring strategy we proposed is simply called EHBH-ICC in the later section. The overall study process is illustrated in **Figure 1**.

Tumor, Node, Metastasis Stage

The 8th edition of AJCC and the LCSGJ staging manual in patients who underwent operations were adopted as baseline models for performance comparison (1, 20).

Machine Learning

In the process of machine learning modeling, we chose the XGBoost, RF, and GBDT for the variable selection, which are capable of dealing with missing values under certain assumptions and do not require data imputation. Since our data was derived from real-world settings with a small number of missing values, machine learning methods with incomplete data learning ability are necessary. We performed these three algorithms using Scikit-learn: a machine learning framework (https://www.scikit-learn.org/stable/) in Python 3.6.8. In order to achieve their best performance, the AutoML (https://github.com/ClimbsRocks/auto_ml) method was adopted to automatically select their hyperparameters.

Statistical and Survival Analysis

Data statistics were characterized as quantity (%) or median (interquartile range, IQR). Mann-Whitney U test and chi-square were used on continuous variables and categorical variables respectively, and $p<0.05$ was considered statistically significant. Relevant prognostic predictors were evaluated by the Cox proportional hazard model using backward stepwise regression (Wald-test, $p<0.05$ represents a significant difference). We ensured comparability of the training and internal validation cohorts, a random distribution was applied in a ratio of 8:2. To estimate the influence of prognostic factors, the hazard ratio (HR) was calculated. Kaplan-Meier analysis was used in survival analysis and log-rank test was adopted to compare significant differences. The Concordance Index (C-index) and Brier Score (BS) were utilized to evaluate the discrimination ability and predictive performance of the staging methods. The higher C-index indicates, the better discrimination ability of the model. BS was an important measure of model calibration, i.e., the mean squared difference between the predicted probability and the actual outcome. The lower BS value indicates the higher prediction accuracy of the model. Statistical analysis and

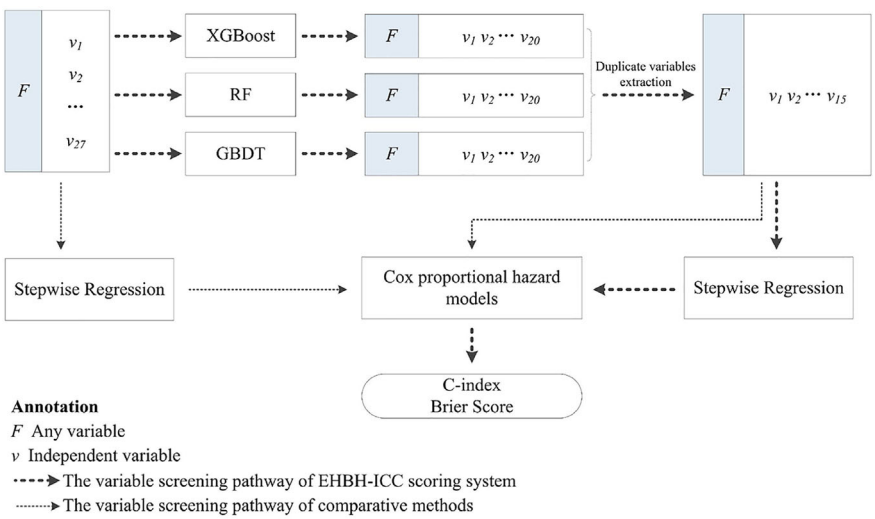


FIGURE 1 | The workflow of this study.

modeling were performed using Python (version 3.6.8) and R Studio (version 1.1.463).

RESULTS

Clinicopathologic Characteristics of Patients

A total of 1,390 patients underwent surgical resection for ICC during the study period. Twenty-seven types of variables included in the primary entire cohort were sorted out and input into the models, patients' demographic information, medical history, tumor information, and examination information were contained in modeling and reported in **Table 1**. The median survival time was 15.5 months (IQR 7.7 to 27.7 months). Of all ICC patients in this study, there were 560 of them (40.3%) having a survival of less than 1 year, 576 patients (41.4%) died between 1 and 3 years after surgery, while 254 (18.2%) died after 3 years. There were 939 females (67.6%) and 451 males (32.4%) enrolled in the study, with a male-to-female ratio of 1:2.1. Among study population, 316 patients

TABLE 1 | Clinicopathologic characteristics of study patients.

Variable types	Variable name		Entire dataset (n=1,390)
Basic information and medical record	Age, year, median (IQR)		55 (46–62)
	Sex, female, n (%)		939 (67.6%)
	Blood type, n (%)	A	413 (29.7%)
		B	379 (27.3%)
		AB	136 (9.8%)
		O	462 (33.2%)
	Jaundice, n (%)		160 (11.5%)
	History of stone, n (%)		264 (19.0%)
	History of tumor, n (%)		101 (7.3%)
	HBV, n (%)		316 (22.7%)
	Smoking, n (%)		374 (26.9%)
Tumor information	Tumor size, cm, median (IQR)		6.0 (4.0–8.2)
	T classification (AJCC 8th), n (%)	T1a and T1b	277 (19.9%)
		T2	186 (13.4%)
		T3	544 (39.1%)
		T4	383 (27.6%)
	N classification (AJCC 8th), n (%)	N0	966 (69.5%)
		N1	424 (30.5%)
	M classification (AJCC 8th), n (%)	M0	1,214(87.3%)
		M1	176 (12.7%)
	TNM stage (AJCC 8th), n (%)	IA and IB	237 (17.1%)
		II	147 (10.5%)
		IIIA	376 (27.1%)
		IIIB	456 (32.8%)
		IV	174 (12.5%)
	T classification (LCSGJ), n (%)	T1	28 (2.0%)
		T2	562 (40.4%)
		T3	540 (38.9%)
		T4	260 (18.7%)
	TNM stage (LCSGJ), n (%)	I	27 (1.9%)
		II	472 (34.0%)
		III	317 (22.8%)
		IVA	108 (7.8%)
		IVB	466 (33.5%)
	Excision, n (%)	R0	1,253 (90.1%)
		R1	54 (3.9%)
		R2	83 (6.0%)
Laboratory results	CA19-9, U/ml, median (IQR)		55.0 (18.0–490.4)
	γ-GT, U/l, median (IQR)		84.0 (44.0–177.0)
	Alb, g/l, median (IQR)		42.1 (39.3–44.6)
	ALT, U/l, median (IQR)		27.1 (17.7–44.8)
	ALP, U/l, median (IQR)		110.0 (83.0–153.0)
	PA, mg/l, median (IQR)		212.0 (170.0–257.0)
	AST, U/l, median (IQR)		29.0 (21.9–42.3)
	CEA, μg/l, median (IQR)		2.9 (1.6–6.0)
	AFP, μg/l, median (IQR)		3.5 (2.29.0)
	DBIL, umol/l, median (IQR)		4.7 (3.5–6.5)
	TBIL, umol/l, median (IQR)		12.6 (9.5–17.2)

IQR, interquartile range; HBV, hepatitis B virus; AJCC 8th, the 8th edition of the American Joint Committee on Cancer staging system; LCSGJ, the Liver Cancer Study Group of Japan staging system; CA19-9, carbohydrate antigen; γ-GT, γ-glutamyltranspeptidase; Alb, albumin; ALT, alanine aminotransferase; ALP, alkaline phosphatase; PA, prealbumin; AST, aspartate aminotransferase; CEA, carcinoembryonic antigen; AFP, alpha-fetoprotein; DBIL, direct bilirubin; TBIL, total bilirubin.

(22.7%) had HBV infection. TNM staging and T classification of AJCC 8th and LCSGJ were evaluated. The T classification (AJCC 8th) includes the extents or existence of tumor diameter, vascular invasion, solitary or multiple tumors, perforation of the visceral peritoneum, and direct invasion of local extrahepatic structures. Nodal and metastasis categories' conditions between the two staging systems were similar, so we counted them together. Only one patient was diagnosed with T1b, that is, had a tumor size larger than 5 cm and without vascular invasion, T1a and T1b tumors were combined in the following study.

Selection and Comparison of Prognostic Factors

The IS of variables, most relevant to patient OS for 3 years were calculated by XGBoost, RF, and GBDT, the top 20 important variables selected from which were assembled in **Table 2**. Then we extracted the intersection of the above variables, and the retained 15 important variables were ALP, γ-GT, N, T, Alb, tumor size, AST, DBIL, TBIL, PA, ALT, AFP, CEA, CA19-9, and age. Among the variables, IS of T staging of AJCC 8th were higher than that of LCSGJ staging system, therefore T (AJCC 8th) was adopted and used "T" as a general name in the following analysis. Variables screened by machine learning participated in developing the Cox proportional hazard regression model. **Table 3** counted the variables in training cohort (n=1,112) used for modeling and the internal validation cohort (n=278) used for verification. The median survival time (months) of training cohort and internal validation cohort was 15.6 (IQR: 7.9–27.7) and 15.3 (IQR: 7.1–27.4), respectively. The data distribution among all factors in cohorts had relative equilibrium ($p>0.05$).

The data sets in **Table 3** were used to perform the Cox regression model, and further screened through backward stepwise regression ($p<0.05$). The results of backward stepwise regression are demonstrated in **Table 4**. The natural logarithmic transformation was applied on the continuous variables to avoid deviation of data distribution. Multivariate analysis by stepwise regression revealed that T classification of AJCC 8th (HR, 1.204; 95% CI, 1.142–1.270), N (HR, 1.927; 95% CI, 1.655–2.243), ln (CEA) (HR, 1.158; 95% CI, 1.098–1.221), ln (CA19-9) (HR, 1.127; 95% CI, 1.085–1.171), ln (AFP) (HR, 1.057; 95% CI, 1.019–1.096), and ln (PA) (HR, 0.830; 95% CI, 0.714–0.964) were determined to be independent predictors of 3-year OS in ICC patients.

Variable Selection Methods Comparison

The Cox regression models with stepwise selection were commonly used in similar studies to select variables, which significantly associated with the prognostic outcome after ICC resection. To verify whether the variable selection incorporated machine learning algorithms can improve the model accuracy or not, we performed three approaches for comparison: only by Cox proportional hazards model with backward stepwise regression (namely SR), only by machine learning (namely ML), and combining both methods (SR+ML) (**Figure 2**). By establishing the survival prediction models, the C-index (**Figure 2A**) and BS (**Figure 2B**) of the above three approaches were obtained, and the results demonstrated that SR+ML (C-index, 0.693; BS, 0.115) had better performance in the most of survival time than only

TABLE 2 | The important variables calculated by XGBoost, random forest (RF), and gradient boosted decision tree (GBDT), and their intersection variables.

No.	XGBoost	IS	RF	IS	GBDT	IS	Intersection variables
1	ALP	0.0792	CA19-9	0.0948	CA19-9	0.1201	ALP
2	Alb	0.0774	ALP	0.0744	T (AJCC8th)[a]	0.1023	γ-GT
3	Age	0.0738	PA	0.0711	ALP	0.0897	N
4	CA19-9	0.0725	γ-GT	0.0645	PA	0.0788	T
5	CEA	0.0724	Tumor size	0.0643	ALT	0.0693	Alb
6	AFP	0.0719	CEA	0.0624	γ-GT	0.0646	Tumor size
7	ALT	0.0707	AST	0.0615	Tumor size	0.0600	AST
8	PA	0.0671	AFP	0.0591	AFP	0.0593	DBIL
9	TBIL	0.0660	Alb	0.0591	CEA	0.0535	TBIL
10	γ-GT	0.0659	ALT	0.0582	Alb	0.0506	PA
11	AST	0.0653	T (AJCC8th)[a]	0.0549	Age	0.0456	ALT
12	Tumor size	0.0617	TBIL	0.0531	DBIL	0.0435	AFP
13	DBIL	0.0615	Age	0.0530	AST	0.0394	CEA
14	T (AJCC8th)[a]	0.0243	DBIL	0.0521	TBIL	0.0388	CA19-9
15	T (LCSGJ)[a]	0.0122	T (LCSGJ)[a]	0.0264	N	0.0329	Age
16	N	0.0114	N	0.0176	T (LCSGJ)[a]	0.0194	
17	M	0.0084	Smoking	0.0098	History of stone	0.0057	
18	Smoking	0.0069	Blood type B	0.0094	M	0.0055	
19	Blood type A	0.0065	Gender	0.0085	History of tumor	0.0036	
20	History of stone	0.0058	Blood type A	0.0082	Blood type AB	0.0036	

[a]*Since the importance score of T (AJCC 8th) in the three models is greater than T (LCSGJ), stage T is merged and only expressed as T in the intersection variables and following article.*
XGBoost, eXtreme Gradient Boosting; RF, random forest; GBDT, gradient boosted decision tree; IS, importance score; ALP, alkaline phosphatase; Alb, albumin; CA19-9, carbohydrate antigen 19-9; CEA, carcinoembryonic antigen; AFP, alpha-fetoprotein; ALT, alanine aminotransferase; PA, prealbumin; TBIL, total bilirubin; γ-GT, γ-glutamyltranspeptidase; AST, aspartate aminotransferase; DBIL, direct bilirubin.

TABLE 3 | Variable characteristic statistics of the training cohort and internal validation cohort.

Variables	Training cohort (n=1,112)	Internal validation cohort (n=278)	*p value*
Age, years, median (IQR)	55 (46–62)	55 (45–62)	0.412
T, n (%)			0.111
T1a and T1b	225 (20.2%)	52 (18.7%)	
T2	154 (13.9%)	32 (11.5%)	
T3	438 (39.4%)	106 (38.1%)	
T4	295 (26.5%)	88 (31.7%)	
N, n (%)			0.046*
N0	787 (70.8%)	179 (64.4%)	
N1	325 (29.2%)	99 (35.6%)	
Tumor size, cm, median (IQR)	5.8 (4.0–8.1)	6.0 (4.1–8.2)	0.209
CA19-9, U/ml, median (IQR)	54.7 (18.4–489.1)	58.3 (16.4–499.8)	0.380
γ-GT, U/l, median (IQR)	85.0 (44.8–178.4)	79.0 (43.0–174.5)	0.253
Alb, g/l, median (IQR)	42.1 (39.4–44.7)	41.8 (38.6–44.3)	0.046*
ALT, U/l, median (IQR)	26.9 (17.6–44.8)	28.4 (18.2–44.7)	0.287
ALP, U/l, median (IQR)	109.0 (83.0–156.0)	113.0 (82.0–148.0)	0.250
PA, mg/l, median (IQR)	212.0 (170.0–259.0)	210.0 (162.3–250.0)	0.070
AST, U/l, median (IQR)	28.8 (21.9–41.1)	29.5 (22.2–45.7)	0.090
CEA, μg/l, median (IQR)	3.0 (1.7–5.9)	2.7 (1.6–6.7)	0.323
AFP, μg/l, median (IQR)	3.6 (2.2–9.2)	3.3 (2.3–8.6)	0.338
DBIL, μmol/l, median (IQR)	4.7 (3.5–6.4)	4.7 (3.5–6.7)	0.173
TBIL, μmol/l, median (IQR)	12.6 (9.4–17.1)	12.7 (9.7–17.5)	0.299

*T and N indicates the staging results of AJCC 8th; IQR, interquartile range; CA19-9, carbohydrate antigen 19-9; γ-GT, γ-glutamyltranspeptidase; Alb, albumin; ALT, alanine aminotransferase; ALP, alkaline phosphatase; PA, prealbumin; AST, aspartate aminotransferase; CEA, carcinoembryonic antigen; AFP, alpha-fetoprotein; DBIL, direct bilirubin; TBIL, total bilirubin. *p < 0.05.*

TABLE 4 | Multivariate regression analysis in the training cohort (n=1,112).

Variables	β	SE (β)	Waldχ^2	HR	95% CI	*p-value*
T	0.186	0.027	6.844	1.204	1.142–1.270	<0.001***
N	0.656	0.078	8.433	1.927	1.655–2.243	<0.001***
ln (CEA)	0.147	0.027	5.426	1.158	1.098–1.221	<0.001***
ln (CA19-9)	0.120	0.019	6.166	1.127	1.085–1.171	<0.001***
ln (AFP)	0.055	0.019	2.972	1.057	1.019–1.096	0.003**
ln (PA)	–0.187	0.077	–2.439	0.830	0.714–0.964	0.015*

*T and N indicates the staging results of AJCC 8th; β, regression coefficient; SE, standard error; HR, hazard ratio; CI, confidence interval; CEA, carcinoembryonic antigen; CA19-9, carbohydrate antigen 19-9; AFP, alpha-fetoprotein; PA, prealbumin. 0.01 < *p < 0.05, 0.001 < **p < 0.01, ***p < 0.001.*

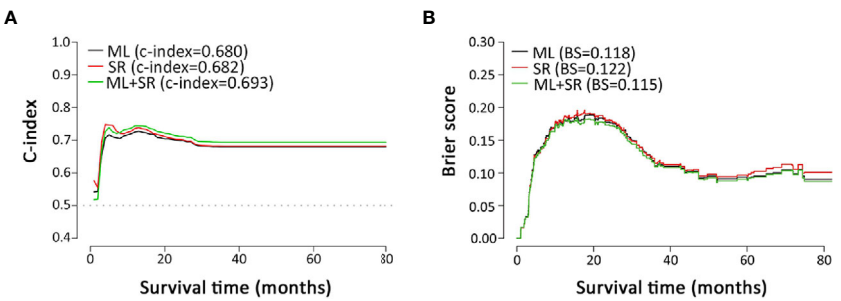


FIGURE 2 | Metrics comparison of models based on different multivariate analysis approaches. **(A, B)** are C-index and brier score comparisons of models based on multivariate analysis by ML, SR, and ML+SR, respectively. ML, machine learning; SR, stepwise regression.

ML and only SR. Therefore, machine learning was proven to capture the prognostic predictors of postoperative outcome more accurately during variable processing, consequently improving the prediction performance of the model. The influenced factors selected *via* only SR including: sex, age, history of stone, smoking habit, HBV, T, N, M, CA19-9, PA, CEA, DBIL, TBIL, excision, and the blood type A. The variables screening results of SR *via* Cox analysis were summarized in **Supplemental Table 2**.

Establishment and Evaluation of Eastern Hepatobiliary Surgery Hospital-Intrahepatic Cholangiocarcinoma Scoring System

Based on the Cox regression, the range of the prognostic index for each individual is from −1.2 to 2.4. In order to adjust the score in our proposed scoring system into positive, we obtained the EHBH-ICC scoring formula as follows:

$$EHBH-ICC\_score = 10 \times \begin{pmatrix} 1.2 + 0.186 \times T + 0.656 \times N + 0.147 \times 1n(CEA) \\ + 0.120 \times 1n(CA19-9) + \\ 0.055 \times 1n(AFP) - 0.187 \times 1n(PA) \end{pmatrix}$$

Histograms of survival risk score distribution for training cohort and internal validation cohort were built based on our EHBH-ICC score (**Figures 3A, B**). According to the score distribution, we divided patients into four risk groups: low (0–10), moderate (11–20), high (21–30), and extremely high (>30). The median risk scores in training and internal validation cohorts were 16.3 and 17.0, respectively. **Figure 4A** displays the good prognostic stratification for patients between stages in internal validation cohort (log rank p<0.001).

Comparison of Predictive Accuracy for Overall Survival in Eastern Hepatobiliary Surgery Hospital-Intrahepatic Cholangiocarcinoma, American Joint Committee on Cancer 8th and the Liver Cancer Study Group of Japan Staging System

Further, we made a comparison of the EHBH-ICC staging system with AJCC 8th and the LCSGJ staging systems. Since

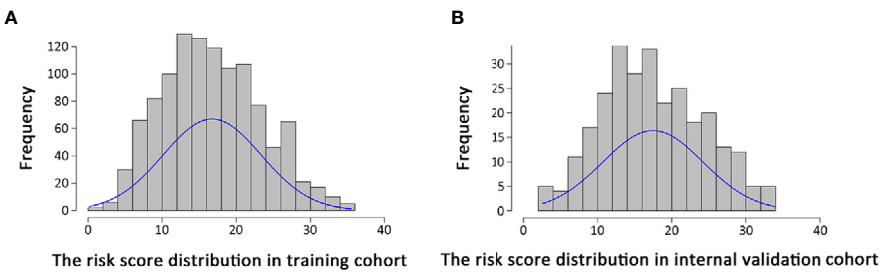


FIGURE 3 | Distribution of risk scores in patients using Eastern Hepatobiliary Surgery Hospital-intrahepatic cholangiocarcinoma (EHBH-ICC) scoring system. **(A, B)** are risk score distributions in training cohort (n=1,112, median=16.3) and internal validation cohort (n=278, median=17.0), respectively.

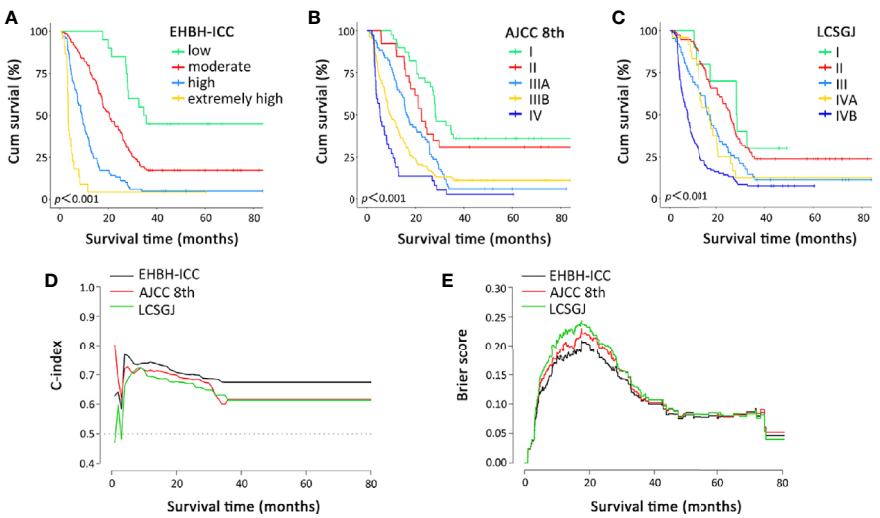


FIGURE 4 | Overall survival curves and prognostic performance indicator curves in the Eastern Hepatobiliary Surgery Hospital-intrahepatic cholangiocarcinoma (EHBH-ICC), American Joint Committee on Cancer (AJCC) 8th, and the Liver Cancer Study Group of Japan (LCSGJ) staging systems. **(A–C)** depict the overall survival according to the three staging systems in internal validation cohort, all log rank p<0.001. **(D, E)** present the C-index and brier score change in long-term survival, respectively.

time-to-mortality and time-to-event were crucial to interpret the results, **Figures 4A–C** depict the Kaplan-Meier curves of the three different staging systems. All of three systems in our study appeared a progressive decrease in OS during the study period. The log-rank test proved that all these staging methods have $p<0.001$.

The discrimination ability and prediction performance of EHBH-ICC score model in internal validation cohort and external validation cohort were respectively indicated with higher C-index of 0.693 (95% CI, 0.663–0.723) and 0.671 (95% CI, 0.602–0.740) than the AJCC 8th and LCSGJ staging systems, which were then confirmed with lower probability calibration of BS (0.103 in internal validation cohort and 0.169 external validation cohort). Detailed C-index and BS results are presented in **Table 5** and **Figures 4D, E**. The model evaluation results show that the EHBH-ICC score was the most precise in predicting the survival after resection in this study.

DISCUSSION

ICC is the second most common primary hepatic malignancies after HCC with increasing incidence and mortality worldwide (21, 22). Hepatectomy is considered as the mainstay of curative option for ICC (23). Accurate tumor staging provides the prognostic details, evaluates the risk level appropriately, as well as assists the choice of adjuvant therapeutic options.

At present, the most commonly used staging systems for ICC are the TNM classification systems, among which, the AJCC 8th and LCSGJ are widely approbatory. With relentless efforts of AJCC to improve the prognostic staging of ICC, there are still research evidences that it is inadequate. T1b with single lesion larger than 5 cm without vascular invasion in AJCC was often rare in clinical treatments. And some recent studies indicated that stage II and stage IIIA for ICC patients in AJCC edition failed to show significant prognostic differentiation. Survival time for intrahepatic metastases was sometimes lower than in patients with serous membrane protruding tumors; however, these patients were only at T2 stage. Some recent studies assessed the prognostic performance of the 7th and 8th edition versions of AJCC staging system, proving that there was no remarkable improvement in overall prognostic discrimination, especially in the staging of T3 category (14, 24, 25). While the LCSGJ focuses on the HCC which has distinct differences in biological behaviors and postoperative outcomes. Some modified staging systems for resectable ICC reserved the prognostic factors in TNM classification or combined these two systems as one of the predictors (19, 26). In our investigation, we analyzed the diagnoses of both staging systems above as separate independent variables. We hypothesized that pathology factors are important prognostic factors for postoperative ICC patients but are only partially relevant. Our study was based on multi-dimensional clinical real-world data in relatively larger population, thus we could seek factors affecting postoperative survival of ICC patients with a wider perspective.

We derived 15 important factors by three algorithms concurrently (**Table 2**), and further identified T (AJCC 8th) and N classifications, CEA, CA19-9, AFP, PA as the prognostic predictive factors. Multiple potential tumor biomarkers have been used in evaluating the prognosis of ICC (27–29). For now, many researches have constructed some new assessment systems with diagnostic biomarkers to predict the survival of patients, such as CA19-9, AFP, CEA, ALP, and PA (17, 19, 30). These factors were confirmed by our results and were involved in the outcome scoring of ICC patients. Serum CA 19-9 and CEA were most investigated in prognosis of ICC (17, 18, 31). Jaklitsch et al. had proven that the inclusion of preoperative CA 19-9 and CEA in AJCC and LCSGJ staging systems improved the prognostic survival prediction after resection for ICC (32). Serum AFP is a widely used tumor marker of HCC (33), and the positive serum AFP (>20 ng/ml) is seen in approximately 19% of ICC patients (34). Zhou et al. showed that the lymph node metastasis rate was low in ICC patients with positive AFP (35). PA generated by liver is commonly regarded as a sensitive marker of nutritional status. A study reported that patients with lower PA have poorer outcomes in ICC (19), which is consistent with our result that PA level is negatively associated with the score. Compared with pathological factors, clinical parameters are easier to obtain and can also provide valuable reference. In our EHBH-ICC scoring system, the diagnosis of T and N and the laboratory results can be directly substituted into the calculation to obtain the corresponding risk level scores.

To our knowledge, our report is the first ICC staging method developed based on machine learning models. In recent years,

TABLE 5 | The comparison of Eastern Hepatobiliary Surgery Hospital (EHBH)-intrahepatic cholangiocarcinoma (ICC), American Joint Committee on Cancer (AJCC) 8th and the Liver Cancer Study Group of Japan (LCSGJ) staging system in internal and external validation cohorts.

Cohorts	Models	C-index (95% CI)	BS
Internal validation	**EHBH-ICC**	**0.693 (0.663–0.723)**	**0.103**
	AJCC 8th	0.675 (0.642–0.708)	0.110
	LCSGJ	0.665 (0.632–0.698)	0.114
External validation	**EHBH-ICC**	**0.671 (0.602–0.740)**	**0.169**
	AJCC 8th	0.648 (0.578–0.718)	0.198
	LCSGJ	0.539 (0.455–0.623)	0.189

CI, confidence interval; BS, brier score; EHBH-ICC, the prognostic scoring system for postoperative intrahepatic cholangiocarcinoma proposed by Eastern Hepatobiliary Surgery Hospital; AJCC 8th, the 8th edition of the American Joint Committee on Cancer staging system; LCSGJ, the Liver Cancer Study Group of Japan staging system.

The discrimination ability and prediction performance of EHBH-ICC score model in internal validation cohort and external validation cohort were respectively indicated with higher C-index of 0.693 (95% CI, 0.663–0.723) and 0.671 (95% CI, 0.602–0.740) than the AJCC 8th and LCSGJ staging systems, which were then confirmed with lower probability calibration of BS (0.103 in internal validation cohort and 0.169 external validation cohort).

machine learning-based methods are widely used in diagnosis, treatment and outcome prediction such as prostate cancer (36), renal cancer (37), non-small cell lung cancer (38), and cardiovascular event prediction (39). Machine learning can deal with different data types even if data are incomplete or incoherent comparing with traditional statistics. Many studies have demonstrated the advantages of machine learning algorithms over traditional statistical methods (40).

According to the EHBH-ICC scoring system, patients are divided into four survival risk grades (low to extremely high). This is a scoring approach to predict the outcome of resectable ICC in Chinese population. The other scoring approach, for instance, the Fudan scoring system was only conducted for 344 patients with multivariate Cox regression. Compared with the Fudan scoring system, the EHBH-ICC has different calculation methods and key prognostic factors. A similarity between Fudan scoring system and our system was the discovery and application of the prognostic value of readily available clinical parameters. Our ultimate validation methods of discrimination ability and performance were C-index and BS. The EHBH-ICC scoring system (C-index, 0.693; BS, 0.103) has more accurate prognostic prediction for ICC patients *via* comparison with the AJCC 8th and LCSGJ edition (**Figures 4D, E**).

In our study, patients' tumor diversity was well reflected. With the continuously increasing sample size, the evaluation system will be more optimized to predict the prognosis of patients more accurately to make decision of the treatment. We cannot only obtain the proportion of risk factors in the prognosis of patients, but also accurately predict the prognosis of patients with the increasing score *via* machine learning.

However, there are limitations in our study. Our study is a retrospective study in one single center. More medical centers and samples could be added to optimize our evaluation system and solve the limitation. In conclusion, the EHBH-ICC scoring system shows good predictive ability for ICC patients who underwent surgical operation *via* evaluation and comparison with existing staging systems (the AJCC 8th and LCSGJ). The machine learning-based EHBH-ICC scoring system can effectively evaluate the ICC prognosis after resections and be used in clinical practice.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The protocol of this study has been approved by the Ethics Committee of the Eastern Hepatobiliary Surgery Hospital, and the informed consent has been exempted in the Ethical approval documents.

AUTHOR CONTRIBUTIONS

ZL and LY conceptualized the study. JS and ZW contributed to the methodology. JS and CZ conducted the formal analysis and investigation. ZL, YW, and XH wrote and prepared the original draft. FG and XJ provided the resources and supervised the study. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.576901/full#supplementary-material.

REFERENCES

1. Liver Cancer Study Group of Japan. *General rules for the clinical and pathological study of primary liver cancer*. *Frist English edition*. Tokyo: Kanehara & Co Ltd (1997).
2. Nathan H, Pawlik TM, Wolfgang CL, Choti MA, Cameron JL, Schulick RD. Trends in survival after surgery for cholangiocarcinoma: a 30 -year population based SEER database analysis. *J Gastrointest Surg* (2007) 11:1488–96. doi: 10.1007/s11605-007-0282-0
3. Njei B. Changing pattern of epidemiology in intrahepatic cholangiocarcinoma. *Hepatology* (2014) 60:1107–8. doi: 10.1002/hep.26958
4. Saha SK, Zhu AX, Fuchs CS, Brooks GA. Forty-year trends in cholangiocarcinoma incidence in the U.S.: Intrahepatic disease on the rise. *Oncologist* (2016) 21:594–9. doi: 10.1634/theoncologist.2015-0446
5. Bridgewater J, Galle PR, Khan SA, Park JW, Patel T, Pawlik TM, et al. Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. *J Hepatol* (2014) 60:1268–89. doi: 10.1016/j.jhep.2014.01.021
6. Doussot A, Groot-Koerkamp B, Wiggers JK, Chou J, Gonen M, DeMatteo RP, et al. Outcomes after resection of intrahepatic cholangiocarcinoma: External validation and comparison of prognostic models. *J Am Coll Surg* (2015) 221:452–61. doi: 10.1016/j.jamcollsurg.2015.04.009
7. Guglielmi A, Ruzzenente A, Campagnaro T, Pachera S, Valdegamberi A, Nicoli P, et al. Intrahepatic cholangiocarcinoma: prognostic factorsafter surgical resection. *World J Surg* (2009) 33:1247–54. doi: 10.1007/s00268-009-9970-0
8. Lang H, Sotiropoulos GC, Sgourakis G, Schmitz KJ, Paul A, Hilgard P, et al. Operations for intrahepatic cholangiocarcinoma: single-institution experience of 158 patients. *J Am Coll Surg* (2009) 208:218–28. doi: 10.1016/j.jamcollsurg.2008.10.017
9. Lin XH, Luo JC. The risk factors and prognostic factors of intrahepatic cholangiocarcinoma. *J Chin Med Assoc* (2017) 80:121–2. doi: 10.1016/j.jcma.2016.12.001
10. Farges O, Fuks D. Clinical presentation andmanagement of intrahepatic cholangiocarcinoma. *Gastroenterol Clin Biol* (2010) 34:191–9. doi: 10.1016/j.gcb.2010.01.006
11. Ohtsuka M, Ito H, Kimura F, Shimizu H, Togawa A, Yoshidome H, et al. Results ofsurgical treatment for intrahepatic cholangiocarcinoma and clinicopathological factors influencing survival. *Br J Surg* (2002) 89:1525–31. doi: 10.1046/j.1365-2168.2002.02268.x
12. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: continuing to build a bridge from a population -based to a more "personalized" approach to cancer staging. *CA Cancer J Clin* (2017) 67:93–9. doi: 10.3322/caac.21388

13. Kudo M, Matsui O, Izumi N, Iijima H, Kadoya M, Imai Y, et al. Surveillance and diagnostic algorithm for hepatocellular carcinoma proposed by the Liver Cancer Study Group of Japan: 2014 Update. *Oncology* (2014) 87 Suppl 1:7–21. doi: 10.1159/000368141
14. Spolverato G, Bagante F, Weiss M, Alexandrescu S, Marques HP, Aldrighetti L, et al. Comparative performances of the 7th and the8th editions of the American Joint Committee on Cancer staging systems for intrahepatic cholangiocarcinoma. *J Surg Oncol* (2017) 115:696–703. doi: 10.1002/jso.24569
15. Lee AJ, Chun YS. Intrahepatic cholangiocarcinoma: the AJCC/UICC 8th edition updates. *Chin Clin Oncol* (2018) 7:52. doi: 10.21037/cco.2018.07.03
16. Uenishi T, Ariizumi S, Aoki T, Ebata T, Ohtsuka M, Tanaka E, et al. Proposal of a new staging system for mass-forming intrahepatic cholangiocarcinoma: a multicenter analysis by the Study Group for Hepatic Surgery of the Japanese Society of Hepato-Biliary-Pancreatic Surgery. *J Hepatobiliary Pancreat Sci* (2014) 21:499–508. doi: 10.1002/jhbp.92
17. Jiang W, Zeng ZC, Tang ZY, Fan J, Sun HC, Zhou J, et al. A prognostic scoring system based on clinical features of intrahepatic cholangiocarcinoma: the Fudan score. *Ann Oncol* (2011) 22:1644–52. doi: 10.1093/annonc/mdq650
18. Wang Y, Li J, Xia Y, Gong R, Wang K, Yan Z, et al. Prognostic nomogram for intrahepatic cholangiocarcinoma after hepatectomy. *J Clin Oncol* (2013) 31:1188–95. doi: 10.1200/JCO.2012.41.5984
19. Zhou H, Jiang X, Li Q, Hu J, Zhong Z, Wang H, et al. A simple and effective prognostic staging system based on clinicopathologic features of intrahepatic cholangiocarcinoma. *Am J Cancer Res* (2015) 5:1831–43.
20. American Joint Committee on Cancer. *AJCC Cancer Staging Manual. 8th Edition*. MB Amin, S Edge, F Greene, DR Byrd, editors. New York: Springer (2017).
21. Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Dicker D, Pain A, Hamavid H, Moradi-Lakeh M, et al. The global burden of cancer 2013. *JAMA Oncol* (2015) 1:505–27. doi: 10.1001/jamaoncol.2015.0735
22. Aljiffry M, Abdulelah A, Walsh M, Peltekian K, Alwayn I, Molinari M. Evidence-based approach to cholangiocarcinoma: a systematic review of thecurrent literature. *J Am Coll Surg* (2009) 208:134–47. doi: 10.1016/j.jamcollsurg.2008.09.007
23. Dodson RM, Weiss MJ, Cosgrove D, Herman JM, Kamel I, Anders R, et al. Intrahepatic cholangiocarcinoma: management options and emerging therapies. *J Am Coll Surg* (2013) 217:736–50.e4. doi: 10.1016/j.jamcollsurg.2013.05.021
24. Kang SH, Hwang S, Lee YJ, Kim KH, Ahn CS, Moon DB, et al. Prognostic comparison of the 7th and 8th editions of the American Joint Committee on Cancer staging system for intrahepatic cholangiocarcinoma. *J Hepatobiliary Pancreat Sci* (2018) 25:240–8. doi: 10.1002/jhbp.543
25. Kim Y, Moris DP, Zhang XF, Bagante F, Spolverato G, Schmidt C, et al. Evaluation of the 8th edition American Joint Commission on Cancer (AJCC) staging system for patients with intrahepatic cholangiocarcinoma: A surveillance, epidemiology, and end results (SEER) analysis. *J Surg Oncol* (2017) 116:643–50. doi: 10.1002/jso.24720
26. Meng ZW, Pan W, Hong HJ, Chen JZ, Chen YL. Macroscopic types of intrahepatic cholangiocarcinoma and the eighth edition of AJCC/UICC TNM staging system. *Oncotarget* (2017) 8:101165–74. doi: 10.18632/oncotarget.20932
27. Endo I, Gonen M, Yopp AC, Dalal KM, Zhou Q, Klimstra D, et al. Intrahepatic cholangiocarcinoma: rising frequency, improved survival, and determinants of outcome after resection. *Ann Surg* (2008) 248:84–96. doi: 10.1097/SLA.0b013e318176c4d3
28. Paik KY, Jung JC, Heo JS, Choi SH, Choi DW, Kim YI. What prognostic factors are important for resected intrahepatic cholangiocarcinoma? *J Gastroenterol Hepatol* (2008) 23:766–70. doi: 10.1111/j.1440-1746.2007.05040.x
29. Rahnemai-Azar AA, Weisbrod A, Dillhoff M, Schmidt C, Pawlik TM. Intrahepatic cholangiocarcinoma: Molecular markers for diagnosis and prognosis. *Surg Oncol* (2017) 26:125–37. doi: 10.1016/j.suronc.2016.12.009
30. Sasaki K, Margonis GA, Andreatos N, Chen Q, Barbon C, Bagante F, et al. Serum tumor markers enhance the predictive power of the AJCC and LCSGJ staging systems in resectable intrahepatic cholangiocarcinoma. *HPB (Oxford)* (2018) 20:956–65. doi: 10.1016/j.hpb.2018.04.005
31. Cho SY, Park SJ, Kim SH, Han SS, Kim YK, Lee KW, et al. Survival analysis of intrahepatic cholangiocarcinoma after resection. *Ann Surg Oncol* (2010) 17:1823–30. doi: 10.1245/s10434-010-0938-y
32. Jaklitsch M, Petrowsky H. The power to predict with biomarkers: carbohydrate antigen 19-9 (CA 19-9) and carcinoembryonic antigen (CEA) serum markers in intrahepatic cholangiocarcinoma. *Transl Gastroenterol Hepatol* (2019) 4:23. doi: 10.21037/tgh.2019.03.06
33. Kudo M, Kitano M, Sakurai T, Nishida N. General rules for the clinical and pathological study of primary liver cancer, nationwide follow-up survey and clinical practice guidelines: The outstanding achievements of the Liver Cancer Study Group of Japan. *Dig Dis* (2015) 33:765–70. doi: 10.1159/000439101
34. Primary liver cancer in Japan. Clinicopathologic features and results of surgical treatment. Liver Cancer Study Group of Japan. *Ann Surg* (1990) 211:277–87.
35. Zhou YM, Yang JM, Li B, Yin ZF, Xu F, Wang B, et al. Clinicopathologic characteristics of intrahepatic cholangiocarcinoma in patients with positive serum a-fetoprotein. *World J Gastroenterol* (2008) 14:2251–4. doi: 10.3748/wjg.14.2251
36. Kim JK, Yook IH, Choi MJ, Lee JS, Park YH, Lee JY, et al. A performance comparison on the machine learning classifiers in predictive pathology staging of prostate cancer. *Stud Health Technol Inform* (2017) 245:1273.
37. Zheng H, Ji J, Zhao L, Chen M, Shi A, Pan L, et al. Prediction and diagnosis of renal cell carcinoma using nuclear magnetic resonance-based serum metabolomics and self-organizing maps. *Oncotarget* (2016) 7:59189–98. doi: 10.18632/oncotarget.10830
38. Yu KH, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun* (2016) 7:12474. doi: 10.1038/ncomms12474
39. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, et al. Cardiovascular event prediction by machine learning: The multi-ethnic study of atherosclerosis. *Circ Res* (2017) 121:1092–101. doi: 10.1161/CIRCRESAHA.117.311312
40. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L, et al. Missing data and multiple imputation in clinical epidemiological research. *Clin Epidemiol* (2017) 9:157–66. doi: 10.2147/CLEP.S129785

Conflict of Interest: JS and FG were employed by company Beijing Medicinovo Technology Co., Ltd. YW and XH were employed by company Dalian Medicinovo Technology Co., Ltd.

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

frontiers in Oncology

ORIGINAL RESEARCH
published: 01 February 2021
doi: 10.3389/fonc.2020.593741

Development and Validation of a Machine Learning Prognostic Model for Hepatocellular Carcinoma Recurrence After Surgical Resection

Yao Huang[1,2,3†], *Hengkai Chen*[1,2,3†], *Yongyi Zeng*[1,2,3], *Zhiqiang Liu*[2,3], *Handong Ma*[4] *and Jingfeng Liu*[1,2,3*]

[1] Liver Disease Center, The First Affiliated Hospital of Fujian Medical University, Fuzhou, China, [2] The United Innovation of Mengchao Hepatobiliary Technology Key Laboratory of Fujian Province, Mengchao Hepatobiliary Hospital of Fujian Medical University, Fuzhou, China, [3] The Liver Center of Fujian Province, Fujian Medical University, Fuzhou, China, [4] Department of Computer Science, Shanghai Jiao Tong University, Shanghai, China

OPEN ACCESS

Edited by:
Francesco Giovinazzo, Catholic University of the Sacred Heart, Italy

Reviewed by:
Alfonso Recordare, Ospedale dell'Angelo, Italy
Han Wu, Eastern Hepatobiliary Surgery Hospital, China

***Correspondence:**
Jingfeng Liu
drjingfeng@126.com

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Surgical Oncology, a section of the journal Frontiers in Oncology

Received: 13 August 2020
Accepted: 17 December 2020
Published: 01 February 2021

Citation:
Huang Y, Chen H, Zeng Y, Liu Z, Ma H and Liu J (2021) Development and Validation of a Machine Learning Prognostic Model for Hepatocellular Carcinoma Recurrence After Surgical Resection. Front. Oncol. 10:593741. doi: 10.3389/fonc.2020.593741

Surgical resection remains primary curative treatment for patients with hepatocellular carcinoma (HCC) while over 50% of patients experience recurrence, which calls for individualized recurrence prediction and early surveillance. This study aimed to develop a machine learning prognostic model to identify high-risk patients after surgical resection and to review importance of variables in different time intervals. The patients in this study were from two centers including Eastern Hepatobiliary Surgery Hospital (EHSH) and Mengchao Hepatobiliary Hospital (MHH). The best-performed model was determined, validated, and applied to each time interval (0–1 year, 1–2 years, 2–3 years, and 3–5 years). Importance scores were used to illustrate feature importance in different time intervals. In addition, a risk heat map was constructed which visually depicted the risk of recurrence in different years. A total of 7,919 patients from two centers were included, of which 3,359 and 230 patients experienced recurrence, metastasis or died during the follow-up time in the EHSH and MHH datasets, respectively. The XGBoost model achieved the best discrimination with a c-index of 0.713 in internal validation cohort. Kaplan-Meier curves succeed to stratify external validation cohort into different risk groups ($p < 0.05$ in all comparisons). Tumor characteristics contribute more to HCC relapse in 0 to 1 year while HBV infection and smoking affect patients' outcome largely in 3 to 5 years.

Abbreviations: α-7-nAChR, 7-nicotinic acetylcholine receptor; APTT, Activated partial thromboplastin time; ALBI, Albumin-bilirubin grade; AJCC, American Joint Committee on Cancer; CPH, Cox Proportional Hazards Model; DeepSurv, Deep Learning-based Survival Model; ERASL, Early Recurrence After Surgery for Liver tumor model; EHSH, Eastern Hepatobiliary Surgery Hospital, Second Military Medical University; XGBoost, Extreme Gradient Boosting; FLD, Fatty liver disease; GSH, glutathione; HBG, Hemoglobin concentration; HCC, Hepatocellular carcinoma; HBV, Hepatitis B virus; MaVI, Macrovascular invasion; MHH, Mengchao Hepatobiliary Hospital, Fujian Medical University; MVI, Microvascular invasion; PVTT, Portal vein tumor thrombosis; PT, Prothrombin time; RSF, Random survival forest; ROS, Reactive oxygen species; RFS, Recurrence-free survival; AFP, Serum alpha-fetoprotein; SLICER, Singapore Liver Cancer Recurrence score; SS-CLIP, Surgery-Specific Cancer of the Liver Italian Program; TT, Thrombin time; TBIL, Total bilirubin.

Based on machine learning prediction model, the peak of recurrence can be predicted for individual HCC patients. Therefore, clinicians can apply it to personalize the management of postoperative survival.

Keywords: hepatocellular carcinoma, recurrence, machine learning, modeling, prognosis

INTRODUCTION

Hepatocellular carcinoma (HCC) is the most common primary liver cancer and ranks as the fourth leading cause of cancer-related mortality (8.2%) worldwide (1). Surgical resection remains the primary curative treatment for patients with adequate liver function (2). However, 50% to 70% of patients who undergo complete tumor resection still suffer from frequent recurrence and disease progression, ultimately leading to unfavorable prognoses (3). Therefore, the identification of patients at high risk of recurrence after surgical resection is essential for clinicians to provide appropriate surveillance and therapy.

During the past decade, researchers have primarily focused on prognosis-predictive models based on biological, demographic, and clinical factors. The most acknowledged system of the American Joint Committee on Cancer (AJCC) tumor-node-metastasis (TNM) is commonly used to determine the staging of liver cancer. However, its prognostic value in predicting tumor recurrence is widely debated (4). Recent models, including the Singapore Liver Cancer Recurrence (SLICER) score, Surgery-Specific Cancer of the Liver Italian Program (SS-CLIP), and the Korean model, were designed to detect tumor recurrence in specific groups of patients. Due to the inaccuracy and diversity of these models, they have not been widely implemented (5–7). In addition, the Early Recurrence After Surgery for Liver tumor (ERASL) model, which is based on Cox regression analysis, has been established to predict early tumor recurrence after liver resection. Despite its better discriminatory performances than other tools, the limited clinical parameters and the prediction for 2-year recurrence restrict its application in the full HCC survivorship management (8).

Machine learning, a field of computer science in which machines mimic, recognize, and learn cognitive functions of the human mind to make empirical predictions, is gaining more and more attention in recent years (9). For cancer, machine learning demonstrates the advantages of image recognition and feature selection compared to traditional methods (10, 11). Recently, automated machine learning algorithms have been developed to detect metastasis in sentinel lymph nodes of women with breast cancer, and showed better diagnostic performance than pathologists (12). In patients with bladder cancer, a novel predictive model based on machine learning algorithms was also created. In the model, disease recurrence after cystectomy was predicted with more than 70% sensitivity and specificity (13). However, few studies have applied a machine learning framework to identify HCC patients with the potential risk of recurrence after curative treatment.

Briefly, we aimed to utilize machine learning algorithms to develop a risk prediction model to predict HCC recurrence among patients who underwent surgical resection. We also explored feature importance in this process, verifying the important prognostic factors for tumor relapse. In addition, a risk heat map covering five years that visually depicts the risk of recurrence was constructed. In this way, we hope to improve the performance of HCC recurrence predictive models using big data and to provide evidential support for individualized management.

MATERIALS AND METHODS

This analysis was reported according to the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) guidelines (14).

Patients

The database was retrospectively derived from patients with HCC who underwent hepatic resection at Eastern Hepatobiliary Surgery Hospital, Second Military Medical University (EHSH) (n = 7,411, from May 2008 to Sept. 2018) or Mengchao Hepatobiliary Hospital, Fujian Medical University (MHH) (n = 508, from Nov. 2014 to Nov. 2018). The patients in this study met the inclusion criteria as follows: (1) pathological confirmation of HCC, (2) Child-Pugh A/B before surgery, (3) R0 surgical resection of tumor with curative intent. However, patients who (1) died within 30 days after surgery or lost to follow-up, (2) received preoperative neoadjuvant treatment (3) diagnosed with extrahepatic cancers, HCC relapse, or metastasis (4) younger than 18 years old were excluded from this study. Inclusion and exclusion of patients and following analysis can be found in **Supplementary Figure 1**.

Different models were constructed on the EHSH dataset, which was randomly divided into derivation and internal validation cohorts at a ratio of 8:2. The models were validated externally using the dataset from MHH. The study was approved by the Ethics Committee of the two centers, and the requirement of written informed consent was waived. All procedures were performed in accordance with the Declaration of Helsinki.

Clinical Variables

The demographics, laboratory tests, and HCC etiologies were collected from the database. The laboratory tests included various parameters of blood examination, liver and coagulation function, and hepatitis virus markers. Tumor characteristics included, but were not limited to, the number of tumors, the diameter of the largest nodule, differentiation, capsule, cirrhosis

in non-cancerous tissues, and vascular invasion. Macrovascular invasion was defined as tumor invasion of large vessels, which can be detected by Computed Tomography/Magnetic Resonance Imaging (CT/MRI) (8). Microvascular invasion refers to the histologically microscopic presence of cancer cell clusters in the blood vessels lined with endothelial cells (15). Thirty-five variables were selected by health professionals based on literature review and clinical expertise.

Follow-up and Outcome

During the follow-up, serum alpha-fetoprotein (AFP) levels were measured, as well as ultrasonography, CT, or MRI of the chest and abdomen once every two months for six months, and then once every three months for the next 1.5 years. For patients who were free of cancer recurrence two years after surgery, a 6-month interval surveillance was carried out. The outcome of this study, recurrence-free survival (RFS), was defined as the time from surgery to the detection of recurrence, metastasis, or death.

General Statistical Principle

After preliminary data cleaning, multiple imputation was performed in R (v3.6.2) based on the Multivariate Imputation by Chained Equations (MICE, v3.8.0). Continuous variables, which were tested for normality by Anderson-Darling tests, were abnormally distributed. Therefore, the variables were summarized by median (IQR), and Wilcoxon rank-sum tests were used for between-groups comparisons. Categorical variables were expressed as frequency (%), and Chi-squared tests or Fisher's exact tests were applied, as appropriate. All statistical analyses above were two-sided, while $p < 0.05$ was considered statistically significant, and conducted in Python (v3.7) with Scipy (v1.4.0) package.

Model Development

Cox Proportional Hazards Model (CPH)

The clinicopathologic parameters of HCC recurrence were fitted by the Cox regression using the Survival package (v3.1) in R-language. Univariable Cox regression was firstly conducted to identify potential predictors ($p < 0.1$). Variables identified in univariable cox model were then applied in multivariable cox regression with stepwise selection method.

Machine Learning Models

Three machine learning models, including Deep Learning-based Survival Model (DeepSurv), Extreme Gradient Boosting (XGBoost), and Random survival forest (RSF) were applied to perform the task of predicting HCC recurrence using all 35 variables preselected. DeepSurv is a multi-layer feed-forward neural network that predicts the effects of diverse variables on their hazard rate parameterized by the weights of the network (16). Based on its algorithm principle, we redeveloped DeepSurv in Python under Pytorch deep learning framework (version 1.3.1, CPU version) and optimized the hyper-parameter search. XGBoost is an improved supervised learning algorithm based on the Gradient Boosting Decision Tree algorithm, which can deal with survival problems by setting partial likelihood functions of the optimization object and log-rank tests as node split criteria (17). Our XGBoost model was implemented in Python using the XGBoost (v.0.9) package. RSF is another machine learning approach for survival analysis that eliminates the proportional hazard assumption and can fit a more general spectrum of survival problems, which conducted in R (randomForestSRC v2.9.3) (18).

Model Discrimination and Calibration

The discrimination performance among the four models in both derivation and validation sets were measured by Harrell's c-index. Comparison of c-index among different models in each cohort was conducted afterwards (19).

As suggested by previous study Kaplan-Meier survival curves for various risk groups were used as informal evidence of discriminative ability (20). Kaplan-Meier curve for the external validation cohort after calibration allows a visual comparison of discrimination among different risk groups at the cut-off of 50th and 84th centiles.

Calibration plots of XGBoost were applied to the derivation and validation sets to determine whether each patient's predicted risk was consistent with the actual outcome. We followed the practice of Chan et al. to draw the calibration plots (8) at 1, 2, 3, and 5 years.

Models in Different Time Intervals and Predictive Heat Map

Inspired by lifetable methodology, we applied XGBoost to different time intervals, including 0 to 1 year, 1 to 2 years, 2 to 3 years, and 3 to 5 years, with the same software. Importance scores were exported, and the Harrell's c-index of each interval were reported at the same time. Furthermore, fifty patients from the external validation cohort were randomly selected to create a heat map for visually illustrating the risk of recurrence within five years after surgery, with aim of providing guidance and support in clinical practice.

RESULTS

Clinicopathologic Features and Outcome

A total of 7,919 patients who underwent surgical resection from two centers were included in the study. 80% of EHSH cohort was assigned as the derivation set (n = 5,928) and the rest was designated as internal validation set (n = 1,483). By the time of data analysis, 3,359 and 230 patients experienced recurrence, metastasis or died during the follow-up time in the EHSH dataset and MHH datasets, respectively. Median follow-up period for two datasets were 3.51 (IQR: 0.41–8.32) and 2.04 (IQR: 0.23–3.88) years. Detailed outcome descriptions are provided in **Supplementary Table 1**.

Thirty-five predictors were included in the final analysis. Preoperative clinical and postoperative pathologic characteristics of the three cohorts are shown in **Table 1**.

Predictive Performance

The discriminatory performance of the four models was assessed with the Harrell's c-index (**Table 2**). The c-index of the Cox regression model in three cohorts were 0.704 (EHSH derivation),

TABLE 1 | Baseline characteristics of patient.

	EHSH derivation (n = 5,928)	EHSH validation (n = 1,483)	MHH validation (n = 508)	p-value*
Gender, male, n (%)	5096 (86.0%)	1305 (88.0%)	437 (86.0%)	0.825
Age (years), median (IQR)	52.0 (44.0–60.0)	51.0 (44.0–59.0)	56.0 (48.0–63.2)	<0.001
Smoking, n (%)	2278 (38.4%)	587 (39.6%)	121 (23.8%)	<0.001
Alcohol consumption, n (%)	1215 (20.5%)	293 (19.8%)	53 (10.4%)	<0.001
FLD, n (%)	209 (3.5%)	72 (4.9%)	96 (18.9%)	<0.001
Ascites, n (%)	149 (2.5%)	45 (3.0%)	43 (8.5%)	<0.001
Cirrhosis, n (%)	5126 (86.5%)	1261 (85.0%)	497 (97.8%)	<0.001
ALBI grade, n (%)				<0.001
1	4546 (76.7%)	1122 (75.7%)	215 (42.3%)	
2	1379 (23.3%)	361 (24.3%)	293 (57.7%)	
3	3 (0.1%)	0 (0.0%)	0 (0.0%)	
Child-Pugh score				0.003
A	5843 (98.6%)	1468 (99.0%)	493 (97.0%)	
B	85 (1.4%)	15 (1.0%)	15 (3.0%)	
HBV history, n (%)	5307 (89.5%)	1334 (90.0%)	206 (40.6%)	<0.001
HBV-DNA load (IU/ml), median (IQR)	1000.0 (1000.0–58000.0)	1000.0 (1000.0–56100.0)	1280.0 (500.0–48400.0)	<0.001
HBsAg, n (%)	5028 (84.8%)	1258 (84.8%)	206 (40.6%)	<0.001
HBsAb, n (%)	809 (13.6%)	207 (14.0%)	55 (10.8%)	0.066
HbcAb, n (%)	5789 (97.7%)	1449 (97.7%)	493 (97.0%)	0.376
HBeAg, n (%)	1474 (24.9%)	365 (24.6%)	114 (22.4%)	0.230
HBeAb, n (%)	4222 (71.2%)	1075 (72.5%)	358 (70.5%)	0.629
AFP (ng/ml), median (IQR)	95.7 (6.5–1210.0)	76.5 (6.7–1210.0)	61.8 (6.2–905.9)	0.148
GGT (IU/L), median (IQR)	66.0 (37.0–120.0)	64.0 (36.0–113.4)	53.0 (30.8–102.0)	<0.001
TBIL (μmol/L), median (IQR)	13.4 (10.3–17.2)	13.4 (10.4–17.5)	15.7 (11.1–21.5)	<0.001
Albumin (g/L), median (IQR)	41.8 (39.4–44.2)	42.0 (39.3–44.2)	39.0 (36.0–42.0)	<0.001
HBG (g/L), median (IQR)	143.0 (132.0–152.0)	144.0 (134.0–152.0)	144.0 (132.0–152.0)	0.586
Prealbumin (mg/L), median (IQR)	218.0 (177.0–264.0)	225.0 (178.0–269.0)	198.0 (151.0–240.2)	<0.001
Platelet (10^9/L), median (IQR)	156.0 (116.0–202.0)	161.0 (121.0–204.0)	169.5 (120.8–219.0)	0.002
PT (s), median (IQR)	11.9 (11.3–12.6)	11.9 (11.4–12.6)	13.5 (13.0–14.2)	<0.001
TT (s), median (IQR)	19.3 (18.3–20.3)	19.3 (18.3–20.3)	17.6 (16.9–18.3)	<0.001
Fibrinogen (mg/dl), median (IQR)	2.4 (2.0–3.0)	2.4 (2.0–3.0)	2.8 (2.4–3.4)	<0.001
APTT (s), median (IQR)	27.4 (25.3–29.9)	27.3 (25.4–29.8)	37.2 (34.7–40.0)	<0.001
Tumor number				<0.001
1	4749 (80.1%)	1198 (80.8%)	437 (86.0%)	
2	729 (12.3%)	181 (12.2%)	54 (10.6%)	
3	170 (2.9%)	36 (2.4%)	0 (0.0%)	
4	63 (1.1%)	17 (1.1%)	0 (0.0%)	
5	217 (3.7%)	51 (3.4%)	17 (3.3%)	
Tumor diameter (cm), median (IQR)	5.2 (3.4–8.5)	5.3 (3.5–8.3)	4.5 (3.0–7.5)	<0.001
Tumor capsule, n (%)	4309 (72.7%)	1070 (72.2%)	399 (78.5%)	0.003
Tumor differentiation, Ⅰ/Ⅱ, n (%)	4845 (81.7%)	1230 (82.9%)	207 (40.7%)	<0.001
Tumor thrombus, n (%)	802 (13.5%)	181 (12.2%)	95 (18.7%)	<0.001
Satellite nodules, n (%)	2594 (43.8%)	650 (43.8%)	139 (27.4%)	<0.001
MaVI, n (%)	1004 (16.9%)	228 (15.4%)	142 (28.0%)	<0.001
MVI, n (%)	2279 (38.4%)	562 (37.9%)	314 (61.8%)	<0.001
Major resection, n (%)	4728 (79.8%)	1169 (78.8%)	425 (83.7%)	0.026
Blood transfusion, n (%)	645 (10.9%)	158 (10.7%)	70 (13.8%)	0.040

IQR, interquartile range (25%–75%).

EHSH, Eastern Hepatobiliary Surgery Hospital; MHH, Mengchao Hepatobiliary Hospital; FLD, fatty liver disease; HBV, hepatitis B virus; AFP, alpha-fetoprotein; TBIL, total bilirubin; PT, prothrombin time; HBG, hemoglobin concentration; GGT, gamma-glutamyl transpeptidase; TT, thrombin time; APTT, activated partial thromboplastin time; MaVI, macrovascular invasion; MVI, microvascular invasion; ALBI grade, albumin-bilirubin grade.

**Comparison between EHSH and MHH cohorts.*

0.700 (EHSH validation), and 0.703 (MHH validation). Among four models, XGBoost achieved the highest c-index in the internal validation cohort (c-index: 0.713, $P < 0.05$, all comparisons). The c-index of XGBoost in the external validation cohort of MHH is 0.697, no statistically significant difference from those of CPH, DeepSurv, and RSF (0.703, $P = 0.470$; 0.700, $P = 0.616$; and 0.699, $P = 0.672$; respectively). Meanwhile, XGBoost model outperformed the Early Recurrence After Surgery for Liver tumor (ERASL) model (c-index: 0.672, $P < 0.001$; 0.673, $P < 0.001$; and 0.679, $P = 0.185$) in all three cohorts with our dataset. Thus, XGBoost was employed for the following demonstration and analysis. KM curves of the external validation dataset (**Figure 1**) indicated good discriminative ability of XGBoost to categorize patients into three risk groups after resection: low risk, intermediate risk ($p < 0.001$ in comparison to the low-risk group), high risk ($p < 0.001$ in comparison to the intermediate-risk group).

As shown in **Figure 2**, the calibration plots demonstrated a satisfying agreement between predictions made by XGBoost and actual patient outcomes in all datasets.

TABLE 2 | Predictive performance (c-index with 95% CI) of the different models.

	EHSH derivation	EHSH validation	MHH validation
CPH	0.704 (0.694–0.712)	0.700 (0.683–0.719)	0.703 (0.671–0.733)
DeepSurv	0.697 (0.687–0.707)	0.698 (0.682–0.718)	0.700 (0.663–0.737)
RSF	0.702 (0.691–0.713)	0.704 (0.685–0.722)	0.699 (0.665–0.730)
XGBoost	0.704 (0.695–0.714)	0.713* (0.698–0.731)	0.697 (0.661–0.728)
ERASL	0.672 (0.663–0.681)	0.673 (0.654–0.690)	0.679 (0.636–0.714)

EHSH, Eastern Hepatobiliary Surgery Hospital; MHH, Mengchao Hepatobiliary Hospital; CPH, Cox Proportional Hazards Regression; DeepSurv, Deep Learning-Based Survival Model; RSF, Random Survival Forest; XGBoost, Extreme Gradient Boosting; ERASL, Early Recurrence After Surgery for Liver tumor models
**p < 0.001 in comparison to DeepSurv and RSF models, p = 0.008 in comparison to CPH model.*

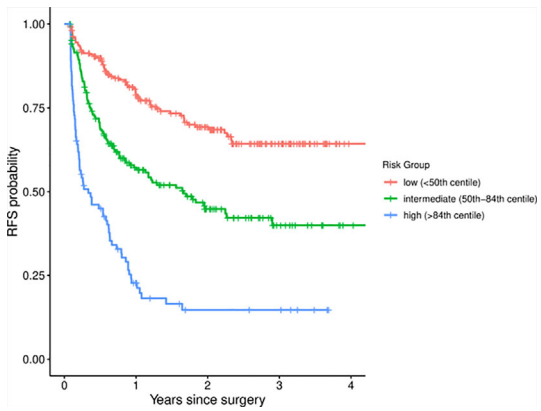


FIGURE 1 | Kaplan-Meier curves for different risk groups among MHH patients. MHH, Mengchao Hepatobiliary Hospital.

Models and Feature Importance in Different Time Intervals

We established the XBGoost model in different time intervals, including 0 to 1 year, 1 to 2 year, 2 to 3 year, and 3 to 5 years, to examine the dynamics of feature importance in HCC patients. The specific predictive performance measurements using c-index and 95% CI for each time slot are listed in **Table 3**.

The variables with the top 10 importance scores are shown in **Table 4**. During 0 to 1 year after resection, the importance score of tumor thrombus (defined as the tumor extending into a vessel, typically portal vein) was 103.01, substantially higher than scores of other factors, such as tumor diameter (33.94), gamma-glutamyl transpeptidase (GGT) (20.25), and tumor capsule (19.22). For 1 to 2 year, tumor number (13.39) was the most important variable related with patient outcomes, followed by resection type (major resection 13.22), tumor thrombus (13.04), and tumor diameter (12.36). In the latter two intervals, apart from tumor number, HBV infection was found to be a relatively important variable. HBV-DNA load has the third highest importance score for 2 to 3 years and HBsAg ranked first in the last period. Furthermore, smoking, an unhealthy lifestyle, was also associated with late recurrence.

The Pattern of Recurrence Risk

Using the XGBoost model in different time intervals, a risk heat map covering four time intervals was developed that visually depicts a patient's risk of tumor recurrence, metastasis or death after undergoing curative liver resection. In general, individual heat map indicated a trend of relatively high recurrence risk in 0 to 1 year and 3 to 5 years after surgical resection (**Figure 3**).

DISCUSSION

HCC is one of the most common malignancies worldwide. Though curative resection offers the best prognosis for patients, disease recurrence remains a major obstacle to the long-term survival of patients (21). Moreover, little is known about the potential risk and peak time periods of HCC recurrence after curative surgery (22, 23). We therefore conducted this research to mediate this gap. In this study, the risk prediction model based on the XGBoost algorithm showed the best c-index in the EHSH validation set. To observe the recurrence risk of individual patients at different time intervals post-surgery, a heat map was constructed based on the XGBoost model for 50 randomly selected HCC patients. The majority of patients had a similar trend of postoperative recurrence that risks in 0 to 1 and 3 to 5 years after surgery were higher than those in 1 to 2 and 2 to 3 years.

In the past few years, several scoring systems have been developed for estimating HCC recurrence risk and stratifying patients. These systems have primarily selected significant clinical parameters through multivariate analyses and constructed conventional Cox proportional hazard models based on the limited risk factors (24–26). One of the important assumptions for Cox proportional hazards regression is that each variable makes linear contribution to model. However, in clinical studies, multiple risk factors usually have non-linear effects with recurrence-free survival, especially in cancer studies (16, 27, 28). Due to this reason, the previous models might fail to show goodness-of-fit and to make accurate prediction. Machine learning algorithms are probably superior than conventional CPH because they can fit more sophisticated non-linear relationship. According to our attempts of building different models, the XGBoost model did better prediction of liver recurrence.

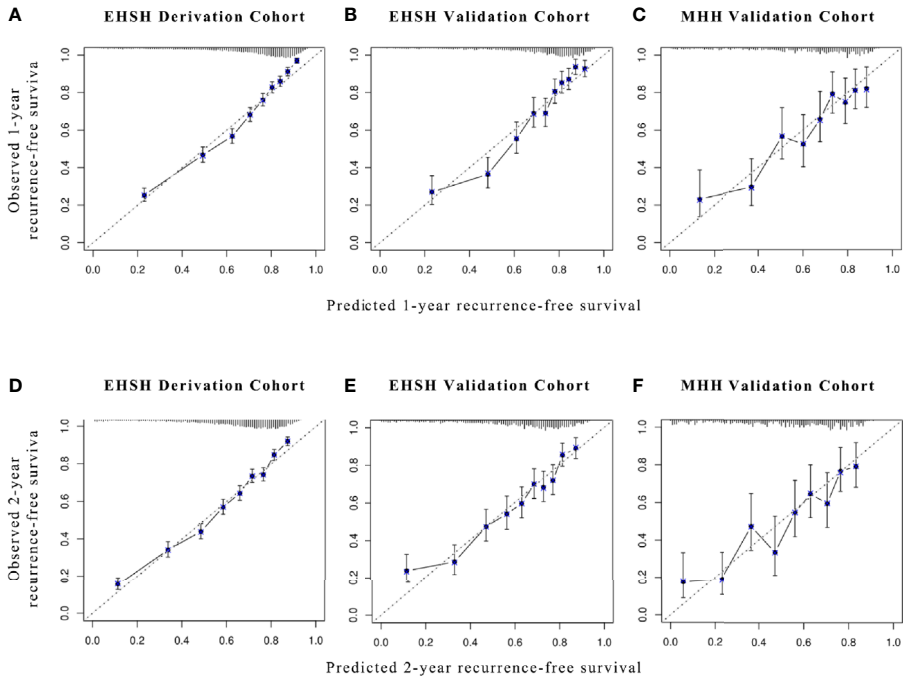


FIGURE 2 | Calibration plots for XGBoost models in predicting 1- and 2-year RFS. Calibration plots for **(A, D)** EHSH derivation cohort, **(B, E)** EHSH validation cohort, and **(C, F)** MHH validation cohort in predicting 1-year **(A–C)** and 2-year RFS **(D–F)**. RFS, recurrence-free survival.

TABLE 3 | Predictive performance (c-index with 95% CI) of XGBoost in four time intervals.

Time Intervals	EHSH derivation	EHSH validation	MHH validation
0–1 year	0.736 (0.726–0.748)	0.751 (0.731–0.772)	0.712 (0.671–0.751)
1–2 years	0.608 (0.579–0.632)	0.551 (0.498–0.604)	0.667 (0.553–0.757)
2–3 years	0.581 (0.545–0.622)	0.571 (0.508–0.641)	NA
3–5 years	0.565 (0.530–0.605)	0.689 (0.625–0.751)	NA

EHSH, Eastern Hepatobiliary Surgery Hospital; MHH, Mengchao Hepatobiliary Hospital; NA, not available.

TABLE 4 | Feature importance during the four time intervals.

No.	0–1 year		1–2 years		2–3 years		3–5 years	
	Features	Importance Score	Features	Importance Score	Features	Importance Score	Features	Importance Score
1	Tumor thrombus	103.01	Tumor number	13.39	Tumor number	8.00	HBsAg	13.26
2	MaVI	37.47	Major resection	13.22	Smoking	7.99	Prealbumin	11.28
3	Tumor diameter	33.94	Tumor thrombus	13.04	HBV-DNA load	7.48	Smoking	8.94
4	MVI	33.63	Tumor diameter	12.36	HBeAg	7.20	Tumor number	8.67
5	GGT	20.25	Satellite nodules	12.01	Major resection	7.14	Age	8.41
6	AFP	19.55	HBV-DNA load	11.89	MaVI	6.95	Platelet	8.40
7	Tumor capsule	19.22	GGT	11.89	Alcohol consumption	6.70	AFP	8.26
8	Blood transfusion	18.21	Albumin	9.94	MVI	6.68	PT	8.21
9	Major resection	17.57	Tumor capsule	9.58	Platelet	6.58	Tumor diameter	8.17
10	Tumor number	15.10	Platelet	8.98	Tumor diameter	6.52	MaVI	8.13

AFP, serum alpha-fetoprotein; GGT, gamma-glutamyl transpeptidase; HBV, hepatitis B Virus; MaVI, macrovascular invasion; MVI, microvascular invasion; PT, prothrombin time.

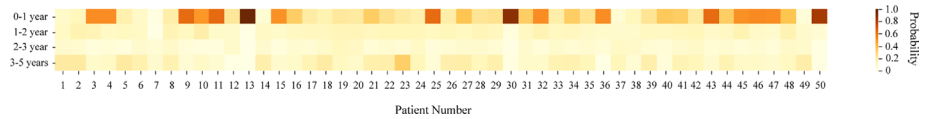


FIGURE 3 | Risk heat map for 50 randomly selected patients.

Apart from an individualized heatmap for illustrating recurrence risk, a feature importance analysis was conducted based on the XGBoost model and was used to evaluate dynamics of variables contributing to the interesting outcome. Specifically, tumor characteristics, such as tumor thrombus, tumor number, tumor size, and tumor differentiation, contributed more to the model's predictive performance in our study. In addition, macrovascular invasion (MaVI), microvascular invasion (MVI), gamma-glutamyl transpeptidase (GGT), intraoperative blood transfusion and major resection also showed a more significant contribution to the predictive performance of the model. Furthermore, smoking as an unhealthy lifestyle also hampered prognosis of HCC patients. These findings are supported by previous research as follows.

Firstly, previous studies found that patients with portal vein tumor thrombosis (PVTT) usually decreased liver function reserves, which was a high-risk factor for disease progression and recurrence (29, 30). In addition to tumor thrombus, tumor volume is also associated with HCC recurrence. In another study, tumor volume was shown to be a predictor of HCC recurrence after liver transplantation (31). A clinical study in Korea confirmed that the maximal size of HCC and the number of tumors were significantly correlated with the recurrence of HCC after liver transplantation (32). In line with our results, MVI was also a unique parameter assessed in the ERASL, SLICER, SS-CLIP, and Korean models (5–8). The dissemination and spread of tumors through micro-vessels may explain the advanced tumor stage, tumor progression, and worse outcomes (33–35).

Secondly, perioperative blood transfusions were independently associated with survival and cancer recurrence after surgical resection (36). A meta-analysis found that allogeneic blood transfusions were associated with poor clinical prognoses in patients with HCC who underwent radical hepatectomy (37). The association between major resection and blood loss as well as RFS of HCC patients has been examined: the more complicated hepatectomy is, the more likely patients are to suffer from intraoperative blood loss, leading to shorter time to recurrence (38).

Thirdly, liver function presented by GGT was another crucial prognostic factor to predict tumor recurrence (39). GGT was first found to modulate the metabolism of glutathione (GSH) and facilitate amino-acid recovery for GSH synthesis (40). Recently, GGT was reported to be involved in tumor initiation, progression, and invasion. As such, GGT may induce the production of endogenous reactive oxygen species (ROS), leaving cells exposed to persistent oxidative stress, leading to DNA damage and tumor growth (41, 42).

Moreover, smoking was associated with an increased risk of HCC (43, 44) and disease-free survival of patients who underwent resection (45). In the current study, we found that smoking was associated with a recurrence risk of 2 to 3 and 3 to 5 years after HCC. The underlying mechanism might be that nicotine increases the expression of α-7-nicotinic acetylcholine receptor (α-7-nAChR), leading to recurrence through the JAK2/STAT3 signaling pathway (46). A previous study found that the history and amount of smoking were both risk factors for the progressive recurrence of HBV-related HCC (47).

Finally, early disease recurrence (0–1 year) is often thought to be a result of intrahepatic metastases, while late recurrence is more likely to result from newly-onset tumors with multicenter origins (48, 49). In accordance with this theory, HBV-DNA load and HBsAg contribute significantly to HCC recurrence from two to five years in our study, which likely induce genomic alternations and pro-oncotic signaling for *de novo* HCC in the long term (50).

Our results suggest that clinicians can provide personalized management of recurrence risk after surgical resection in HCC patients based on information provided by heat maps and feature importance, which may improve postoperative survival outcomes. The risk heat map allows clinical teams to detect patients most at risk of HCC recurrence, schedule appointments for them in the "heat zones" that most likely for recurrence, and take interventions as needed. For example, clinicians may give greater attention to malignant characteristics of tumors, including the presence of tumor thrombus, larger tumor sizes, multiple tumor nodules, and micro- or macro-vascular invasion, if the heat map indicates a high risk within one year after surgery.

There are certain underlying limitations to our study. Firstly, our model is primarily based on two Chinese institutions of patients with HCC in hepatitis B virus-endemic areas. It is necessary to validate our model in international cohorts to extend our results to patients with HCC of various etiologies. Second, some other variables that may be associated with the prognosis of HCC patients, such as postoperative adjunctive therapies and serum inflammatory markers, were not evaluated in this study. In addition, further prospective studies with longer follow-ups are essential to extend the performance of our model further.

In summary, we have developed a model based on a machine learning algorithm that better predicts the risk of disease recurrence in individual patients following hepatic resection in a large population. We further applied this model to four time periods to describe patterns of HCC relapse, and to explore important risk factors. The heat map offers clinicians a decision support tool to identify individuals prone to recurrence, while also allowing clinicians to identify the prognostic factors, which are clinically useful in terms of individualized patient

monitoring, surveillance, and management. Future prospective studies are needed to verify our conclusions.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, upon reasonable request.

AUTHOR CONTRIBUTIONS

YH and HC contributed equally to the manuscript as they both took charge of study design and implementation, as well as drafting manuscript. YZ also participated in study design and literature review. Both ZL and HM conducted statistical and machine learning analysis. Corresponding author, JL, made a huge contribution to the manuscript in terms of revising the draft and reviewing the final version. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by Startup Fund for scientific research, Fujian Medical University (Grant Number: 2019QH1297).

ACKNOWLEDGMENTS

Thanks to all the staff of Mengchao Hepatobiliary Hospital who contributed to the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.593741/full#supplementary-material

Supplementary Figure 1 | Flowchart of inclusion and exclusion criteria and analysis strategy.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2018) 68(6):394–424. doi: 10.3322/caac.21492
2. Netbook N.C.C. *Hepatobiliary Cancers1* (Version 4.2019). (2019). Available at: https://www.nccn.org/professionals/physician_gls/pdf/hepatobiliary.pdf.
3. Tabrizian P, Jibara G, Shrager B, Schwartz M, Roayaie S. Recurrence of hepatocellular cancer after resection: patterns, treatments, and prognosis. *Ann Surg* (2015) 261(5):947–55. doi: 10.1097/SLA.0000000000000710
4. Zhang JF, Shu ZJ, Xie CY, Li Q, Jin XH, GU W, et al. Prognosis of unresectable hepatocellular carcinoma: comparison of seven staging systems (TNM, Okuda, BCLC, CLIP, CUPI, JIS, CIS) in a Chinese cohort. *PloS One* (2014) 9(3):e88182. doi: 10.1371/journal.pone.0088182
5. Ang SF, Ng ESH, Li H, Ong YH, Choo SP, Ngeow J, et al. The Singapore Liver Cancer Recurrence (SLICER) Score for relapse prediction in patients with surgically resected hepatocellular carcinoma. *PloS One* (2015) 10(4):e0118658. doi: 10.1371/journal.pone.0118658
6. Huang S, Huang GQ, Zhu GQ, Liu WY, You J, Shi KQ, et al. Establishment and Validation of SSCLIP Scoring System to Estimate Survival in Hepatocellular Carcinoma Patients Who Received Curative Liver Resection. *PloS One* (2015) 10(6):e0129000. doi: 10.1371/journal.pone.0129000
7. Shim JH, Jun MJ, Han S, Lee YJ, Lee SG, Kim KM, et al. Prognostic nomograms for prediction of recurrence and survival after curative liver resection for hepatocellular carcinoma. *Ann Surg* (2015) 261(5):939–46. doi: 10.1097/SLA.0000000000000747
8. Chan AWH, Zhong J, Berhane S, Toyoda H, Cucchetti A, Shi K, et al. Development of pre and post-operative models to predict early recurrence of hepatocellular carcinoma after surgical resection. *J Hepatol* (2018) 69 (6):1284–93. doi: 10.1016/j.jhep.2018.08.027
9. Camacho DM, Collins KM, Powers RK, Costello JC, Collins JJ. Next-Generation Machine Learning for Biological Networks. *Cell* (2018) 173 (7):1581–92. doi: 10.1016/j.cell.2018.05.015
10. Deo RC. Machine Learning in Medicine. *Circulation* (2015) 132(20):1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593
11. Singal AG, Mukherjee A, Elmunzer BJ, Higgins PD, Lok AS, Zhu J, et al. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. *Am J Gastroenterol* (2013) 108(11):1723–30. doi: 10.1038/ajg.2013.332
12. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *Jama* (2017) 318(22):2199–210. doi: 10.1001/jama.2017.14580
13. Hasnain Z, Mason J, Gill K, Miranda G, Gill IS, Kuhn P, et al. Machine learning models for predicting post-cystectomy recurrence and survival in bladder cancer patients. *PloS One* (2019) 14: (2):e0210976. doi: 10.1371/journal.pone.0210976
14. Moons KM, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Internal Med* (2015) 162(1):1. doi: 10.7326/M14-0698
15. Zhou J, Sun HC, Wang Z, Cong WM, Wang JH, Zeng MS, et al. Guidelines for diagnosis and treatment of primary liver cancer in China (2017 Edition). *Liver Cancer* (2018) 7(3):235–60. doi: 10.1159/000488035
16. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y, et al. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med. Res. Methodol.* (2018) 18: (1):24. doi: 10.1186/s12874-018-0482-1
17. Babajide Mustapha I, Saeed F. Bioactive Molecule Prediction Using Extreme Gradient Boosting. *Molecules* (2016) 21(8):983. doi: 10.3390/molecules21080983
18. Wang H, Zhou L. Random survival forest with space extensions for censored data. *Artif Intell Med* (2017) 79:52–61. doi: 10.1016/j.artmed.2017.06.005
19. Kang L, Chen W, Petrick NA, Gallas BD. Comparing two correlated C indices with right-censored survival outcome: a one-shot nonparametric approach. *Stat Med* (2015) 34(4):685–703. doi: 10.1002/sim.6370
20. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Method* (2013) 13(1):33–3. doi: 10.1186/1471-2288-13-33
21. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet* (2018) 391 (10127):1301–14. doi: 10.1016/S0140-6736(18)30010-2
22. Bruix J, Reig M, Sherman M. Evidence-Based Diagnosis, Staging, and Treatment of Patients With Hepatocellular Carcinoma. *Gastroenterology* (2016) 150(4):835–53. doi: 10.1053/j.gastro.2015.12.041
23. Mahvi DA, Liu R, Grinstaff MW, Colson YL, Raut CP. Local Cancer Recurrence: The Realities, Challenges, and Opportunities for New Therapies. *CA Cancer J Clin* (2018) 68(6):488–505. doi: 10.3322/caac.21498
24. Kamarajah SK, Frankel TL, Sonnenday C, Cho CS, Natahan H. Critical evaluation of the American Joint Commission on Cancer (AJCC) 8th

edition staging system for patients with Hepatocellular Carcinoma (HCC): A Surveillance, Epidemiology, End Results (SEER) analysis. *J Surg Oncol* (2018) 117(4):644–50. doi: 10.1002/jso.24908
25. Cho HJ, Kim B, Lee JD, Kang DR, Kim JK, Lee JH, et al. Development of Risk Prediction Model for Hepatocellular Carcinoma Progression of Indeterminate Nodules in Hepatitis B Virus-Related Cirrhotic Liver. *Am J Gastroenterol* (2017) 112(3):460–70. doi: 10.1038/ajg.2016.480
26. Sasaki K, Firl DJ, Hashimoto K, Fujiki M, Diago-Uso T, Quintini C, et al. Development and validation of the HALT-HCC score to predict mortality in liver transplant recipients with hepatocellular carcinoma: a retrospective cohort analysis. *Lancet Gastroenterol Hepatol* (2017) 2(8):595–603. doi: 10.1016/S2468-1253(17)30106-1
27. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Engl J Med* (2019) 380(14):1347–58. doi: 10.1056/NEJMra1814259
28. Austin PC, Pencinca MJ, Steyerberg EW. Predictive accuracy of novel risk factors and markers: A simulation study of the sensitivity of different performance measures for the Cox proportional hazards regression model. *Stat Methods Med Res* (2017) 26(3):1053–77. doi: 10.1177/0962280214567141
29. Liu PH, Huo TI, Miksad RA. Hepatocellular Carcinoma with Portal Vein Tumor Involvement: Best Management Strategies. *Semin Liver Dis* (2018) 38 (3):242–51. doi: 10.1055/s-0038-1666805
30. Li M, Zhao Y, Liu X, Zhang S, Jiang Y, Yang Z. Early risk warning system for distant metastasis of hepatitis B virus-associated hepatocellular carcinoma with portal vein tumor thrombus. *Oncol Lett* (2020) 19(4):3249–57. doi: 10.3892/ol.2020.11423
31. Kashkoush S, El Moghazy W, Kawahara T, Gala-Lopez B, Toso C, Kneteman NM. Three-dimensional tumor volume and serum alpha-fetoprotein are predictors of hepatocellular carcinoma recurrence after liver transplantation: refined selection criteria. *Clin Transplant* (2014) 28(6):728–36. doi: 10.1111/ctr.12373
32. Chang Y, Cho Y, Lee JH, Lee YB, Cho EJ, Yu SJ, et al. Comparison of Models for Tumor Recurrence after Liver Transplantation for the Patients with Hepatocellular Carcinoma: A Multicenter Long-Term Follow-Up Study. *Cancers (Basel)* (2019) 11(9):1295. doi: 10.3390/cancers11091295
33. Hou YF, Wei YG, Yang JY, Wen TF, Xu MQ, Yan LN, et al. Microvascular invasion patterns affect survival in hepatocellular carcinoma patients after second hepatectomy. *J Surg Res* (2016) 200(1):82–90. doi: 10.1016/j.jss.2015.06.069
34. Feng LH, Dong H, Lau WY, Yu H, Zhu YY, Zhao Y, et al. Novel microvascular invasion-based prognostic nomograms to predict survival outcomes in patients after R0 resection for hepatocellular carcinoma. *J Cancer Res Clin Oncol* (2017) 143(2):293–303. doi: 10.1007/s00432-016-2286-1
35. Lim KC, Chow PKH, Allen JC, Chia GSM, Lim M, Cheow PC, et al. Microvascular invasion is a better predictor of tumor recurrence and overall survival following surgical resection for hepatocellular carcinoma compared to the Milan criteria. *Ann Surg* (2011) 254(1):108–13. doi: 10.1097/SLA.0b013e31821ad884
36. Tai YH, Wu HL, Mandell MS, TSou MY, Chang KY. The association of allogeneic blood transfusion and the recurrence of hepatic cancer after surgical resection. *Anaesthesia* (2020) 75(4):464–71. doi: 10.1111/anae.14862
37. Xun Y, Tian H, Hu L, Yan P, Yang K, Guo T. The impact of perioperative allogeneic blood transfusion on prognosis of hepatocellular carcinoma after radical hepatectomy: A systematic review and meta-analysis of cohort studies. *Medicine (Baltimore)* (2018) 97(43):e12911. doi: 10.1097/MD.0000000000012911
38. Katz SJ, Shia J, Liau KH, Gonen M, Ruo L, Jarnagin W, et al. Operative Blood Loss Independently Predicts Recurrence and Survival After Resection of Hepatocellular Carcinoma. *Ann Surg* (2009) 249(4):617–23. doi: 10.1097/SLA.0b013e31819ed22f
39. Corti A, Franzini M, Paolicchi A, Pompella A. Gamma-glutamyltransferase of cancer cells at the crossroads of tumor progression, drug resistance and drug targeting. *Anticancer Res* (2010) 30(4):1169–81.
40. Kunutsor SK, Apekey TA, Seddoh D. Gamma glutamyltransferase and metabolic syndrome risk: a systematic review and dose-response meta-analysis. *Int J Clin Pract* (2015) 69(1):136–44. doi: 10.1111/ijcp.12507
41. Koenig G, Seneff S. Gamma-Glutamyltransferase: A Predictive Biomarker of Cellular Antioxidant Inadequacy and Disease Risk. *Dis Markers* (2015) 2015:818570. doi: 10.1155/2015/818570
42. Moon DO, Kim BY, Jang JH, Kim MO, Jayasooriya RG, Kang CH, et al. K-RAS transformation in prostate epithelial cell overcomes H2O2-induced apoptosis via upregulation of gamma-glutamyltransferase-2. *Toxicol In Vitro* (2012) 26(3):429–34. doi: 10.1016/j.tiv.2012.01.013
43. Koh WP, Robien K, Wang R, Govindarajan S, Yuan JM, Yu MC. Smoking as an independent risk factor for hepatocellular carcinoma: the Singapore Chinese Health Study. *Br J Cancer* (2011) 105(9):1430–5. doi: 10.1038/bjc.2011.360
44. Koh WP, Yuan JM, Sun CL, Lee HP. Middle-aged and older Chinese men and women in Singapore who smoke have less healthy diets and lifestyles than nonsmokers. *J Nutr* (2005) 135(10):2473–7. doi: 10.1093/jn/135.10.2473
45. Zhang TT, Zhao XQ, Liu Z, Mao ZY, Bai L. Factors affecting the recurrence and survival of hepatocellular carcinoma after hepatectomy: a retrospective study of 601 Chinese patients. *Clin Transl Oncol* (2016) 18(8):831–40. doi: 10.1007/s12094-015-1446-0
46. Li CL, Lin YK, Chen HA, Huang CY, Huang MT, Chang YJ. Smoking as an Independent Risk Factor for Hepatocellular Carcinoma Due to the α7-Nachr Modulating the JAK2/STAT3 Signaling Axis. *J Clin Med* (2019) 8(9):1391. doi: 10.3390/jcm8091391
47. Zhang XF, Wei T, Liu XM, Liu C, Lv Y. Impact of cigarette smoking on outcome of hepatocellular carcinoma after surgery in patients with hepatitis B. *PloS One* (2014) 9(1):e85077. doi: 10.1371/journal.pone.0085077
48. Poon RT, Fan ST, Ng IO, Lo CM, Liu CL, Wong J. Different risk factors and prognosis for early and late intrahepatic recurrence after resection of hepatocellular carcinoma. *Cancer* (2000) 89(3):500–7. doi: 10.1002/1097-0142(20000801)89:3<500::AID-CNCR4>3.0.CO;2-O
49. Furuta M, Ueno M, Fujimoto A, Hiyami S, Yasukawa S, Kojima F, et al. Whole genome sequencing discriminates hepatocellular carcinoma with intrahepatic metastasis from multi-centric tumors. *J Hepatol* (2017) 66(2):363–73. doi: 10.1016/j.jhep.2016.09.021
50. Utsunomiya T, Shimada M, Kudo M, Ichida T, Matsui O, Izumi N, et al. A comparison of the surgical outcomes among patients with HBV-positive, HCV-positive, and non-B non-C hepatocellular carcinoma: a nationwide study of 11,950 patients. *Ann Surg* (2015) 261(3):513–20. doi: 10.1097/SLA.0000000000000821

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

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 22 August 2022
DOI 10.3389/fonc.2022.967758

OPEN ACCESS

EDITED BY
Marco Scarpa,
University Hospital of Padua, Italy

REVIEWED BY
Giovanni Grignani,
Institute for Cancer Research and Treatment (IRCC), Italy
Junjiong Zheng,
Department of Urology, Sun Yat-sen Memorial Hospital, China

*CORRESPONDENCE
Jianghai Chen
chenjianghai@hust.edu.cn
Yuxiong Weng
yxweng1218@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Surgical Oncology, a section of the journal Frontiers in Oncology

RECEIVED 13 June 2022
ACCEPTED 29 July 2022
PUBLISHED 22 August 2022

CITATION
Yan L, Gao N, Ai F, Zhao Y, Kang Y, Chen J and Weng Y (2022) Deep learning models for predicting the survival of patients with chondrosarcoma based on a surveillance, epidemiology, and end results analysis.
Front. Oncol. 12:967758.
doi: 10.3389/fonc.2022.967758

Deep learning models for predicting the survival of patients with chondrosarcoma based on a surveillance, epidemiology, and end results analysis

Lizhao Yan[1†], Nan Gao[1†], Fangxing Ai[1], Yingsong Zhao[2], Yu Kang[1], Jianghai Chen[1*] and Yuxiong Weng[1*]

[1]Department of Hand Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2]Department of Orthopaedics, Liyuan Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Background: Accurate prediction of prognosis is critical for therapeutic decisions in chondrosarcoma patients. Several prognostic models have been created utilizing multivariate Cox regression or binary classification-based machine learning approaches to predict the 3- and 5-year survival of patients with chondrosarcoma, but few studies have investigated the results of combining deep learning with time-to-event prediction. Compared with simplifying the prediction as a binary classification problem, modeling the probability of an event as a function of time by combining it with deep learning can provide better accuracy and flexibility.

Materials and methods: Patients with the diagnosis of chondrosarcoma between 2000 and 2018 were extracted from the Surveillance, Epidemiology, and End Results (SEER) registry. Three algorithms—two based on neural networks (DeepSurv, neural multi-task logistic regression [NMTLR]) and one on ensemble learning (random survival forest [RSF])—were selected for training. Meanwhile, a multivariate Cox proportional hazards (CoxPH) model was also constructed for comparison. The dataset was randomly divided into training and testing datasets at a ratio of 7:3. Hyperparameter tuning was conducted through a 1000-repeated random search with 5-fold cross-validation on the training dataset. The model performance was assessed using the concordance index (C-index), Brier score, and Integrated Brier Score (IBS). The accuracy of predicting 1-, 3-, 5- and 10-year survival was evaluated using receiver operating characteristic curves (ROC), calibration curves, and the area under the ROC curves (AUC).

Results: A total of 3145 patients were finally enrolled in our study. The mean age at diagnosis was 52 ± 18 years, 1662 of the 3145 patients were male (53%), and mean survival time was 83 ± 67 months. Two deep learning models

outperformed the RSF and classical CoxPH models, with the C-index on test datasets achieving values of 0.832 (DeepSurv) and 0.821 (NMTLR). The DeepSurv model produced better accuracy and calibrated survival estimates in predicting 1-, 3- 5- and 10-year survival (AUC:0.895-0.937). We deployed the DeepSurv model as a web application for use in clinical practice; it can be accessed through https://share.streamlit.io/whuh-ml/chondrosarcoma/Predict/app.py.

Conclusions: Time-to-event prediction models based on deep learning algorithms are successful in predicting chondrosarcoma prognosis, with DeepSurv producing the best discriminative performance and calibration.

KEYWORDS

chondrosarcoma, survival analysis, machine learning, DeepSurv, deep learning

Introduction

Chondrosarcoma accounts for 20-30% of primary bone tumors in adulthood and is the second most frequently occurring bone sarcoma behind osteosarcoma (1). Compared to Ewing sarcoma and osteosarcoma, chondrosarcoma is a less malignant disease, with most patients living for 10 years following standard therapy (2). The clinical presentation of chondrosarcoma varies. 90% are conventional chondrosarcomas and 90% of these are low to intermediate-grade tumors. These tumors are slow growing, less likely to metastasize and relatively insensitive to both chemotherapy and radiotherapy (3). The remaining 10-8% of non-conventional tumors are further classified into five subtypes: myxoid, mesenchymal, dedifferentiated, juxtacortical, and clear cell. Those sarcomas (including 5-10% of high-grade conventional chondrosarcomas) can be highly malignant and aggressive, with a higher probability of metastasis, leading to poorer outcomes for patients (4).

Several prognostic models have been created utilizing multivariate Cox regression or machine-learning approaches to predict the 3- and 5-year survival of patients with chondrosarcoma (5–8). Among these models, the nomogram is a frequently used method for integrating and measuring different significant clinical variables of patients when assessing the odds of occurrence of events using the Cox proportional hazards (CoxPH) model. However, one of the underlying assumptions regarding the CoxPH model is that each predictor variable has the same effect at each follow-up time point; however, this overlooks changes in the effect of predictor factors on individual patients at different time points. Additionally, these models use linearity assumptions rather than conducting nonlinear analyses that represent clinical aspects in the real world. As a result, improved solutions focusing on nonlinear variables are required. The Skeletal Oncology Research Group (SORG) algorithm was proposed (5), which trained several binary classification-based machine learning models using the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) data to predict 5-year survival, with the highest AUC being 0.868. The algorithm was subsequently validated on data from two external datasets (9, 10) and showed good performance. Although the SORG algorithm achieves better prediction performance than traditional methods by assessing the nonlinear relationships between variables, its limitations are also obvious. Firstly, it applied a machine learning method to survival data by simplifying the prediction as a binary classification problem; this approach lacks the interpretability and flexibility provided by modeling the probabilities of events as a function of time (11). Secondly, it was trained using data from the SEER database between 2004 and 2010, but data from 2011 to 2018 are already available in the SEER database. Since treatment strategies have evolved in recent years, the patient's clinical characteristics may have changed. Thirdly, the surgical treatment of patients (one of its input features) is not classified in detail. However, the type of surgery may be associated with survival rates (5).

In order to address all of the above-mentioned issues concerning survival predictions, new approaches for combining machine learning methods with survival models have been proposed. Katzman et al. (12) integrated the Cox proportional hazards model with neural networks (DeepSurv) and showed that this novel approach was able to outperform classical Cox models (13, 14). The DeepSurv model used the negative log partial likelihood function to assess patients' survival hazards, utilizing a core hierarchical structure composed of fully connected feed-forward neural networks with a single output node. Yu et al. (15) proposed the Linear Multi-Task Logistic Regression (MTLR) model—an extension of binomial log-likelihood—for jointly modeling a series of binary

labels representing event indicators. It is a collection of logistic regression models constructed at several different time intervals that can be used to assess the probability that the event of interest occurred within each interval. The neural MTLR (N-MTLR) (16) model is based on the MTLR technique but utilizes a deep learning architecture that considers nonlinear relationships in datasets; this method has been shown to outperform the MTLR model in the majority of cases (16). The random survival forest (RSF) model is an extension of the random forest model that takes censoring into account and has been used as a benchmark for method comparison in many pieces of literature (11).

This study aimed to develop models for predicting the overall survival (OS) of patients with chondrosarcoma using the Cox proportional hazards model and three machine learning algorithms and compared the predictive performance of these methods. In addition, the best algorithm will be deployed as an accessible web-based app for clinical use.

Methods

Patient population and data collection

Patients were identified from the SEER database for the period 2000-2018 for this retrospective cohort study. The SEER database collects information from 18 cancer registries and covers approximately 28% of the total US population. SEER*Stat software (Version 8.4.0; National Cancer Institute, Bethesda, MD) was used to extract information from the SEER database. We collected the baseline information of cases (year of diagnosis, gender, age), tumor characteristics (size, number, histologic type, grade, primary site, tumor extension, distant metastasis site, and stage) and treatment details (surgical type, radiotherapy and chemotherapy). The inclusion criteria were as follows: (1) patients have a confirmed diagnosis of chondrosarcoma according to the third edition of the International Classification of Diseases for Oncology (ICD-O-3), morphological code (9220, 9240); (2) bones and joints are the primary site (site recode ICD-O-3/WHO 2008 = Bones and Joints). The exclusion criteria were as follows: (1) survival time is unknown or less than one month; (2) chondrosarcoma was not identified as the primary tumor (first malignant primary indicator = No). A flowchart of the detailed selection process is presented in Figure 1.

Variable's definitions

The following variables are extracted from the SEER database: Year of diagnosis, Age, Gender, Histological type, Primary site, Stage, Grade, Surgery, Radiotherapy, Chemotherapy, Tumor size, Number of tumors, Tumor extension, Distant metastasis, Survival months, Status. The original name of variables in the SEER database and the specific details of each categorical variable was shown in Supplementary Material E1, section S1. Until 2018, The grading system in SEER has been consistent throughout all the years of data collection and consists of a four-tier system with grade IV corresponding to undifferentiated tumors in addition to the common grades I (well), II (moderate) and III (poorly). The new grading strategy "Grade Clinical (2018+)" has been implemented in the SEER database since 2018, which consists of three grades and explicitly mentions that Grade 3 includes undifferentiated tumors.

Deep learning model design

The source code of model development is available on GitHub (https://github.com/WHUH-ML/Chondrosarcoma).

Feature selection

Collinearity occurs when two features have a strong association with one another. Highly correlated features should be avoided since they increase computational cost and effort and they overfit the model. Thus, the cor function in the stats R package was used to calculate correlations between features, with a Pearson's correlation value of 0.7 indicating that features are highly collinear. In addition, univariate and multivariate Cox regression were used to assess the potential features.

Data preprocessing

Binary categorical features were coded as 0 and 1. Ordinal features were encoded as ordinal numeric values, and categorical features were one-hot encoded. We implemented the nonparametric missForest imputation method for handling missing data, which imputes missing values based on random forest predictions. Continuous features were standardized using the StandardScaler function from the sklearn preprocessing library.

Model development

The primary predicted outcome was overall survival (OS). Three algorithms—two based on neural networks (DeepSurv, NMLTR) and one on ensemble learning (RSF)—were selected for training. Meanwhile, a multivariate CoxPH model was also constructed for comparison. The dataset was randomly divided into training and testing datasets at a ratio of 7:3.

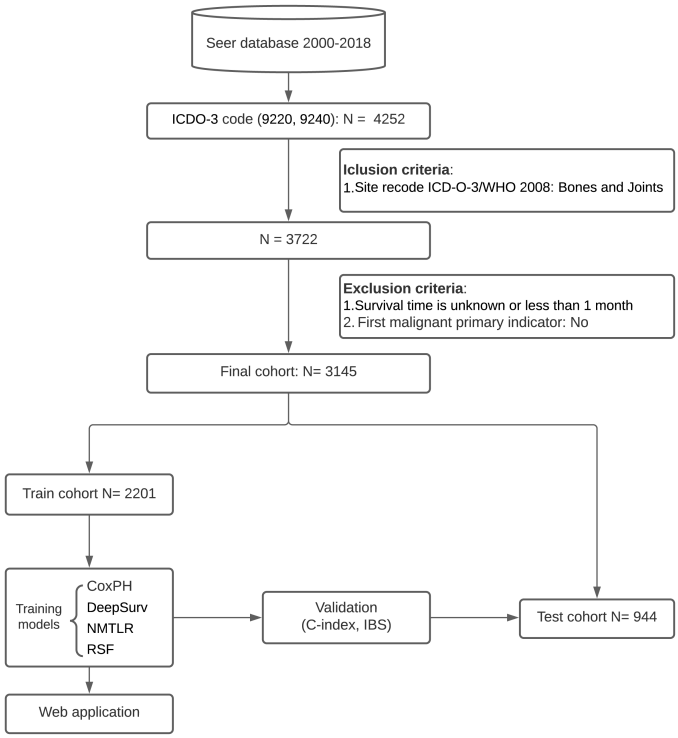


FIGURE 1
Study profile and analysis pipeline.

Hyperparameter tuning

It was essential to find the best configuration for our proposed network, including network architecture and hyperparameter values. Hyperparameter tuning was conducted through a 1000-repeated random search with 5-fold cross-validation on the training dataset. The concordance index (C-index) was used to evaluate the performance of models with different combinations of hyperparameters.

Model evaluation

The accuracy of models was determined using C-index, which is a correlation coefficient between predicted survival risks and observed survival times. A C-index value of 0.5 indicates that the prediction is random, whereas a C-index value of 1.0 indicates excellent prediction. The difference between the two models' C-index was tested using Kang's method (17). Brier scores were also obtained; they indicate the mean square difference between observed patient status and predicted survival probability and are always between 0 and 1, with 0 being the best possible result. A model with a Brier score of less than 0.25 is considered useful in practice. The Integrated Brier Score (IBS) was also calculated to determine the models' overall performance across all available periods. The 1-, 3-, 5- and 10-year OS were calibrated using a calibration curve, comparing expected and observed survival. In order to assess the time-dependent sensitivities and specificities of the models, receiver operating characteristic (ROC) curves were generated, and the area under the curve (AUC) values were calculated for 1-, 3-, 5- and 10-year survival.

Feature importance

To determine the association between individual features and model performance, we estimated the importance of each feature within the test set by replacing the feature data with random numbers (18). The performance of the models, as measured by the concordance index, was then computed using the data after replacement to assess the importance of each feature.

Model deployment

The algorithm with the best performance was deployed using the Streamlit package in Python to create an interactive web-based tool for practical use.

Statistical analysis

All continuous variables in clinical data are displayed as the mean value ± standard deviation (SD). Frequencies and percentages are used to characterize categorical variables. The chi-square test and unpaired two-side t-test were utilized to examine the differences in variables across groups. The R programming language (version 4.1.2) was used to carry out data preprocessing and plotting. The machine learning models were constructed using the PySurvival package in the Python programming language (version 3.6.8).

Results

Basic characteristics

A total of 3145 chondrosarcoma patients registered in the SEER database from 2004 to 2015 were finally enrolled in this study. The patient demographic characteristics are shown in Table 1. 1483 cases were female (47%), and 1662 were male (53%); the mean age was 52 ± 18 years. In terms of the primary site of tumors, 1595 of them were in the extremities (51%), 702 in the axial skeleton (22%), and 848 in other joints and bones (27%). 1033 cases were well-differentiated (39%), 1099 were moderately differentiated (41%), 319 were poorly differentiated (12%), and 208 were undifferentiated (7.8%). 393 cases did not undergo surgery (13%), 1066 underwent a local treatment (35%), 1243 underwent a radical excision with limb salvage (41%), and 358 underwent amputation surgery (12%). The mean overall survival (OS) was 83 ± 67 months, and 904 patients died (29%).

Feature selection and data preprocessing

In the univariate Cox regression, OS was significantly associated with most features except for the year of diagnosis and the number of tumors (Table 1). For the multivariate Cox regression, age, gender, histological type, primary site, grade, surgery, tumor size, tumor extension, and distant metastasis were independent factors for OS ($P<0.05$). Results of the collinearity analysis showed high collinearity between stage and distant metastasis, and between stage and grade (Figure 2). Considered together, we ultimately included nine features (age, gender, histological type, primary site, grade, surgery, tumor size, tumor extension and distant metastasis) in the model development. The dataset was divided into two subsets—training set and testing set; 2203 cases were used for the training set, and the remaining 942 cases were used for the test set (Table 2).

Hyperparameter tuning

After a 1000-repeated random search with 5-fold cross-validation on the training dataset, we selected those parameters showing the highest average C-index in cross-validation as the optimal parameters. The graph of the loss function for the two neural network models (DeepSurv, and NMTLR) is shown in Figure 3. The search space and optimal parameter combinations for models' hyperparameters are displayed in our open-source code on GitHub (https://github.com/WHUH-ML/Chondrosarcoma).

Model comparisons

The predictive performance of the machine learning and CoxPH models is shown in Table 3. In the test dataset, the three machine learning models showed significant ($P < 0.01$) better discrimination (C-index of DeepSurv: 0.832; NMLTR: 0.821; RSF: 0.803) compared with the standard CoxPH model (C-index: 0.773); of the three, DeepSurv had the highest C-index of 0.832. The IBS of the four models were 0.108 (DeepSurv), 0.115 (NMLTR), 0.128 (RSF) and 0.126 (CoxPH) (Figure 4). There is little difference between the C-index obtained from the training data set (DeepSurv: 0.854; NMLTR: 0.850; RSF: 0.829; CoxPH: 0.782) and that from the test set, indicating that the models do not suffer from overfitting.

The calibration plots showed that the consistency between the model's prediction and the actual observation in terms of the 1-, 3-, 5- and 10-year overall survival rates were best for the DeepSurv model, followed by the NMTLR, CoxPH, and RSF models (Figure 5). The AUC was larger for the DeepSurv model than for the three other models (1-year-AUC of DeepSurv: 0.937, NMLTR: 0.896, RSF: 0.900, CoxPH: 0.879; 3-year-AUC of DeepSurv: 0.907, NMLTR: 0.896, RSF: 0.900, CoxPH: 0.879; 5-year-AUC of DeepSurv: 0.895, NMLTR: 0.889, RSF: 0.889, CoxPH: 0.865; 10-year-AUC of DeepSurv: 0.896, NMLTR: 0.890, RSF: 0.885, CoxPH: 0.870) (Figure 5). The results showed that the deep learning models—especially the DeepSurv model—were more accurate in predicting the survival prognosis of chondrosarcoma patients than the RSF and classical CoxPH models.

10.3389/fonc.2022.967758

TABLE 1 Patient demographic, disease, treatment characteristics, and Cox regression analysis.

	Overall	Univariate Cox			Multivariate Cox		
Characteristic	**N = 3,145[1]**	**HR[2]**	**95% CI[2]**	**P-value**	**HR[2]**	**95% CI[2]**	**P-value**
Year of diagnosis				0.23			0.17
2004-2010	1,768 (56%)	—	—		—	—	
2011-2015	1,377 (44%)	1.10	0.94, 1.27		0.85	0.68, 1.07	
Age	52 (18)	1.05	1.05, 1.06	**<0.001**	1.04	1.03, 1.05	**<0.001**
Gender				**<0.001**			**<0.001**
Female	1,483 (47%)	—	—		—	—	
Male	1,662 (53%)	1.48	1.29, 1.69		1.58	1.27, 1.96	
Histological type				**<0.001**			**<0.001**
Conventional	2,879 (92%)	—	—		—	—	
Dedifferentiated	266 (8.5%)	6.30	5.34, 7.42		1.96	1.42, 2.69	
Primary site				**<0.001**			**0.018**
Extremity	1,595 (51%)	—	—		—	—	
Axial skeleton	702 (22%)	1.60	1.37, 1.86		1.09	0.84, 1.42	
Other	848 (27%)	0.77	0.65, 0.91		0.72	0.54, 0.95	
Stage				**<0.001**			0.80
I	1,083 (73%)	—	—		—	—	
II	249 (17%)	3.36	2.68, 4.22		1.21	0.69, 2.14	
III	15 (1.0%)	1.33	0.49, 3.57		0.73	0.21, 2.49	
IV	140 (9.4%)	12.8	10.2, 16.2		1.33	0.46, 3.83	
Missing	1,658						
Grade				**<0.001**			**0.007**
Well differentiated	1,033 (39%)	—	—		—	—	
Moderately differentiated	1,099 (41%)	1.75	1.45, 2.11		1.40	1.05, 1.88	
Poorly differentiated	319 (12%)	4.18	3.36, 5.22		1.73	0.94, 3.20	
Undifferentiated	208 (7.8%)	10.4	8.31, 13.0		2.63	1.38, 5.03	
Missing	486						
Surgery				**<0.001**			**0.002**
No	393 (13%)	—	—		—	—	
Local treatment	1,066 (35%)	0.24	0.20, 0.29		0.54	0.37, 0.80	
Radical excision with limb salvage	1,243 (41%)	0.33	0.28, 0.39		0.48	0.33, 0.68	
Amputation	358 (12%)	0.65	0.53, 0.80		0.62	0.42, 0.90	
Missing	85						
Radiotherapy				**<0.001**			0.39
No	2,822 (90%)	—	—		—	—	
Yes	323 (10%)	1.42	1.17, 1.72		1.15	0.84, 1.56	
Chemotherapy				**<0.001**			0.18
No	2,905 (92%)	—	—		—	—	
Yes	240 (7.6%)	4.92	4.14, 5.83		1.26	0.90, 1.75	
Tumor size, mm	81 (60)	1.00	1.00, 1.01	**<0.001**	1.00	1.00, 1.00	**<0.001**
Missing	1,552						
Number of tumors				0.28			0.23
1	2,867 (91%)	—	—		—	—	
> 1	278 (8.8%)	1.12	0.91, 1.37		0.82	0.59, 1.14	
Tumor extension				**<0.001**			**0.002**
No break in periosteum	553 (29%)	—	—		—	—	
Extension beyond periosteum	1,251 (67%)	2.27	1.81, 2.85		1.50	1.12, 2.00	
Further extension	75 (4.0%)	4.73	3.28, 6.82		2.30	1.41, 3.75	

(Continued)

TABLE 1 Continued

Characteristic	Overall N = 3,145[1]	Univariate Cox HR[2]	95% CI[2]	P-value	Multivariate Cox HR[2]	95% CI[2]	P-value
Missing	1,266						
Distant metastasis				**<0.001**			**0.012**
No	1,792 (93%)	—	—		—	—	
Yes	128 (6.7%)	9.98	8.07, 12.4		3.15	1.11, 8.93	
Missing	1,225						
Survival months	83 (67)						
Status							
Alive	2,241 (71%)						
Dead	904 (29%)						

[1]n (%); Mean (SD).
[2]HR = Hazard Ratio, CI = Confidence Interval.
P values are bolded to indicate they are less than 0.05.

Feature importance

The assessment of feature importance (Figure 6) identified features important to model accuracy for prognosis, with a more than 1% mean reduction in the concordance index with replacement data of age, tumor size, distant metastasis, histological type, grade, tumor extension and primary site.

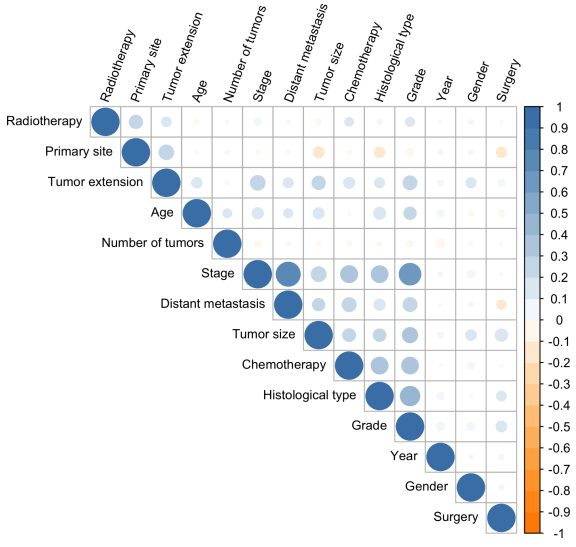


FIGURE 2
Correlation coefficients for each pair of variables in the data set. The estimated correlation values are distributed within the range of -1 to +1. They are represented by color depth, with a number closer to either end value implying a stronger negative correlation or positive correlation.

Algorithm deployment

A visual representation of the functionality and output of the application is presented in Figure 7. The web application, which is primarily for research or informational purposes, can be publicly accessed at https://share.streamlit.io/whuh-ml/chondrosarcoma/Predict/app.py.

Discussion

Accurate prediction for chondrosarcoma survival is crucial for the counseling, follow-up, and treatment planning of patients. Previous studies have revealed various prognostic factors influencing the survival times of patients with chondrosarcoma, including patient age, tumor size, histological type, tumor grade, and metastasis (6, 19–21).. At the same time, increasing amounts of imaging (22, 23) and genetic data (2, 24) are being mined for survival analysis of chondrosarcoma patients. In the face of high-dimensional data, the limitations of the linear relationship between variables assumed by the classical CoxPH model are evident (11). Deep learning is applied to survival analysis due to its ability to comprehensively reveal potential nonlinear relationships in data. In recent years, this method has been gradually improved and successfully applied to clinical (25–27), imaging (28, 29), and genetic data (27). As far as we know, this approach has not been applied to bone tumors. Therefore, we constructed two deep learning models to predict the OS of chondrosarcoma patients and compared the models' performance with two classical models.

By gathering potentially significant characteristics from the SEER database, this study constructed different models for predicting the survival rates of chondrosarcoma patients. We firstly used Cox proportional hazards regression to identify variables related to the prognosis of 3145 individuals with chondrosarcoma. Age, gender, histological type, original

10.3389/fonc.2022.967758

TABLE 2 Characteristic distribution of data in training sets and test sets.

	Level	Overall	Train	Test	P-value
Total		3145	2203	942	
Age (mean (SD))		51.58 (17.53)	51.70 (17.41)	51.29 (17.82)	0.547
Gender (%)	Female	1483 (47.2)	1036 (47.0)	447 (47.5)	0.857
	Male	1662 (52.8)	1167 (53.0)	495 (52.5)	
Histological type (%)	Conventional	2879 (91.5)	2025 (91.9)	854 (90.7)	0.274
	Dedifferentiated	266 (8.5)	178 (8.1)	88 (9.3)	
Primary site (%)	Extremity	1595 (50.7)	1121 (50.9)	474 (50.3)	0.395
	Axial skeleton	702 (22.3)	502 (22.8)	200 (21.2)	
	Other	848 (27.0)	580 (26.3)	268 (28.5)	
Grade (%)	Well differentiated	1033 (38.8)	725 (38.6)	308 (39.5)	0.933
	Moderately differentiated	1099 (41.3)	782 (41.6)	317 (40.7)	
	Poorly differentiated	319 (12.0)	228 (12.1)	91 (11.7)	
	Undifferentiated	208 (7.8)	145 (7.7)	63 (8.1)	
Surgery (%)	None	393 (12.8)	266 (12.4)	127 (14.0)	0.571
	Local treatment	1066 (34.8)	762 (35.4)	304 (33.5)	
	Radical excision with limb salvage	1243 (40.6)	874 (40.6)	369 (40.7)	
	Amputation	358 (11.7)	251 (11.7)	107 (11.8)	
Tumor size, mm (mean (SD))		80.65 (60.19)	80.96 (62.00)	79.88 (55.47)	0.746
Tumor extension (%)	No break in periosteum	553 (29.4)	389 (28.9)	164 (30.8)	0.425
	Extension beyond periosteum	1251 (66.6)	900 (66.8)	351 (66.0)	
	Further extension	75 (4.0)	58 (4.3)	17 (3.2)	
Distant metastasis (%)	Not	1792 (93.3)	1280 (93.5)	512 (92.9)	0.721
	Yes	128 (6.7)	89 (6.5)	39 (7.1)	
Survival months (mean (SD))		83.16 (66.93)	84.50 (66.86)	80.04 (67.01)	0.087
Status (%)	Alive	2241 (71.3)	1572 (71.4)	669 (71.0)	0.882
	Dead	904 (28.7)	631 (28.6)	273 (29.0)	

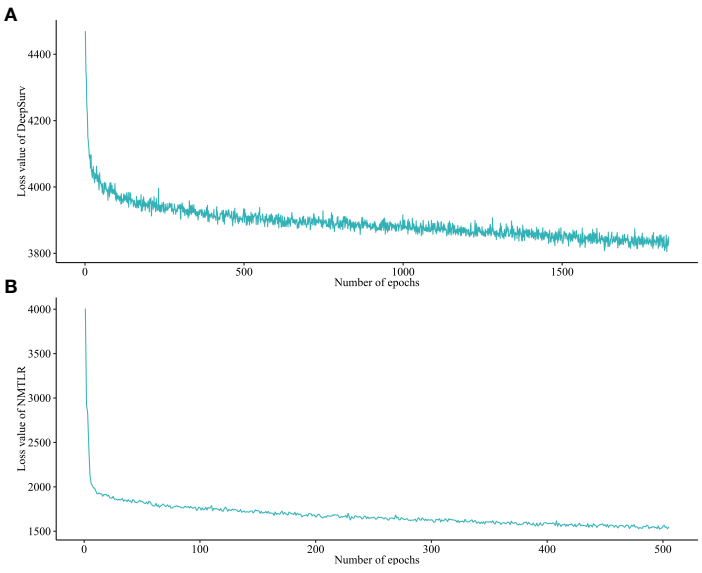


FIGURE 3
Loss convergence graph for **(A)** DeepSurv, **(B)** neural network multitask logistic regression (N-MLTR) models.

TABLE 3 Performance of four survival models.

Models	C index[a]		IBS[a]	1-year AUC[a]	3-year AUC	5-year AUC	10-year AUC
	Train[b]	Test[b]					
CoxPH[a]	0.782	0.773	0.126	0.923 (0.897-0.948)	0.879 (0.852-0.906)	0.865 (0.836-0.893)	0.870 (0.841-0.899)
DeepSurv[a]	**0.854**	**0.832**	**0.108**	**0.937** (0.911-0.962)	**0.907** (0.883-0.931)	**0.895** (0.870-0.920)	**0.896** (0.870-0.921)
NMTLR[a]	0.850	0.821	0.115	0.928 (0.900-0.956)	0.896 (0.870-0.922)	0.889 (0.862-0.915)	0.890 (0.863-0.917)
RSF[a]	0.829	0.803	0.128	0.931 (0.905-0.958)	0.900 (0.873-0.926)	0.889 (0.862-0.916)	0.885 (0.857-0.913)

[a]CoxPH, standard cox proportional hazards; NMLTR, neural multi-task logistic regression; RSF, random survival forest; IBS, Integrated Brier Score; AUC, area under receiver operating characteristic curve. C index, concordance index.
[b]C index in train and test dataset are calculated separately, other metrics are calculated only in the test set.
Bolded metrics indicate that the metric is the best of the fourgroups.

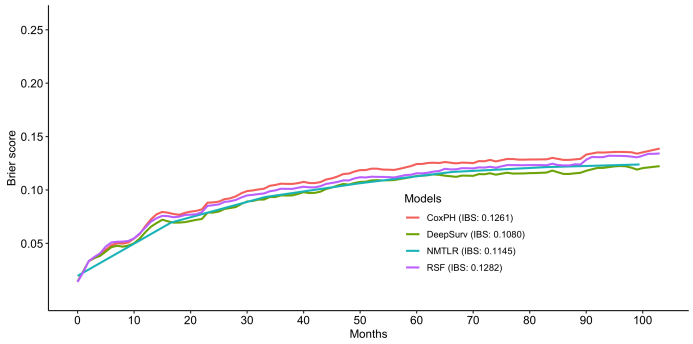


FIGURE 4
Prediction error curve. As a benchmark, a useful model will have a Brier score below 0.25.

location, tumor grade, surgery, tumor size, tumor extension, and distant metastasis were selected to incorporate in the modeling ($p<0.05$) (Table 1). The two-layer neural network DeepSurv model performed the best, followed by NMTLR, RSF and CoxPH. The C-index values for the DeepSurv model were 0.854 for the training dataset and 0.832 for the test dataset. Roc curves and calibration curves further validated DeepSurv's performance in terms of discrimination and calibration for predicting 1 -, 3 -, 5 - and 10-year survival. By combining deep learning methods to model the probabilities of events as a function of time, the DeepSurv model outperforms other models when dealing with large samples, multiple variables, and nonlinearity. The best-performing DeepSurv model was incorporated into a user-friendly web-based application that can be accessed for free at https://share.streamlit.io/whuh-ml/chondrosarcoma/Predict/app.py.

Compared to previous studies predicting chondrosarcoma survival, our study showed advantages in terms of discrimination and flexibility. Song (6) used a nomogram to fit data from chondrosarcoma patients in the SEER database prior to 2011 to predict OS, with a c-index of 0.753 for the validation set. In our study, the discrimination of the CoxPH model was slightly improved (0.773), which may be related to the fact that we included more cases and a more detailed classification of surgical procedures. The SORG algorithm proposed by Thio (10) made progress under the task of predicting 5-year survival in chondrosarcoma, with an AUC of 0.87 in the internal validation dataset. Although our DeepSurv model slightly outperformed the SORG algorithm in predicting 5-year survival (AUC of DeepSurv: 0.895), what makes our study more significant is that the influence of time on events is considered. Unlike SORG, which can only predict the binary outcome of 5-year survival, the DeepSurv model is more flexible and able to directly predict the patient's survival function, thereby obtaining the probability of survival at any point in time. In addition, the neural network embedded in the DeepSurv model has great potential to learn from high-dimensional data and can be further enhanced by fitting images and genetic data, or by using multimodal information fusion techniques.

There are several limitations to consider in our study. Firstly, with the removal of one-third of the data used for internal validation, only 2,203 pieces of data were used for model training. Since chondrosarcoma tumors are mostly early-stage tumors (distant metastasis occurred in 128 of the 2203 patients), deep learning may not fully learn the characteristics of patients with advanced tumors. The prediction error curve also shows that the prediction performance of the DeepSurv model is

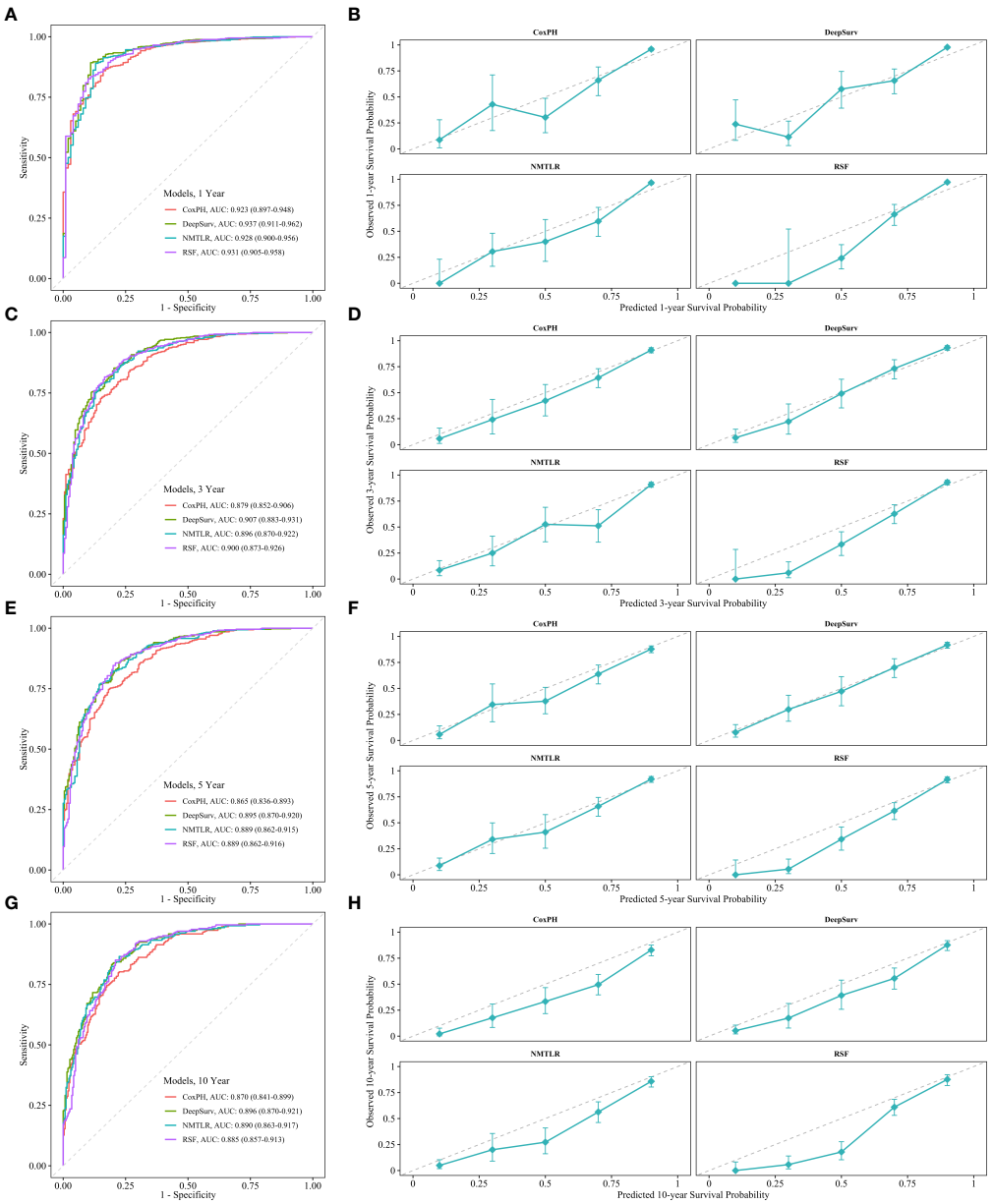


FIGURE 5
The receiver operating curves (ROC) and calibration curves for 1-, 3-, 5-, 10-year survival predictions. ROC curves for **(A)** 1-, **(C)** 3-, **(E)** 5-, **(G)** 10-year survival predictions. calibration curves for **(B)** 1-, **(D)** 3-, **(F)** 5-, **(H)** 10- year survival predictions.

significantly better than that of other models for patients with longer survival (Figures 4, 5). Secondly, since the data are from national databases, some known prognostic factors [such as pathologic fracture (6) and biomarkers (2)] were not available. Thirdly, the model in this study has not been externally validated. Although we have adopted measures such as data segmentation and cross-validation in model development, the generalization and reliability of the model need to be further validated using other data sets. Fourthly, personalized treatment recommendations are another advantage of the DeepSurv algorithm (12, 18) but were not validated in this study because of the lack of treatment data. Due to the linear fitting of variables by the classical Cox model, the model recommended a constant treatment plan for all patients according to the calculated hazard ratio (HR) value. However, DeepSurv can make personalized treatment recommendations for different patients based on the complex non-linear relationship between the variables fitted by the model (12), which is more in line with real-world rules. For example, the use of chemotherapy in patients

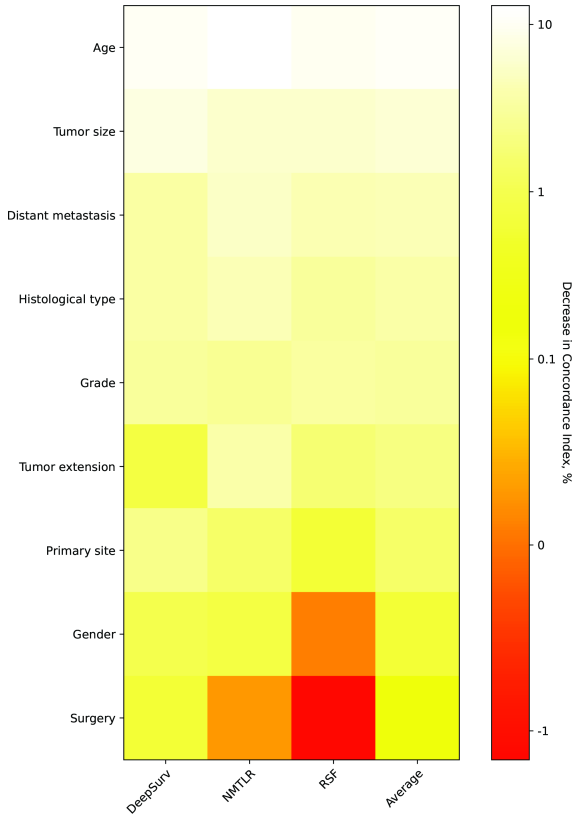


FIGURE 6
Heatmap of feature importance for DeepSurv, neural network multitask logistic regression (N-MLTR) and random survival forest (RSF) models. The values are expressed as a percentage reduction in the C-index after the value of a feature has been replaced by random numbers. Higher values suggest that a feature is more important in influencing the predictive accuracy of the corresponding deep learning model.

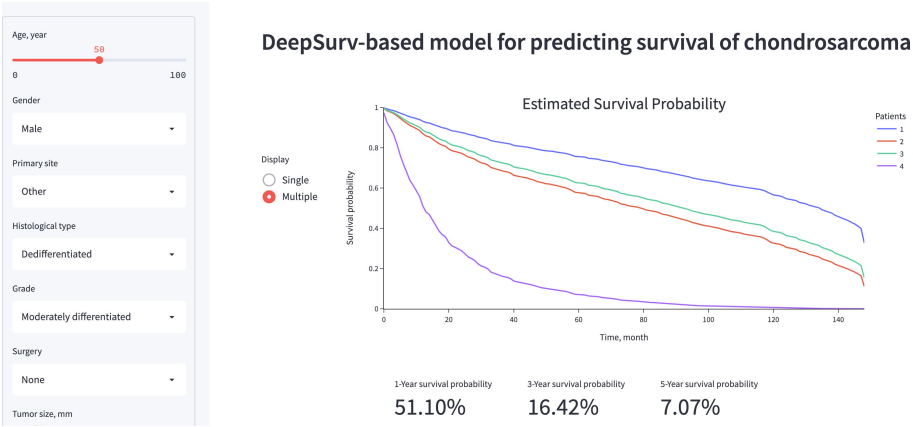


FIGURE 7
A screenshot of the online web-based application of DeepSurv model.

with chondrosarcoma is still controversial (1). By fitting the complex factors that affect the efficacy of chemotherapy, a treatment recommendation system based on deep learning may suggest the appropriate treatment for each individual.

To conclude, this study evaluated and compared the performance of two deep learning-based algorithms and two conventional methods for predicting overall survival in patients with chondrosarcoma. Overall, deep learning algorithms showed excellent discriminating capabilities, calibration, and stability in survival prediction. DeepSurv performed best in terms of discrimination and model calibration and was incorporated into a web-based application for clinical use. Further extension of the models developed in this work—considering specific aspects such as prognostic biomarkers, and image data—is necessary for future studies in order to encourage their widespread use in orthopedic oncology clinics for customized treatment planning and monitoring.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://seer.cancer.gov/.

Ethics statement

Because the SEER database is a publicly available database of de-identified patient data, no ethics committee review was required for its use in this project.

Author contributions

YW, JC, and LY contributed to the conception and design of the study. LY organized the database. LY and NG performed the statistical analysis. LY, FA, YK, and YZ wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version. The first two authors contributed equally to this work. The last two authors contributed equally to this work.

Funding

This study was supported by the National Key R&D Program of China (2020YFC2006004-05).

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.967758/full#supplementary-material

References

1. Cranmer LD, Chau B, Mantilla JG, Loggers ET, Pollack SM, Kim TS, et al. Is chemotherapy associated with improved overall survival in patients with dedifferentiated chondrosarcoma? *A SEER Database Anal Clin Orthop Relat Res* (2022) 480:748–58. doi: 10.1097/CORR.0000000000002011

2. Lyskjaer I, Davies C, Strobl AC, Hindley J, James S, Lalam RK, et al. Circulating tumour DNA is a promising biomarker for risk stratification of central chondrosarcoma with IDH1/2 and GNAS mutations. *Mol Oncol* (2021) 15:3679–90. doi: 10.1002/1878-0261.13102

3. Angelini A, Guerra G, Mavrogenis AF, Pala E, Picci P, Ruggieri P. Clinical outcome of central conventional chondrosarcoma. *J Surg Oncol* (2012) 106:929–37. doi: 10.1002/jso.23173

4. Amer KM, Munn M, Congiusta D, Abraham JA, Basu Mallick A. Survival and prognosis of chondrosarcoma subtypes: SEER database analysis. *J Orthop Res* (2020) 38:311–9. doi: 10.1002/jor.24463

5. Thio QCBS, Karhade AV, Ogink PT, Raskin KA, De Amorim Bernstein K, Lozano Calderon SA, et al. Can machine-learning techniques be used for 5-year survival prediction of patients with chondrosarcoma? *Clin Orthop Relat Res* (2018) 476:2040–8. doi: 10.1097/CORR.0000000000000433

6. Song K, Shi X, Wang H, Zou F, Lu F, Ma X, et al. Can a nomogram help to predict the overall and cancer-specific survival of patients with chondrosarcoma? *Clin Orthop Relat Res* (2018) 476:987–96. doi: 10.1007/s11999.0000000000000152

7. Dong Y, Xie L, Kang H, Peng R, Guo Q, Song K, et al. A competing risk-based prognostic model to predict cancer-specific death of patients with spinal and pelvic chondrosarcoma. *Spine (Phila Pa 1976)* (2021) 46:E1192–201. doi: 10.1097/BRS.0000000000004073

8. Wu X, Wang Y, Sun W, Tan M. Prognostic factors and a nomogram predicting overall survival in patients with limb chondrosarcomas: A population-based study. *BioMed Res Int* (2021) 2021:4510423. doi: 10.1155/2021/4510423

9. Bongers MER, Karhade AV, Setola E, Gambarotti M, Groot OQ, Erdoğan KE, et al. How does the skeletal oncology research group algorithm's prediction of 5-year survival in patients with chondrosarcoma perform on international validation? *Clin Orthop Relat Res* (2020) 478:2300–8. doi: 10.1097/CORR.0000000000001305

10. Bongers MER, Thio QCBS, Karhade AV, Stor ML, Raskin KA, Lozano Calderon SA, et al. Does the SORG algorithm predict 5-year survival in patients with chondrosarcoma? an external validation. *Clin Orthop Relat Res* (2019) 477:2296–303. doi: 10.1097/CORR.0000000000000748

11. Kvamme H, Borgan Ø., Scheel I. Time-to-event prediction with neural networks and cox regression. *arXiv* (2019) 20(129).

12. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC Med Res Method* (2018) 18:1–12. doi: 10.1186/s12874-018-0482-1

13. Lee C, Light A, Alaa A, Thurtle D, van der Schaar M, Gnanapragasam VJ. Application of a novel machine learning framework for predicting non-metastatic prostate cancer-specific mortality in men using the surveillance, epidemiology, and end results (SEER) database. *Lancet Digit Health* (2021) 3:e158–65. doi: 10.1016/S2589-7500(20)30314-9

14. She Y, Jin Z, Wu J, Deng J, Zhang L, Su H, et al. Development and validation of a deep learning model for non-small cell lung cancer survival. *JAMA Netw Open* (2020) 3:e205842. doi: 10.1001/jamanetworkopen.2020.5842

15. Yu C-N, Greiner R, Lin H-C, Baracos V. Learning patient-specific cancer survival distributions as a sequence of dependent regressors. *Adv Neural Inf Process Syst* (2011) 24:1845–53.

16. Fotso S. Deep neural networks for survival analysis based on a multi-task framework. *arXiv: Mach Learn* (2018) arXiv:1801.05512.

17. Kang L, Chen W, Petrick NA, Gallas BD. Comparing two correlated c indices with right-censored survival outcome: a one-shot nonparametric approach. *Stat Med* (2015) 34:685–703. doi: 10.1002/sim.6370

18. Howard FM, Kochanny S, Koshy M, Spiotto M, Pearson AT. Machine learning-guided adjuvant treatment of head and neck cancer. *JAMA Netw Open* (2020) 3:e2025881. doi: 10.1001/jamanetworkopen.2020.25881

19. Bruns J, Elbracht M, Niggemeyer O. Chondrosarcoma of bone: an oncological and functional follow-up study. *Ann Oncol* (2001) 12:859–64. doi: 10.1023/A:1011162118869

20. Giuffrida AY, Burgueno JE, Koniaris LG, Gutierrez JC, Duncan R, Scully SP. Chondrosarcoma in the united states (1973 to 2003): an analysis of 2890 cases from the SEER database. *JBJS* (2009) 91:1063–72. doi: 10.2106/JBJS.H.00416

21. Nota SP, Braun Y, Schwab JH, van Dijk CN, Bramer JA. The identification of prognostic factors and survival statistics of conventional central chondrosarcoma. *Sarcoma* (2015) 2015:623746. doi: 10.1155/2015/623746

22. Gitto S, Cuocolo R, Annovazzi A, Anelli V, Acquasanta M, Cincotta A, et al. CT radiomics-based machine learning classification of atypical cartilaginous tumours and appendicular chondrosarcomas. *EBioMedicine* (2021) 68:103407. doi: 10.1016/j.ebiom.2021.103407

23. Gitto S, Cuocolo R, van Langevelde K, van de Sande MAJ, Parafioriti A, Luzzati A, et al. MRI Radiomics-based machine learning classification of atypical cartilaginous tumour and grade II chondrosarcoma of long bones. *EBioMedicine* (2022) 75:103757. doi: 10.1016/j.ebiom.2021.103757

24. Righi A, Pacheco M, Cocchi S, Asioli S, Gambarotti M, Donati DM, et al. Secondary peripheral chondrosarcoma arising in solitary osteochondroma: variables influencing prognosis and survival. *Orphanet J Rare Dis* (2022) 17:74. doi: 10.1186/s13023-022-02210-2

25. Ivanics T, Nelson W, Patel MS, Claasen M, Lau L, Gorgen A, et al. The Toronto postliver transplantation hepatocellular carcinoma recurrence calculator: A machine learning approach. *Liver Transpl* (2022) 28:593–602. doi: 10.1002/lt.26332

26. Hadanny A, Shouval R, Wu J, Gale CP, Unger R, Zahger D, et al. Machine learning-based prediction of 1-year mortality for acute coronary syndrome(☆). *J Cardiol* (2022) 79:342–51. doi: 10.1016/j.jjcc.2021.11.006

27. Kim B, Jang YJ, Cho HR, Kim SY, Jeong JE, Shim MK, et al. Predicting completion of clinical trials in pregnant women: Cox proportional hazard and neural network models. *Clin Transl Sci* (2022) 15:691–9. doi: 10.1111/cts.13187

28. Zhong LZ, Fang XL, Dong D, Peng H, Fang MJ, Huang CL, et al. A deep learning MR-based radiomic nomogram may predict survival for nasopharyngeal carcinoma patients with stage T3N1M0. *Radiother Oncol* (2020) 151:1–9. doi: 10.1016/j.radonc.2020.06.050

29. Han W, Qin L, Bay C, Chen X, Yu KH, Miskin N, et al. Deep transfer learning and radiomics feature prediction of survival of patients with high-grade gliomas. *AJNR Am J Neuroradiol* (2020) 41:40–8. doi: 10.3174/ajnr.A6365

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 14 October 2022
DOI 10.3389/fonc.2022.997911

OPEN ACCESS

EDITED BY
Hani J Marcus,
University College London,
United Kingdom

REVIEWED BY
Enrico Checcucci,
IRCCS Candiolo Cancer Institute, Italy
Xu Dongliang,
Shanghai University of Traditional Chinese Medicine, China

*CORRESPONDENCE
Pengfei Shao
spf_urology@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Surgical Oncology, a section of the journal Frontiers in Oncology

RECEIVED 19 July 2022
ACCEPTED 28 September 2022
PUBLISHED 14 October 2022

CITATION
Zhang S, Yang G, Qian J, Zhu X, Li J, Li P, He Y, Xu Y, Shao P and Wang Z (2022) A novel 3D deep learning model to automatically demonstrate renal artery segmentation and its validation in nephron-sparing surgery. *Front. Oncol.* 12:997911. doi: 10.3389/fonc.2022.997911

A novel 3D deep learning model to automatically demonstrate renal artery segmentation and its validation in nephron-sparing surgery

Shaobo Zhang[1†], Guanyu Yang[2†], Jian Qian[1†], Xiaomei Zhu[3], Jie Li[1], Pu Li[1], Yuting He[2], Yi Xu[3], Pengfei Shao[1*] and Zengjun Wang[1]

[1]Department of Urology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China, [2]Key Laboratory of Computer Network and Information Integration, Southeast University, Ministry of Education, Nanjing, China, [3]Department of Radiology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

Purpose: Nephron-sparing surgery (NSS) is a mainstream treatment for localized renal tumors. Segmental renal artery clamping (SRAC) is commonly used in NSS. Automatic and precise segmentations of renal artery trees are required to improve the workflow of SRAC in NSS. In this study, we developed a tridimensional kidney perfusion (TKP) model based on deep learning technique to automatically demonstrate renal artery segmentation, and verified the precision and feasibility during laparoscopic partial nephrectomy (PN).

Methods: The TKP model was established based on convolutional neural network (CNN), and the precision was validated in porcine models. From April 2018 to January 2020, TKP model was applied in laparoscopic PN in 131 patients with T1a tumors. Demographics, perioperative variables, and data from the TKP models were assessed. Indocyanine green (ICG) with near-infrared fluorescence (NIRF) imaging was applied after clamping and dice coefficient was used to evaluate the precision of the model.

Results: The precision of the TKP model was validated in porcine models with the mean dice coefficient of 0.82. Laparoscopic PN was successfully performed in all cases with segmental renal artery clamping (SRAC) under TKP model's guidance. The mean operation time was 100.8 min; the median estimated blood loss was 110 ml. The ischemic regions recorded in NIRF imaging were highly consistent with the perfusion regions in the TKP models (mean dice coefficient = 0.81). Multivariate analysis revealed that the feeding lobar artery number was strongly correlated with tumor size and contact surface area; the supplying segmental arteries number correlated with tumor size.

Conclusions: Using the CNN technique, the TKP model is developed to automatically present the renal artery trees and precisely delineate the perfusion regions of different segmental arteries. The guidance of the TKP model is feasible and effective in nephron-sparing surgery.

KEYWORDS

tridimensional kidney perfusion model, automatic segmentation, deep learning technique, convolutional neural network, nephron-sparing surgery

Introduction

As a minimally invasive nephron-sparing surgery, laparoscopic partial nephrectomy (laparoscopic PN) is a mainstream treatment for cT1a renal tumors (1). In laparoscopic PN, renal artery clamping induces warm ischemic injury (WII) (2), which could be minimized by segmental renal artery clamping (SRAC) technique, converting global parenchymal ischemia to regional ischemia (3–6). To implement the SRAC technique, dual-source computed tomography (DSCT) angiography was applied to reveal a high-quality three-dimensional vasculature model of the renal hilum, and identify the target segmental arteries needed to be clamped if their branches enter or abut the tumor (6, 7). However, DSCT angiography is unable to provide the details of the perfusion regions of different segmental arteries, and the determination of the target arteries is inaccurate occasionally, which could lead to insufficient clamping and result in arterial bleeding (5). Therefore, a more precise clamping strategy is required.

Based on the contrast CT scan, organ segmentation with different kinds of statistical models were reported in several abdominal organs over the years (8–10). Previously, in our center, to meet the requirement of a more precise SRAC technique, a novel functional three-dimensional perfusion model was established to determine the target arteries by depicting the parenchymal perfusion regions of different segmental arteries using the semi-automatic segmentation of the kidney and renal arteries (11). Recently, with the development of medical image processing technology, convolutional neural network (CNN) as a kind of deep learning technique has gradually begun to be applied in the segmentation of organs and vasculature (12). Based on the CNN, we previously created a series of novel methods, which could provide a fully automatic segmentation of kidney, tumor, and renal artery trees (13, 14). In present study, integrating CNN technique and the distance transformation algorithm, a novel three-dimensional perfusion model was established, which was called the tridimensional kidney perfusion (TKP) model.

Materials and methods

The establishment procedure of the TKP model was presented. The accuracy of the TKP model was verified in porcine models, and the feasibility and efficacy of this model were assessed in patients undergoing laparoscopic PN with SRAC.

Establishment of the TKP model

Automatic segmentation of the kidneys and tumors

Our homemade three-dimensional fully-supervised convolutional neural (FCN) network with a pyramid-pooling module (PPM) (3D_FCN_PPM) was proposed for segmentation of kidneys and tumors previously (13) (Figure 1). During the establishment of the 3D_FCN_PPM network, the abdominal CT images of 140 patients were recruited from the department of radiology after the informed consent was obtained. And the images were obtained and analyzed in Dicom format. Ninety images were used for the training set, and the remaining 50 images were used for testing. The 3D_FCN_PPM network was demonstrated to be efficient and precise in segmentation with the dice coefficient equal to 0.931 for kidney and 0.802 for renal tumors.

Automatic segmentation of the renal artery trees

Integrating the technologies of deep-priori anatomy (DPA), dense-biased network (DenseBiasNet), and hard-region adaptation loss (HRA loss), we proposed a fine three-dimensional renal artery segmentation framework, called DPA-DenseBiasNet framework (14). The DPA-DenseBiasNet framework was based on a two-stage CNN, including (1) autoencoder (AE) network pre-training and (2) DPA features embedding and DenseBiasNet training. AE is an unsupervised neural network, which can extract anatomical features (15). In this framework, AE is applied to acquire the representation

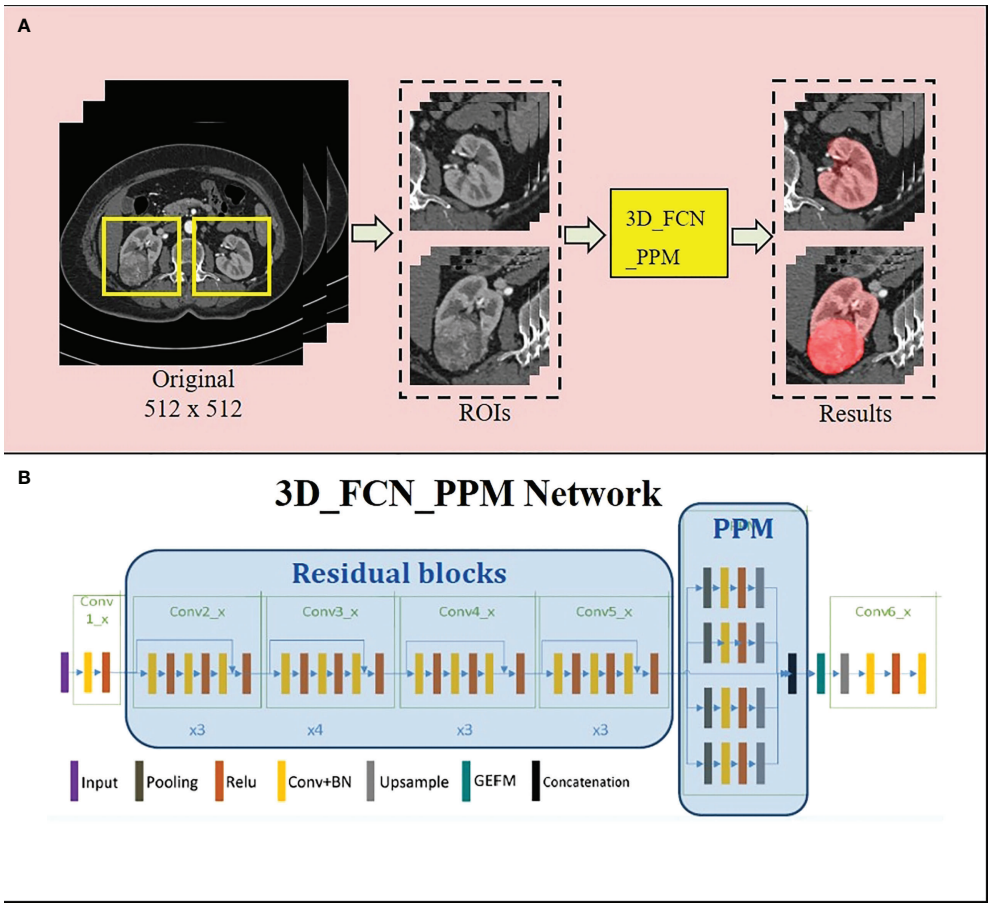


FIGURE 1
The 3D_FCN_PPM network is applied in the automatic segmentation of kidney and tumor. **(A)** the pipeline of kidney and tumor segmentation; **(B)** the architecture of the 3D_FCN_PPM network.

ability of anatomical features (DPA features) through a big unlabeled dataset. In stage 2, extracted DPA features from AE are embedded in the DenseBiasNet system, forming the priori anatomy information, which can adapt anatomical variations. Finally, modified by HRA loss function, a precise tridimensional renal artery segmentation is achieved (Figure 2). During the establishment of the DPA-DenseBiasNet framework, a total of 196 patients with 392 kidney images were recruited. Fifty-two labeled images and 236 unlabeled images were used for training, and 104 labeled images were used for testing. The DPA-DenseBiasNet was demonstrated to have high predictive accuracy in renal artery segmentation with a mean dice coefficient of 0.884.

Estimation of the arterial perfusion regions on the renal parenchyma

After automatic segmentation, the estimation procedure based on the two-step algorithm in C++ programming: (1) set the lobar arteries and their branches to the same category and marked with the same color if they branch out from the same segmental artery (Figure 3A, the arteries with the same color are the same segmental artery subtree); (2) the distance transformation algorithm is used to find the closest lobar arteries or their branches for every voxel point in the renal parenchyma as its blood supply vessel, and the color of this point is marked. All voxel points in the renal parenchyma are categorized according to their colors, and the perfusion regions of different segmental arteries are then depicted (Figure 3B).

By the automatic segmentation and the perfusion region estimation algorithm, the TKP model is finally established.

Validation in animal models

Subjects preparation

The validation procedure in swine was approved by the Animal Use and Management Ethics Committee of Nanjing

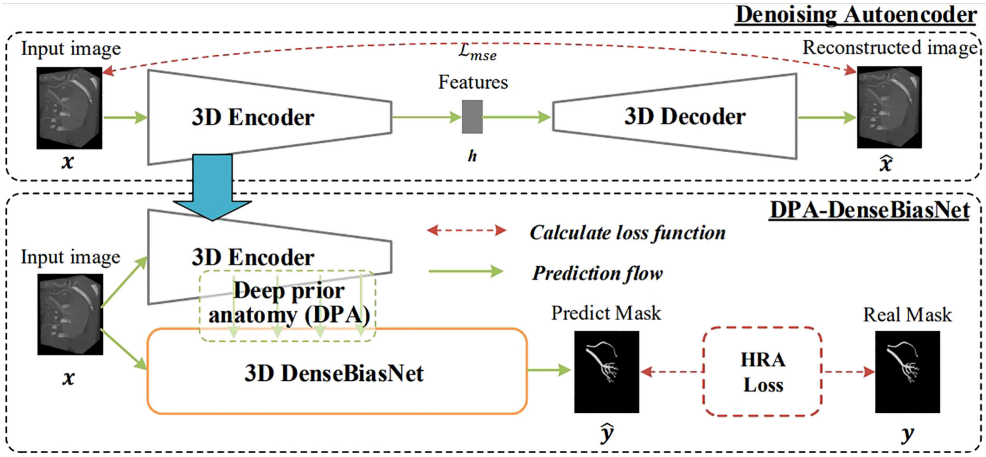


FIGURE 2
The 3D DPA-DensebiasNet framework is applied in the automatic segmentation of renal artery trees. The procedure includes two stages. Stage 1(the upper dotted box): is AE pre-training. The AE is trained by a lot of unlabeled images and DPA features are extracted. Stage 2 (the lower dotted box): extracted DPA features from AE are embedded in the DenseBiasNet system, forming the priori anatomy information, which can adapt anatomical variations. And finally, modified by HRA loss function, a precise tridimensional renal artery segmentation is achieved.

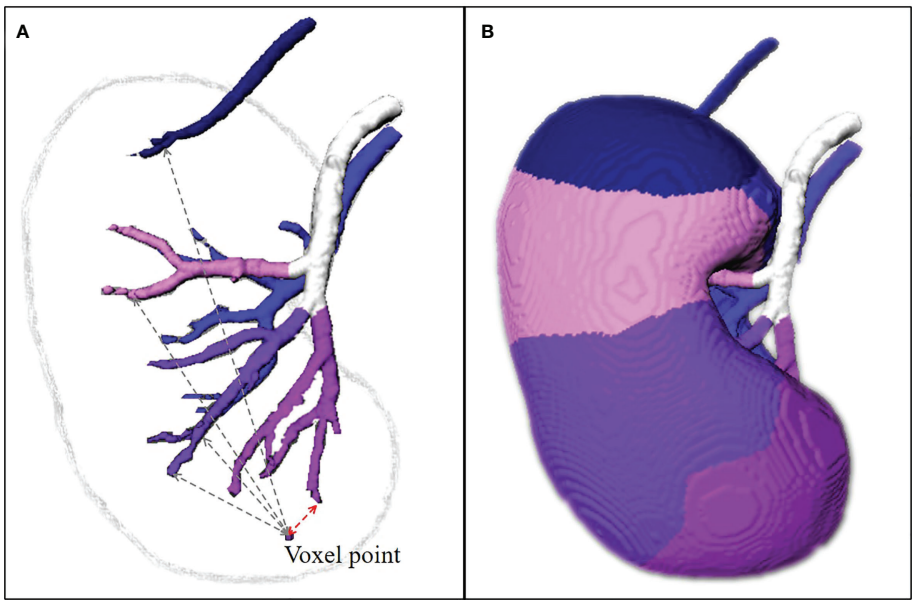


FIGURE 3
The estimation of perfusion regions. **(A)** The lobar arteries and their branches are extracted, set to the same category and marked with the same color if they branch out from the same segmental arteries. The distance transformation algorithm is used to find the closest lobar arteries or their branches for every voxel point in the renal parenchyma as its blood supply vessel, and the color of this point is marked. **(B)** All voxel points in the renal parenchyma are categorized according to their colors, and the perfusion regions of different segmental arteries are then depicted. The TKP model is finally established.

TABLE 1 Patient characteristics and the TKP model information.

Variables Patient characteristics	
Patients, no.	131
Age, yr, mean±SD	56.3±11.4
Male, no. (%)	97 (74.0)
BMI, kg/m^2, mean±SD	25.1±3.3
Hypertension, no. (%)	53 (40.5)
Diabetes mellitus, no. (%)	24 (18.3)
The TKP model information	
R.E.N.A.L score, mean±SD	6.4±1.4
Radius of tumor (maximal diameter), cm, mean±SD	2.5±0.8
Exophytic/endophytic properties, no. (%)	
≥50%	68 (51.9)
<50%	53 (40.5)
Entirely endophytic	10 (7.6)
Location relative to the polar line, no. (%)	
Entirely upper or lower polar	61 (46.6)
Lesion crosses polar line	45 (34.4)
Middle polar (>50% crosses polar line)	25 (19.1)
Nearness to UCS/sinus, no. (%)	
≥7mm	25 (19.1)
<7mm and >4mm	64 (48.9)
≤4mm	42 (32.1)
Contact surface area (CSA), cm^2, mean±SD	13.5±11.3
Feeding lobar artery number, no. (%)	
1	21 (16.0)
2	56 (42.7)
3	43 (32.8)
4	10 (7.7)
5	1 (0.8)
Target segmental artery number, no. (%)	
1	79 (60.3)
2	49 (37.4)
3	3 (2.3)

TKP, tridimensional kidney perfusion; BMI, body mass index; SD, standard deviation; UCS, urinary collecting system.

Medical University. Six domestic female pigs with 11 kidneys were recruited and the median weight was 39.5 kg (Table 1). All swine were intramuscularly injected with xylazine (1.5 mg/kg), atropine (0.02 mg/kg) and diazepam (10 mg) for initial anesthesia and intravenously injected with propofol (25 μg/kg/min) for maintaining anesthesia. Tracheal intubation was conducted and the right femoral artery was punctured.

Validation procedure

All subjects underwent contrast CT scan to establish the TKP models (Figures 4A–D). One candidate segmental artery in each kidney was selected and ligated with a double-strand 1/0 suture during open surgery. The ischemia region was revealed and recorded (Figures 4E, F). To evaluate the accuracy of the TKP model, a second contrast CT scan was performed after open surgery to present the actual ischemia region (Figure 4H). Using the method of dice coefficient, the actual ischemia region from the second contrast CT scan was compared with the perfusion region predicted by the TKP model (Figure 4G).

Clinical application in laparoscopic PN

Patients preparation

Laparoscopic PN with SRAC under the TKP model's guidance was performed in 131 patients from April 2018 to January 2020. All the recruited patients signed a written informed consent form approved by the institutional review board of Nanjing Medical University. Inclusion criteria were: 1) a single localized mass ≤4 cm (clinical T1a); 2) normal renal function (defined as creatinine clearance rate (CCR) ≥90 ml/min). All patients underwent a contrast CT scan to establish the TKP model before operation. The R.E.N.A.L scores were applied to estimate the complexity of tumors, including Radius (maximum tumor diameter), Exophytic/Endophytic, Nearness of the tumor to the collecting system, Anterior/Posterior and Location relative to the polar lines (16, 17). The contact surface area (CSA) of tumors, as another index predicting the tumor complexity (18), could be calculated by area element algorithm according to the model. Additionally, the numbers of target segmental arteries and feeding lobar arteries of tumors could be provided from the model.

Precise determination of the target segmental arteries

In the TKP models, the tumors, segmental renal arteries and their corresponding perfusion regions were automatically presented. The target segmental arteries supplying tumors were determined by the perfusion regions wherein the renal tumors were confined (Figures 5A, G).

Surgical procedure and follow-up

All surgical procedure were conducted by the same surgeon (Pengfei Shao). During laparoscopic PN, target segmental arteries determined by TKP models were clamped by bulldog clamps. Immediately after clamping, 5 mg indocyanine green (ICG) was intravenously injected and the absence of perfusion on renal parenchyma was presented in near-infrared fluorescence (NIRF) imaging (Figures 5D, J). The similarity was evaluated between the absence of perfusion in NIRF imaging and the predicted perfusion region in TKP model by the method of dice coefficient. Tumor resection was conducted and the parenchymal defect was closed. Finally, the mass was retrieved to receive a pathological examination.

The follow-up period was defined as the duration from the date of operation to the date of the most recent examination. For

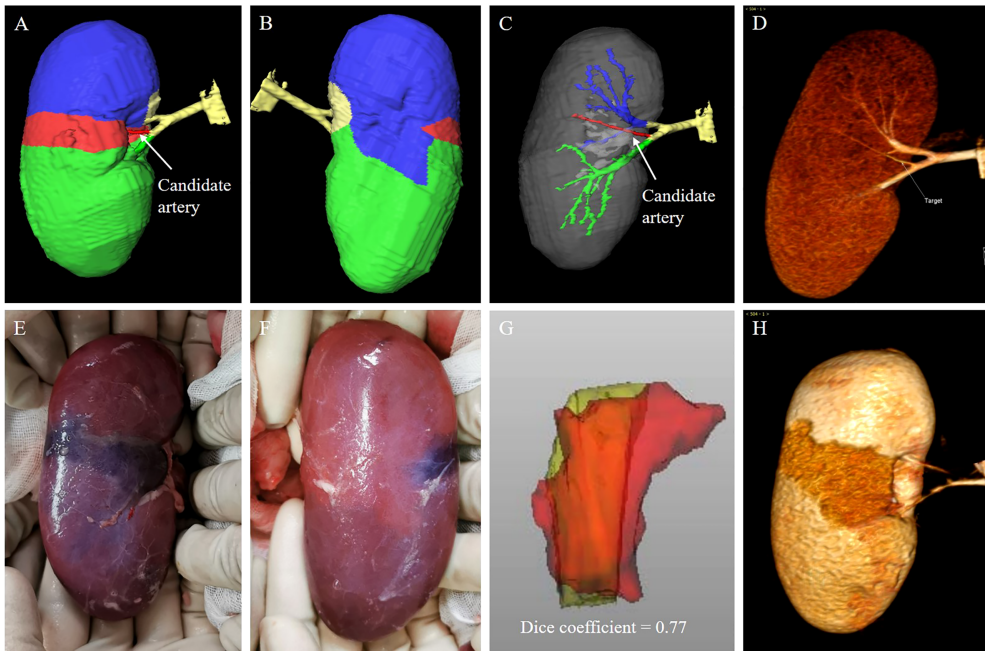


FIGURE 4
animal validation of the TKP model. **(A–D)** The TKP model of a porcine kidney is automatically established based on the first CT scan, and a candidate segmental artery is selected. **(E, F)** The ischemic line on renal parenchyma is visible and recorded after the candidate segmental artery is ligated. **(G, H)** The second CT scan is performed and the similarity between the actual ischemia region from the second CT scan and the perfusion region from the TKP model is calculated using the algorithm of dice coefficient. In this case, the dice coefficient is 0.77.

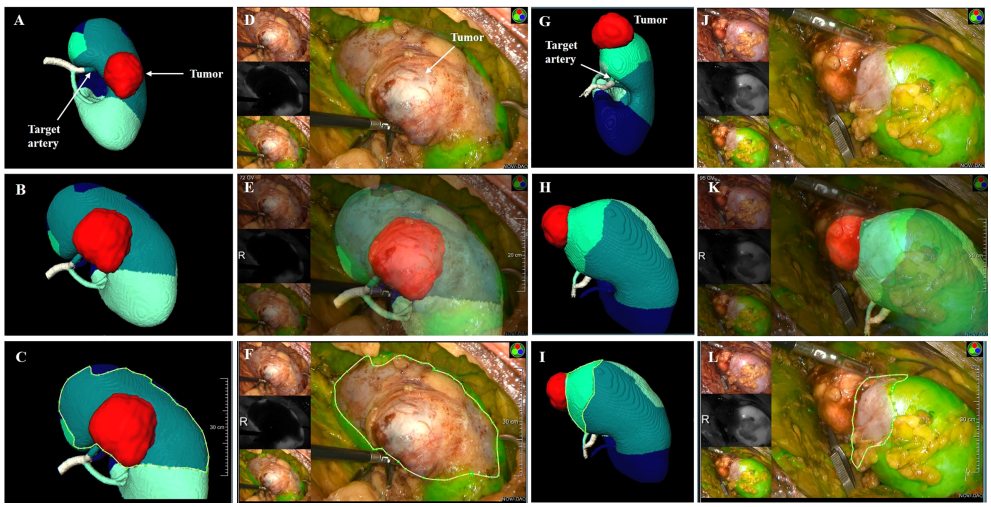


FIGURE 5
the clinical application of TKP model. **(A–F)** Case 1: a male patient with 3.8 cm tumor on the posterior part of the right kidney. The R.E.N.A.L score is 9. **(G–L)** Case 2: a female patient with 3.1 cm tumor on the upper polar of the right kidney. The R.E.N.A.L score is 5. **(A, G)** The TKP model is automatically established based on contrast CT scan and the target segmental artery is determined. **(D–F, J–L)** ICG is injected immediately after clamping, and the absence of perfusion on the renal parenchyma was confirmed with NIRF imaging. (**C** vs. **F**, **I** vs. **L**) The algorithm of dice coefficient is applied in assessing the similarity, and the dice coefficient is 0.92 and 0.81, which indicates that the ischemic region recorded in NIRF imaging is highly consistent with the perfusion region predicted in the TKP model.

follow-up and surveillance, abdominal plain CT scan was performed at 3 and 6 months and every 6 months. Chest CT scan and abdominal contrast CT scans were performed every 6 and 12 months, respectively.

Statistical analysis

Categorical variables were presented as frequencies and percentages. Continuous variables were reported as mean ± SD (normal distribution) or medians and ranges (abnormal distribution). Complications were analyzed according to the Clavien–Dindo system (19). Logistic regression analysis was used to test the correlation of tumor characteristics and the number of arteries supplying the tumor. All statistical analyses were conducted using IBM SPSS v.22 (SPSS Inc., Chicago, IL, USA), and two-sided $p < 0.05$ was considered to be statistically significant. Dice coefficient was applied to evaluate the similarity of regions or volumes in two images, and high similarity was defined as dice coefficient > 0.7.

Results

Of 11 porcine kidneys, the median number of segmental arteries was 3. After clamping, the ischemic regions were located on the upper, middle, and lower polar in three, one, and seven kidneys, respectively (Table 1). As shown in Figure 4, the boundaries of the perfusion regions in the TKP models were consistent with the ischemic lines recorded intraoperatively. It was demonstrated to have high similarity between the actual ischemia region from the post-operative CT scan and the perfusion region from the TKP model (dice coefficient = 0.82) (Table 1).

In clinical procedure, basic characteristics are shown in Table 2. There were 97 males and 34 females, aged 56.3 ± 11.4 years, with a mean body mass index of 25.1 kg/m^2. The mean tumor size (radius of the tumor as maximal diameter) was 2.5 cm, and the R.E.N.A.L score was 6.4 ± 1.4. According to the TKP model, the CSA was 13.5 ± 11.3 cm^2. There were 79, 49, and 3 patients with tumors supplied by one, two, and three target segmental arteries, respectively. Furthermore, subclassified by the numbers of feeding lobar arteries, there were 21, 56, 43, 10, and 1 patients with tumors supplied by one, two, three, four and five lobar arteries, respectively.

Laparoscopic PN with precise SRAC was successfully performed under the TKP models' guidance in all patients. The mean operation time was 100.8 min, with a mean warm ischemic time (WIT) of 27.0 min. The median estimated blood loss (EBL) was 110 ml (40 - 400 ml). There were no patients converting to main renal artery clamping, radical nephrectomy, or open surgery. No arterial bleeding or uncontrolled hemorrhage from the tumor bed occurred during tumor resection (Table 3). The ischemic regions recorded by ICG with NIRF imaging were highly consistent with the perfusion regions in the TKP models, with the dice coefficient of 0.81 (0.72–0.94) (Figure 5 and Table 3). The median length of stay after the operation was 7 days. Postoperative complications occurred in eight (6.1%) patients, including five patients with grade 1 complication (hematuria not requiring intervention), two patients with grade 2 complication (hematuria requiring blood transfusion), and one patient with grade 3a complication (bleeding requiring embolization intervention under local anesthesia). Pathology and follow-up results were also revealed in Table 3. At a median follow-up of 21 months, no patient had tumor recurrence or metastasis.

According to the number of tumor supplying arteries, subjects were divided into subgroups with one and two or more supplying segmental arteries or subgroups with 1–2 and 3–5 supplying lobar arteries. The comparisons of tumor characteristics between these subgroups are conducted in Table 4. Furthermore, logistic regression analysis is applied and presented in Table 5. On multivariate analysis, the numbers of both supplying segmental and lobar arteries strongly correlated with tumor size (OR = 5.92, $p = 0.000$ for segmental arteries and OR = 4.84, $p = 0.002$ for lobar arteries). The larger the tumor size, the more the number of supplying segmental and lobar arteries. The CSA of the tumor had correlation with the number of supplying lobar arteries (OR = 1.11, $p = 0.014$), instead of segmental arteries ($p = 0.815$). None

TABLE 2 Perioperative outcomes and follow-up.

Variables	
Dice coefficient (NIRF imaging vs. TKP model)	0.81 (0.72-0.94)
Operation time, min, mean±SD	100.8±11.2
Warm ischemic time, min, mean±SD	27.0±5.2
EBL, ml, median (range)	110 (40-400)
LOS after operation, days, median (range)	7 (3-17)
Post-operative complications, no. (%)	8 (6.1)
Grade 1 (hematuria not requiring intervention)	5 (3.8)
Grade 2 (hematuria requiring blood transfusion)	2 (1.5)
Grade 3a (bleeding requiring embolization intervention under local anaesthesia)	1 (0.8)
Pathology, no. (%)	
Clear cell carcinoma	112 (85.5)
Perivascular epithelioid cell tumor	7 (5.3)
Papillary renal cell carcinoma	6 (4.6)
Oxyphilic adenoma	3 (2.3)
Chromophobe renal cell carcinoma	3 (2.3)
Follow-up, mo, median (range)	21 (13-33)
Tumor recurrence and metastasis	0

NIRF, near-infrared fluorescence; TKP model, tridimensional kidney perfusion model; SD, standard deviation; EBL, estimated blood loss; LOS, length of stay.

TABLE 3 Validation results of TKP model in porcine models.

Variables	
Subjects/kidneys, no.	6/11
Weight, kg, median (range)	39.5 (28-42)
Number of segmental arteries per kidney, median (range)	3 (2-4)
Location of the perfusion regions of the candidate arteries, no. (%)	
Upper polar	3 (27.3)
Middle polar	1 (9.1)
Lower polar	7 (63.6)
Dice coefficient (TKP model vs. postoperative CT scan)	**0.82 (0.63-0.88)**

TKP model, tridimensional kidney perfusion model.
Bold value, dice coefficient showed high similarity.

of the correlations was found between the other tumor characteristics and the number of supplying arteries.

Discussion

Traditionally, DSCT angiography was utilized to reveal the three-dimensional hilar vasculature during SRAC procedure (5). In DSCT angiography, the target arteries were determined manually, leading to the underestimation of artery branches feeding both tumor and the surrounding normal tissue (7). So measurement bias, insufficient clamping, arterial bleeding or even converting to the main artery clamping during resection might occur. For a more efficient and precise SRAC technique, the TKP model was established using a homemade CNN technology, becoming an automatic tool in the surgical strategy-making of the SRAC during nephron-sparing surgery (14). ICG with NIRF imaging, as an empirical technique of fluorescence guidance (20), was introduced in this study to delineate the real ischemic area after clamping. Our results confirmed that the perfusion regions predicted in the TKP model were highly consistent with the real ischemic area in NIRF imaging during operation. In this study, under the guidance of the TKP models, all surgical procedures were performed successfully, and there was no occurrence of uncontrolled bleeding during tumor resection.

Recently, by constructing various models, researchers have been exploring the navigation technique in PN with SRAC. Ukimura et al. established a 3D model that could present opaque tumors and renal arterial trees by making the renal parenchyma semitransparent (21). By manually segmenting kidney shape, vasculature, collecting system, and tumor, Porpiglia proposed a hyperaccuracy 3D model (22), which used an augmented reality

TABLE 4 Relationship between the tumor characteristics and the number of feeding arteries.

Variables	Target segmental artery, no.		P	Feeding lobar artery, no.		P
	1	2-3		1-2	3-5	
Patients no.	79	52		77	54	
R.E.N.A.L score	6.3 ± 1.4	6.6 ± 1.3	0.270	6.0 ± 1.3	7.0 ± 1.3	**0.000**
Radius, cm, mean ± SD	2.2 ± 0.7	3.1 ± 0.7	**0.000**	2.1 ± 0.7	3.2 ± 0.6	**0.000**
Growth pattern, no. (%)			0.212			**0.035**
Exophytic	43 (54.4)	25 (48.1)		46 (59.7)	22 (40.7)	
Mesophytic	28 (35.4)	25 (48.1)		24 (31.2)	29 (53.7)	
Endophytic	8 (10.1)	2 (3.8)		7 (9.1)	3 (5.6)	
Nearness to UCS/sinus, no. (%)			0.194			**0.001**
≥7 mm	19 (24.1)	6 (11.5)		22 (28.6)	3 (5.6)	
<7 mm and >4 mm	37 (46.8)	27 (51.9)		37 (58.1)	27 (50.0)	
≤4 mm	23 (29.1)	19 (36.5)		18 (23.4)	24 (44.4)	
Location of tumor, no. (%)			0.862			**0.019**
Entirely at the polar	38 (48.1)	23 (44.2)		42 (54.5)	19 (35.2)	
Mostly at the polar	27 (34.2)	18 (34.6)		26 (33.8)	19 (35.2)	
Mostly between polar lines	14 (17.7)	11 (21.2)		9 (11.7)	16 (29.6)	
CSA, cm^2, mean ± SD	9.2 ± 8.2	20.0 ± 12.4	**0.000**	7.6 ± 5.8	22.0 ± 12.0	**0.000**

SD, standard deviation; UCS, urinary collecting system; CSA, contact surface area.
Bold value, p value < 0.05.

TABLE 5 Logistic regression analysis on the numbers of target segmental arteries and feeding lobar arteries.

Variables	Univariate analysis			Multivariate analysis		
	OR	95% CI	P	OR	95% CI	P
TSA no. (2-3 vs 1)						
Radius	6.43	3.32-12.43	**0.000**	5.92	2.29-15.30	**0.000**
CSA	1.10	1.06-1.15	**0.000**	1.01	0.95-1.07	0.815
FLA no. (3-5 vs 1-2)						
Radius	11.88	5.26-26.85	**0.000**	4.84	1.76-13.29	**0.002**
Growth pattern			**0.037**			0.326
Exophytic	ref					
Mesophytic	2.53	1.20-5.31	**0.014**			
Endophytic	0.90	0.21-3.80	0.882			
Nearness to UCS/sinus			**0.004**			0.784
≥7 mm	ref					
<7 mm and >4 mm	5.35	1.45-19.72	**0.012**			
≤4 mm	9.78	2.53-37.80	**0.001**			
Location of tumor			**0.023**			0.318
Entirely at the polar	Ref					
Mostly at the polar	1.62	0.72-3.60	0.241			
Mostly between polar lines	3.93	1.48-10.47	**0.006**			
R.E.N.A.L	1.76	1.31-2.36	**0.000**			0.495
CSA	1.20	1.12-1.29	**0.000**	1.11	1.02-1.20	**0.014**

TSA, target segmental artery; FLA, feeding lobar artery; CSA, contact surface area; UCS, urinary collecting system; OR, odd ratio; CI, confidence interval. Bold value, p value < 0.05.

(AR) technique to guide surgeons during operation (23). Additionally, the goal of purely automatic segmentation of different organs and even renal artery trees using CNNs was achieved (24–26). We had developed a series of novel CNNs, including 3D_FCN_PPM and DPA-DenseBiasNet, providing a precise segmentation of kidney, tumors, renal arteries and their branches (distal to interlobar arteries) (13, 14). These homemade CNNs, along with the distance transformation algorithm, made the establishment of the TKP model fully automated. And the automatic procedure could reduce the manual workload of delineation in the radiological process and significantly improve the efficiency of the preoperative plan of laparoscopic PN with a precise SRAC technique.

Using CNN techniques, automatic segmentation could be accurate to distal interlobar arteries, facilitating a more precise estimation of the arterial perfusion regions. In the future, in combination with the AR technique, it is expected that the TKP model could be implanted into the surgery console and become visual and synchronous. It is beneficial for improving the accuracy and efficacy of SRAC technique and tumor resection.

Leslie and his colleagues presented the CSA as a novel parameter to predict the complexity of renal tumors (18). In our study, tumor size and CSA strongly correlated with the number of feeding arteries. Larger CSA was accompanied by more feeding lobar arteries, instead of the target segmental arteries. In the future, the number of lobar arteries feeding the tumor is expected to predict renal tumor complexity and become an indicator in the scoring system to evaluate the difficulty of PN.

This study is not devoid of limitations. Firstly, we still lack a multi-center research. Secondly, we lack a randomized controlled study to compare the TKP model and the traditional DSCT angiography model since the former is a newly developed technology.

Notwithstanding these limitations, the TKP model was more than satisfactory because of the hyperaccuracy verified during operation. It is expected to become a comprehensive tool with multiple functions, such as preoperative assessment of tumor complexity, automatic planning of surgical strategy and real-time navigation of selective clamping and tumor resection.

Conclusions

Using the CNN technique, the TKP model is developed to automatically present the renal artery trees and precisely delineate the perfusion regions of different segmental arteries. The guidance of the TKP model is feasible and effective in nephron-sparing surgery.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the institutional review board of Nanjing Medical University. The patients/participants provided their written informed consent to participate in this study.The animal study was reviewed and approved by the Animal Use and Management Ethics Committee of Nanjing Medical University. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

SZ, GY, PS and ZW contributed to the study conception and design. The acquisition of data was performed by SZ, XZ, YH and YX. Data analysis and interpretation: GY, XZ, YH, YX. The first draft of the manuscript was written by SZ, JQ, YH and PS, and critical revision of the manuscript was performed by SZ, JL, PL and PS. Supervision of whole procedure was performed by JL, PS and ZW. All authors read and approved the final manuscript.

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Van Poppel H, Becker F, Cadeddu JA, Gill IS, Janetschek G, Jewett MA, et al. Treatment of localised renal cell carcinoma. *Eur Urol* (2011) 60(4):662–72. doi: 10.1016/j.eururo.2011.06.040

2. Porpiglia F, Fiori C, Bertolo R, Morra I, Russo R, Piccoli G, et al. Long-term functional evaluation of the treated kidney in a prospective series of patients who underwent laparoscopic partial nephrectomy for small renal tumors. *Eur Urol* (2012) 62(1):130–5. doi: 10.1016/j.eururo.2012.02.001

3. Li P, Qin C, Cao Q, Li J, Lv Q, Meng X, et al. A retrospective analysis of laparoscopic partial nephrectomy with segmental renal artery clamping and factors that predict postoperative renal function. *BJU Int* (2016) 118(4):610–7. doi: 10.1111/bju.13541

4. Shao P, Qin C, Yin C, Meng X, Ju X, Li J, et al. Laparoscopic partial nephrectomy with segmental renal artery clamping: technique and clinical outcomes. *Eur Urol* (2011) 59(5):849–55. doi: 10.1016/j.eururo.2010.11.037

5. Shao P, Tang L, Li P, Xu Y, Qin C, Cao Q, et al. Precise segmental renal artery clamping under the guidance of dual-source computed tomography angiography during laparoscopic partial nephrectomy. *Eur Urol* (2012) 62(6):1001–8. doi: 10.1016/j.eururo.2012.05.056

6. Shao P, Tang L, Li P, Xu Y, Qin C, Cao Q, et al. Application of a vasculature model and standardization of the renal hilar approach in laparoscopic partial nephrectomy for precise segmental artery clamping. *Eur Urol* (2013) 63(6):1072–81. doi: 10.1016/j.eururo.2012.10.017

7. Xu Y, Shao P, Zhu X, Lv Q, Liu W, Xu H, et al. Three-dimensional renal CT angiography for guiding segmental renal artery clamping during laparoscopic partial nephrectomy. *Clin Radiol* (2013) 68(11):e609–16. doi: 10.1016/j.crad.2013.06.002

8. Okada T, Linguraru MG, Hori M, Summers RM, Tomiyama N, Sato Y. Abdominal multi-organ segmentation from CT images using conditional shape-location and unsupervised intensity priors. *Med Image Anal* (2015) 26(1):1–18. doi: 10.1016/j.media.2015.06.009

9. Freiman M, Kronman A, Esses SJ, Joskowicz L, Sosna J. Non-parametric iterative model constraint graph min-cut for automatic kidney segmentation. *Med Image Comput Comput Assist Interv* (2010) 13(Pt 3):73–80. doi: 10.1007/978-3-642-15711-0_10

10. Heller N, Isensee F, Maier-Hein KH, Hou X, Xie C, Li F, et al. The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge. *Med Image Anal* (2021) 67:101821. doi: 10.1016/j.media.2020.101821

11. Zhang S, Yang G, Tang L, Lv Q, Li J, Xu Y, et al. Application of a Functional3-dimensional perfusion model in laparoscopic partial nephrectomy with precise segmental renal artery clamping. *Urology* (2019) 125:98–103. doi: 10.1016/j.urology.2018.12.023

12. Liskowski P, Krawiec K. Segmenting retinal blood vessels with deep neural networks. *IEEE Trans Med Imaging* (2016) 35(11):2369–80. doi: 10.1109/TMI.2016.2546227

13. Yang G, Li G, Pan T, Kong Y, Wu J, Shu H, et al. (2018). Automatic segmentation of kidney and renal tumor in CT images based on 3D fully convolutional neural network with pyramid pooling module, in: *2018 24th International Conference on Pattern Recognition (ICPR)*.

14. He Y, Yang G, Yang J, Chen Y, Kong Y, Wu J, et al. Dense biased networks with deep priori anatomy and hard region adaptation: Semi-supervised learning for fine renal artery segmentation. *Med Image Anal* (2020) 63:101722. doi: 10.1016/j.media.2020.101722

15. Jonathan Masci UM, Cire¸san D, Schmidhuber J¨urgen. (2011). Stacked convolutional auto-encoders for hierarchical feature extraction, in: *International Conference on Artificial Neural Networks*, . pp. 52–9, Part I(LNCS 6791).

16. Kutikov A, Uzzo RG. The R.E.N.A.L. nephrometry score: a comprehensive standardized system for quantitating renal tumor size, location and depth. *J Urol* (2009) 182(3):844–53. doi: 10.1016/j.juro.2009.05.035

17. Borgmann H, Reiss AK, Kurosch M, Filmann N, Frees S, Mager R, et al. R.E.N.A.L. score outperforms PADUA score, c-index and DAP score for outcome prediction of nephron sparing surgery in a selected cohort. *J Urol* (2016) 196 (3):664–71. doi: 10.1016/j.juro.2016.03.176

18. Leslie S, Gill IS, de Castro Abreu AL, Rahmanuddin S, Gill KS, Nguyen M, et al. Renal tumor contact surface area: a novel parameter for predicting complexity and outcomes of partial nephrectomy. *Eur Urol* (2014) 66(5):884–93. doi: 10.1016/j.eururo.2014.03.010

19. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* (2004) 240(2):205–13. doi: 10.1097/01.sla.0000133083.54934.ae

20. Krane LS, Hemal AK. Emerging technologies to improve techniques and outcomes of robotic partial nephrectomy: Striving toward the pentafecta. *Urol Clin North Am* (2014) 41(4):511–9. doi: 10.1016/j.ucl.2014.07.006

21. Ukimura O, Gill IS. Imaging-assisted endoscopic surgery: Cleveland clinic experience. *J Endourol* (2008) 22(4):803–10. doi: 10.1089/end.2007.9823

22. Porpiglia F, Fiori C, Checcucci E, Amparore D, Bertolo R. Hyperaccuracy three-dimensional reconstruction is able to maximize the efficacy of selective clamping during robot-assisted partial nephrectomy for complex renal masses. *Eur Urol* (2018) 74(5):651–60. doi: 10.1016/j.eururo.2017.12.027

23. Porpiglia F, Checcucci E, Amparore D, Piramide F, Volpi G, Granato S, et al. Three-dimensional augmented reality robot-assisted partial nephrectomy in case of complex tumours (PADUA >/=10): A new intraoperative tool overcoming the ultrasound guidance. *Eur Urol* (2020) 78(2):229–38. doi: 10.1016/j.eururo.2019.11.024

24. Wang C, Roth HR, Kitasaka T, Oda M, Hayashi Y, Yoshino Y, et al. Precise estimation of renal vascular dominant regions using spatially aware fully convolutional networks, tensor-cut and voronoi diagrams. *Comput Med Imaging Graph* (2019) 77:101642. doi: 10.1016/j.compmedimag.2019.101642

25. Dou Q, Yu L, Chen H, Jin Y, Yang X, Qin J, et al. 3D deeply supervised network for automated segmentation of volumetric medical images. *Med Image Anal* (2017) 41:40–54. doi: 10.1016/j.media.2017.05.001

26. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* (2017) 36:61–78. doi: 10.1016/j.media.2016.10.004

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 10 November 2022
DOI 10.3389/fonc.2022.1019009

OPEN ACCESS

EDITED BY
Hani J. Marcus,
University College London,
United Kingdom

REVIEWED BY
Jitao Wang,
Southeast University, China
Shijie Li,
Zhejiang University, China

*CORRESPONDENCE
Jiahong Dong
dongjiahong@mail.tsinghua.edu.cn
Shizhong Yang
ysza02008@btch.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Surgical Oncology, a section of the journal Frontiers in Oncology

RECEIVED 14 August 2022
ACCEPTED 25 October 2022
PUBLISHED 10 November 2022

CITATION
Wang L, Wu M, Zhu C, Li R, Bao S, Yang S and Dong J (2022) Ensemble learning based on efficient features combination can predict the outcome of recurrence-free survival in patients with hepatocellular carcinoma within three years after surgery. *Front. Oncol.* 12:1019009. doi: 10.3389/fonc.2022.1019009

Ensemble learning based on efficient features combination can predict the outcome of recurrence-free survival in patients with hepatocellular carcinoma within three years after surgery

Liyang Wang[1†], Meilong Wu[2†], Chengzhan Zhu[3], Rui Li[3], Shiyun Bao[2], Shizhong Yang[4*] and Jiahong Dong[1,4*]

[1]School of Clinical Medicine, Tsinghua University, Beijing, China, [2]Division of Hepatobiliary and Pancreas Surgery, Department of General Surgery, Shenzhen People's Hospital (The Second Clinical Medical College, Jinan University; The First Affiliated Hospital, Southern University of Science and Technology), Shenzhen, Guangdong, China, [3]Department of Pediatric Surgery, The Affiliated Hospital of Qingdao University, Qingdao, China, [4]Hepato-pancreato-biliary Center, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsing-hua University, Beijing, China

Preoperative prediction of recurrence outcome in hepatocellular carcinoma (HCC) facilitates physicians' clinical decision-making. Preoperative imaging and related clinical baseline data of patients are valuable for evaluating prognosis. With the widespread application of machine learning techniques, the present study proposed the ensemble learning method based on efficient feature representations to predict recurrence outcomes within three years after surgery. Radiomics features during arterial phase (AP) and clinical data were selected for training the ensemble models. In order to improve the efficiency of the process, the lesion area was automatically segmented by 3D U-Net. It was found that the mIoU of the segmentation model was 0.8874, and the Light Gradient Boosting Machine (LightGBM) was the most superior, with an average accuracy of 0.7600, a recall of 0.7673, a F_1 score of 0.7553, and an AUC of 0.8338 when inputting radiomics features during AP and clinical baseline indicators. Studies have shown that the proposed strategy can relatively accurately predict the recurrence outcome within three years, which is helpful for physicians to evaluate individual patients before surgery.

KEYWORDS

recurrence prediction, efficient features, ensemble learning, hepatocellular carcinoma, surgery

1 Introduction

Hepatocellular carcinoma (HCC) accounts for 85%-90% of the main pathological types of primary liver cancer (1–3). It is easy to spread in the liver through the portal vein system to form intrahepatic metastasis, and it is also easy to form tumor thrombus in the portal vein and cause portal hypertension. HCC is mostly found in the middle and late stages, which leads to its generally poor prognosis (4–8). According to statistics, the recurrence rate of HCC after surgery is as high as about 70% (9), and the survival rate is only 15%-40% (10). Fortunately, treatment modalities represented by precision surgery have greatly improved patient prognosis. Liver resection with early diagnosis can improve the survival rate of patients within one year to 91%-98% (11, 12). Therefore, rational clinical decision-making is essential to reduce recurrence and improve survival.

Accurate preoperative prediction of recurrence can help doctors assess the necessity and risk of surgery, so that they can design rational clinical decisions. Early (1-2 years after surgery) (13) and long-term (5 years and beyond) (14) recurrence predictions have been performed in a small number of studies, with encouraging results. It is worth noting that the recurrence rate of HCC within 3 years after surgery is 50-55%, which accounts for about 71%-78% of the total recurrence (15). Three years after surgery is a critical period, and the absence of recurrence within 3 years indicates a relatively good prognosis. There is no doubt that preoperative prediction of the recurrence outcome in patients within 3 years after surgery is also of great significance for evaluating the illness and selecting treatment options.

The rise of artificial intelligence (AI) technology has brought new strategies for the prediction of HCC recurrence, especially novel data processing methods represented by machine learning and radiomics. Studies have shown that patients' preoperative imaging, personal information and clinical manifestations are closely related to prognosis (16, 17). Because of this, some researchers have employed the preoperative performance of patients to predict postoperative recurrence through AI algorithms. Ji et al. (18) collected data on 480 patients undergoing HCC resection from 3 centers. Combined with radiomics characteristics and some biochemical indicators, a Cox-based recurrence risk prediction model was constructed, and the final C-index reached 0.633-0.699. Zeng et al. developed a random survival forest (RSF) model using the 15 characteristics of HCC patients. The model obtained a C-index of 0.725 on the validation set, which was encouraging. Huang et al. (19) developed a machine learning prognostic model to identify high-risk patients after surgical resection. The results show that the eXtreme Gradient Boosting tree (XGBoost) achieved the best discrimination in the internal validation queue. In reference (20), 143 features were extracted, including 26 preoperative clinical features, 5 postoperative pathological features, and 112 imaging features, for predicting early recurrence of HCC. As a result, the area under the receiver operating characteristic curve (AUC) of the preoperative model was 0.739, with relatively strong generalization ability.

Nevertheless, there is still room for improvement in the current related work. For example, the lesion area adopted to extract features in most studies needs to be manually segmented from the original image, which brings great challenges to improving work efficiency and reducing costs. In addition, the features of the input model are often not concise and efficient, which will lead to a decrease in accuracy. It is necessary to explore efficient feature representations and achieve automatic and accurate predictions.

This study aimed to develop an excellent predictive strategy for recurrence-free survival (RFS) outcomes in patients with HCC within 3 years after surgery. A 3-dimension deep learning framework was applied to automate lesion segmentation. Seven feature representation methods were compared to explore the most superior feature combinations, including clinical baseline indicators, radiomics features during arterial phase (AP), portal venous phase (PVP), and delayed phase (DP), and combination of clinical data with radiomics features during each phase. Four novel Boosting ensemble learning models were selected for prediction of recurrence outcome. This work has the following highlights:

- Deep learning was employed for automatic segmentation of regions of interest (ROI), which avoided the drawbacks of manual delineation.
- Seven feature representations were explored to find the best model input.
- The study compared novel Boosting ensemble learning methods to select the model with best performance, which may be applicable in the future.

2 Materials and methods

The workflow of this study is shown in Figure 1.

2.1 Patients

HCC patients who underwent partial hepatectomy in Qingdao University Affiliated Hospital from January 2014 to December 2018 were followed after surgery regularly. The inclusion criteria were as follows: 1. The pathological diagnosis was HCC; 2. The first treatment was partial hepatectomy; 3. Enhanced CT examination was performed within 1 month before surgery, and all periods were completed; 4. The

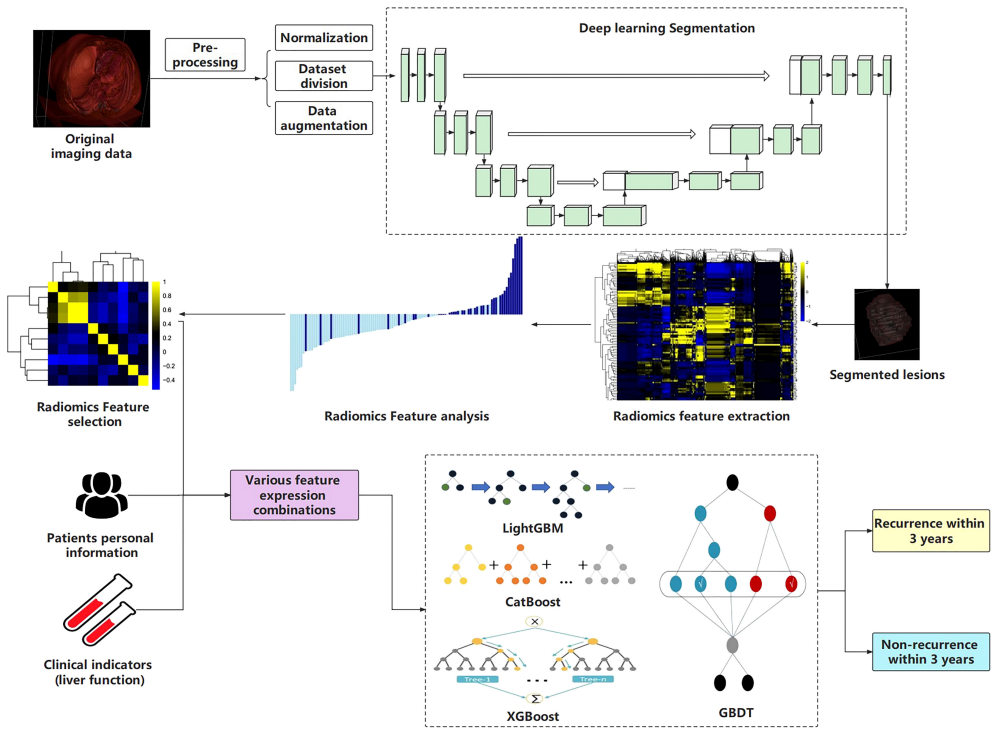


FIGURE 1
The workflow of this study.

patient's personal information and relevant clinical data were complete; 5. It has been confirmed that whether the recurrence occurred within 36 months after surgery. The following were the exclusion criteria: 1. Patients who have received chemotherapy, interventional therapy, targeted therapy, etc. before partial hepatectomy; 2. Patients with a history of other tumors; 3. Patients whose tumors have metastasized; 4. Imaging and clinical data were incomplete; 5. The follow-up data were incomplete or the recurrence outcome within 3 years couldn't be judged. Additionally, all patients included in the study underwent radical hepatectomy. The criteria for radical hepatectomy were: (1) no residual tumor was found at the margin of resection, which was negative; (2) no tumor was found in the remaining liver; (3) tumor markers returned to normal within two months after surgery. Ultimately, 105 patients were selected for the study. RFS period is defined as the time from the date of liver resection to the date of recurrence and within 3 years after surgery is within 36 months from the date of liver resection.

It must be emphasized that the principles of the Declaration of Helsinki were followed and the study was approved by the hospital ethics committee (ethics number: 20001-01). All patients signed an informed consent certificate before surgery.

2.2 Imaging acquisition

The scanning equipment for the detection was the German CT (SOMATOM Definition Flash, Siemens) and the American Discovery CT (GE Healthcare). The scanning method was a three-level contrast-enhanced scan of the upper abdomen, and the scanning range was from the top of the liver to the lower edge of the two kidneys. During the scanning process, the voltage, current, scanning layer thickness, layer spacing, and pixel matrix size were set to 120 kV, 200-350 mA, 5 mm, 5 mm, and 512 × 512, respectively. Workers administered iohexol and 350 mg/m1 of iodine through a peripheral vein at a flow rate of 3.0 ml/s and a dose of 1.5 ml/kg under the action of a pressure syringe. Finally, AP, PVP, and DP images were obtained for the study.

2.3 Lesion segmentation

Generally, studies mostly segment lesions manually, which reduces work efficiency. Based on the previous manual annotation, we built a 3D U-Net deep learning model for automatic and accurate segmentation of lesions.

2.3.1 Manual annotation

This work adopted the supervised learning to automatically segment the ROIs, so manual annotation was required before model training. Two physicians with extensive experience in radiology were selected for this task, one of whom delineated the tumor area of each slice with the help of 3D Slicer (Boston, MA, USA) software without knowing any patient's baseline data, and the other one was responsible for checking the annotation results. Once there was a dispute, return to discuss and re-mark if necessary. All CT images for the three periods were delineated and formed into volumes of interest (VOIs).

2.3.2 Data pre-processing

Considering that some slices in CT images do not contain ROIs, this will increase the computational complexity. Slices without lesions were cropped according to the annotated images and the remainders were studied. Moreover, we normalized the image format to 256×256×48 for better input to the model. In order to expand the amount of data, data augmentation operations were performed on the divided training set, including but not limited to image flipping, rotation, cropping, scaling, and blurring (21, 22).

2.3.3 Construction of segmentation model

CT images have 3D structures, and the traditional method convert them into 2D slices and then send into the 2D segmentation model, which results in the loss of spatial information. In this study, a 3D convolutional neural network (3D U-Net) was constructed to segment lesions directly, which comprehensively preserved the spatial information between slices (23, 24).

Similar to the classic U-Net, the 3D U-Net also consists of Encoder and Decoder, each of which contains four sub-modules. In the Encoder, each sub-module contains two $3 \times 3 \times 3$ convolutional layers, and each convolutional layer is connected to an activation function. After completing the convolution operation, max-pooling with a stride of 2 is performed on each dimension. In Decoder, each sub-module contains an upsampling process (deconvolution operation) with a stride of 2, and then two $3 \times 3 \times 3$ convolutional layers and activation functions are added in turn. It must be emphasized that the padding in the convolutional layer of this module is set to 1, which makes the convolution operation not change the size of the image. Changes in image size are completely controlled by pooling and upsampling. Additionally, the last sub-module of the Decoder consists of a $1 \times 1 \times 1$ convolutional layer, which reduces the number of output feature maps. Batch normalization (BN) was introduced before each activation function.

This work aims to segment liver tumors from other tissues, where the input channel of the model was set to $256 \times 256 \times 48$, and the activation function adopted ReLU. After the construction was completed, the total parameters and the trainable parameters of the neural network reached 4,122,466 and 4,117,570, respectively.

2.4 Radiomics feature extraction and selection

Feature extraction is an essential part of radiomics analysis. In this study, we performed radiomic feature extraction for segmented liver tumors. Using the Pyradiomics 3.0.1 library in Python, a total of 788 dimensional features including Shape, Firstorder, GLCM, GLRLM, GLSZM, and GLDM were extracted. Each type of features was performed 9 transformations including Original, Wavelet-LLH, Wavelet-LHL, Wavelet-LHH, Wavelet-HLL, Wavelet-HLH, Wavelet-HHL, Wavelet-HHH, and Wavelet-LLL. Among them, "Wavelet-XXX" represents the wavelet transform, followed by the corresponding basis function type.

Due to the high dimension of the extracted features, it is easy to cause "dimensionality disaster" and affect the model performance. Therefore, selecting features with large contributions can reduce the dimension as much as possible without affecting the comprehensiveness of the features. This work employed the Least Absolute Shrinkage Selector Operator (Lasso) algorithm to select the extracted features and ranked the contribution of each feature. By constructing a penalty function, Lasso can compress the coefficients of variables and make some regression coefficients 0, so as to achieve the purpose of variable selection. In addition, Lasso can also filter variables and reduce the complexity of the model. The variable screening here refers to not putting all the variables into the model for fitting, but selectively putting the variables into the model to get better performance parameters. Complexity adjustment refers to controlling the complexity of the model through a series of parameters to avoid overfitting. The optimal model was fit and the value of the penalty parameter α was determined based on the sklearn library in Python. For the dimensionality-reduced features, correlation coefficients and cluster heatmaps, as well as the coefficient distribution of each feature are visualized to better interpret the radiomics features.

2.5 Selection of clinical baseline features

This study collected clinical baseline data of HCC patients in addition to CT images, such as personal information and clinical indicators. The gender and age of patients were collected as personal information data. Clinical indicators here were mainly tumor markers and liver function indicators, including alpha-fetoprotein (AFP), hepatitis B surface antigen (HBsAg), albumin (ALB), the total bilirubin (T- BIL), alanine aminotransferase (ALT) and aspartate aminotransferase (AST), etc. It should be noted that positive and negative results were obtained for AFP

and HBsAg, while other liver function indicators were represented as specific test results.

2.6 Construction of recurrence prediction models

A total of seven feature representations, including selected radiomics features during AP, PVP, and DP, clinical baseline features, and their combined features, were input into the recurrence prediction models. The Boosting ensemble learning algorithms were adopted to predict the RFS outcome within 3 years.

2.6.1 Light gradient boosting machine

Gradient Boosting Decision Tree (GBDT) is a classic ensemble algorithm in machine learning. Its main idea is to employ weak classifiers (decision trees) to iteratively train to obtain the optimal model, which has the advantages of good training effect and not easy overfitting. LightGBM (Light Gradient Boosting Machine) is a framework for implementing the GBDT algorithm. It supports efficient parallel training and has faster training speed, lower memory consumption, better accuracy, support for distributed and fast processing of massive data, etc. (25). Currently, this framework has been relatively widely used in the field of medical data processing (26–28), but it has not been attempted in the HCC recurrence prediction task.

A leaf-wise algorithm with a depth limit is adopted in LightGBM. This strategy finds the leaf with the largest split gain from all the current leaves each time, and then splits and loops, which reduce more errors and get better accuracy under the same number of splits. Moreover, the Gradient-based One-Side Sampling (GOSS) operation is proposed to reduce computation and improve accuracy. This method does not calculate the gradient through the sample points used, but calculates the gradient by partial sampling. The Exclusive Feature Bundling (EFB) is also proposed to bundle some features together to reduce the feature dimension, thereby reducing the time-consuming to find the best fork. This study implemented the LightGBM algorithm based on the sklearn library in Python to perform the binary classification task, that is, recurrence or not within 3 years.

2.6.2 Categorical boosting

Categorical Boosting (CatBoost), as a novel ensemble learning algorithm, has been applied to some medical data processing tasks, but has not been used to predict HCC recurrence (29, 30). Catboost adopts the oblivious tree as the base tree model, which is characterized by the same segmentation features in each layer. Leaf nodes can be converted to binary codes, and the value of the node is stored in a floating-point vector of length 2 to the power of d (d is the depth of the tree). One of the advantages of this tree is that the prediction performance is better, and this structure can also weaken the shortcomings of easy fitting in decision trees to a certain extent. When Catboost completes training, it stores the leaf node value of each tree into a vector. When predicting, it can quickly retrieve the corresponding leaf node value by judging which leaf node it is in, so it can improve the prediction efficiency and enhance the model performance. This work selected it for predicting HCC recurrence.

2.6.3 eXtreme gradient boosting

XGBoost has been widely used in the field of medical data analysis since it was proposed in 2014 (31, 32). In the HCC recurrence prediction task, this algorithm was also tried and achieved significant results (19). Its greedy algorithm-based split node calculation and missing value handling techniques are very suitable for data mining. The algorithm was trained to predict RFS outcomes and compared with other models such as LightGBM and CatBoost.

2.6.4 Gradient boosting decision tree

We also employed GBDT as the baseline model for comparison. It is an ensemble learning algorithm based on decision trees that iterates over new learners through gradient descent. In this paper, the classification task was performed, and the Classification And Regression Tree (CART) was selected.

2.7 Statistical analysis

For the analysis of clinical baseline data, the differences involved in this study were compared using student t-test or Mann-Whitney U-test, where the criterion of significant difference was set at $P<0.05$. Mean ± 95% confidence interval (CI) was calculated as results for continuous variables. To reflect the criticality of certain variables, the univariate Kaplan-Meier curve was introduced for survival analysis.

We calculated the mean Intersection overUnion (mIoU), accuracy (Acc), Kappa and Dice coefficients of 3D U-Net to reflect the segmentation effect. Additionally, Acc, recall, precision (Prec), F_1 score, receiver operating characteristic curve (ROC) and corresponding AUC were introduced as performance evaluation criteria for the ensemble learning models. It should be emphasized that the classification threshold was set to 0.5.

2.8 Experimental setup

The image data during the three scanning periods were randomly divided into training set, validation set and test set according to the ratio of 8:1:1. The segmentation model was trained on the training set and validation set, and the test set was employed to demonstrate the performance. All lesions segmented by the model during three periods were acquired and their radiomic features were extracted. For the Lasso regression algorithm, the study obtained the best α value through 10-fold cross-validation to select key features. Considering the small sample size, this study selected the 5-fold cross-validation method to determine the features representation and predict the recurrence outcome, and calculated the mean value of five experiments and the corresponding 95% CI as the results. The relevant computing equipment for this experiment was configured with a CPU AMD Ryzen 7 5800H (16 GB memory) and a GPU NVIDIA® Tesla V100 (32 GB memory) with acceleration support of the compute unified device architecture (CUDA). All work was carried out in the Windows 10 operating system, and the programming language, deep learning framework and key libraries included Python 3.7, Pytorch, Pyradiomics, sklearn, VTK, etc.

3 Results

3.1 Analysis of patients' basic data

During follow-up, 52 patients (49.52%) were found to have recurrence within 3 years after surgery, of which 46 (88.46%) were male and 6 (11.54%) were female; 24 (46.15%) were aged 60 years or older and 28 (53.85%) were younger than 60 years old; 34 (65.38%) were AFP positive, and 18 (34.62%) were negative; 51 (98.08%) were HBsAg positive, and 1 (1.92%) were negative. 53 patients (50.48%) were found to have no recurrence within 3 years after surgery, of which 41 (77.36%) were male and 12 (22.64%) were female; 25 (47.17%) were aged 60 years or order and 28 (52.83%) were younger than 60 years old; 30 (56.60%) were AFP positive, and 23 (43.40%) were negative; 45 (84.91%) were HBsAg positive, and 8 (15.09%) were negative. Based on this, a univariate Cox proportional hazards model was established to judge the influence of different factors on RFS, and the related results were represented by the Kaplan-Meier curves (Figure 2). Through the statistics of gender classification group (HR=1.85, P=0.155) and HBsAg result classification group (HR=6.15, P=0.072), it was found that gender and HBsAg affect RFS to some extent although the differences were not significant, followed by AFP (HR=1.37, P=0.280). Notably, age was not significantly associated with recurrence outcome from the age-categorized group in this study (HR=0.90, P=0.711). However, patient's age is a key factor affecting prognosis from previous studies (20, 33), so we still regarded it as one of the features. Table 1 shows the statistical results of some continuous clinical indicators. It can be found that ALB, T-BIL, ALT, and AST (P=0.149, 0.377, 0.128, and 0.223, respectively) were relatively significantly different or not significantly different between the recurrence and non-recurrence groups.

3.2 Results of lesion segmentation

The training and validation sets during the three periods were input into 3D U-Net for training, and the model performance was optimized through parameter adjustment and continuous iteration. The key hyperparameters were set as follows: Momentum optimizer was selected and set to 0.9, initial learning rate, weight_decay and batch_size were set to 0.001, 4.0×10^{-3}, 2, respectively. After the model iterated for 500 epochs, it fully converged (the loss value of the validation set was lower than 0.001). At this point, we stopped the training and saved the parameters. The performance on the test set was excellent, with mIoU of 0.8874, Acc of 0.9915, Kappa of 0.8738 and Dice coefficient of 0.9360, which indicates that the deep learning model has strong generalization ability for segmenting liver lesions. To visually compare the segmentation effects, this paper presents 3D reconstruction visualization images of the upper abdomen based on CT scans, manually annotated tumors, and deep learning-segmented tumors (Figure 3). The VTK library in Python was adopted as the relevant drawing tool. It must be emphasized that the lesion areas involved in subsequent calculations were automatically segmented by the trained model.

3.3 Results of radiomics feature extraction and selection

A total of 788 radiomic features were extracted in this study, including 100 features from original transform and 688 features from wavelet transform. In the original transform, the extracted contents were 14 shapes, 18 firstorder, 22 GLCM, 16 GLRLM, 16 GLSZM and 14 GLDM features. In the wavelet transform, the contents extracted by Wavelet-LLH, Wavelet-LHL, Wavelet-LHH, Wavelet-HLL, Wavelet-HLH, Wavelet-HHL, Wavelet-HHH, and Wavelet-LLL included 144 firstorder, 176 GLCM, 128 GLRLM, 128 GLSZM and 112 GLDM features. Since high-dimensional features may affect model performance, dimensionality reduction and selection of contributing features is significant.

The Lasso algorithm was used for fitting to obtain the best α values during AP, PVP and DP, respectively. The model was fully converged after 10,000 iterations based on the 10-fold cross-validation. The optimized α values for AP, PVP and DP were calculated as 0.0518, 0.0244 and 0.0202, respectively.

10.3389/fonc.2022.1019009

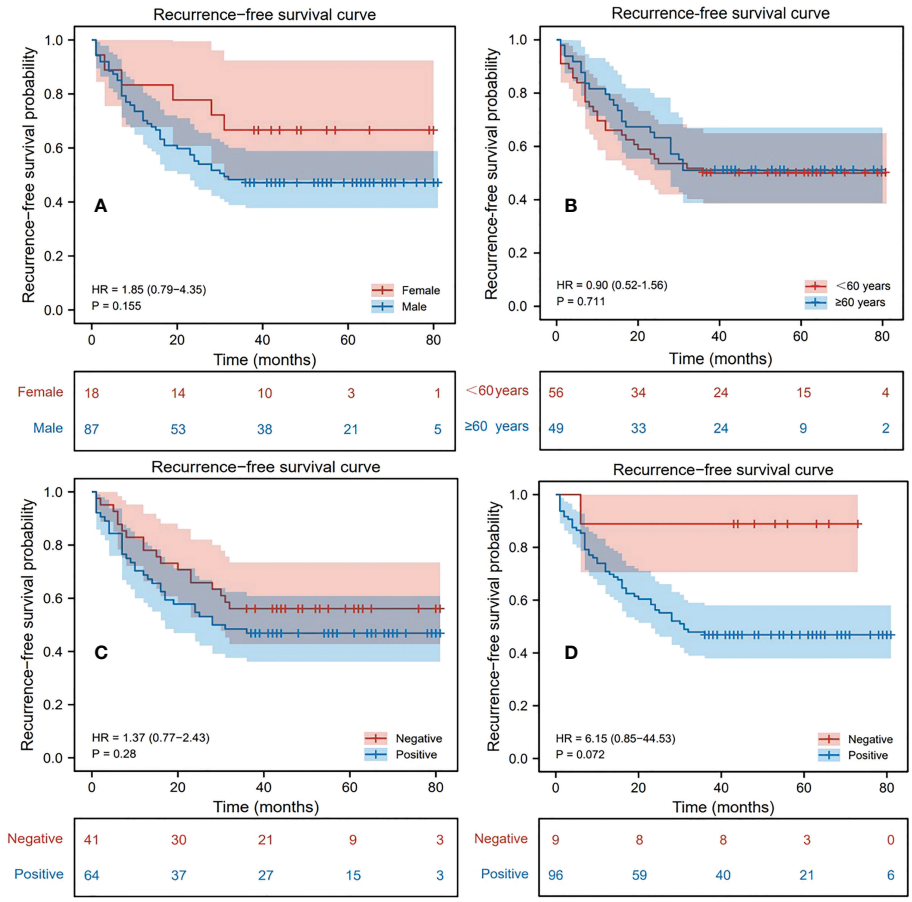


FIGURE 2
Kaplan-Meier survival analysis curve of patients, where the variables in **(A–D)** are gender, age, alpha-fetoprotein (AFP), hepatitis B surface antigen (HBsAg) respectively.

Meanwhile, 22, 38, and 41 features with contribution degrees were selected during the above three periods respectively. Figures 4A, B, and C show the selected feature names and the corresponding coefficients distribution in AP, PVP, and DP, respectively. Figure 5 shows the correlation coefficient between the features and the clustering results through heatmaps (the color depth represents the correlation strength).

3.4 Results of recurrence prediction

3.4.1 Comparison of different feature representations

Seven feature representation methods for evaluating the prognosis of HCC were considered, including clinical baseline features, radiomics features of AP, radiomics features of PVP,

TABLE 1 Statistical results of 4 clinical indicators.

Clinicalindicator	Total dataset (N = 105)		P-value
	Recurrence (N =52)	Non-recurrence (N = 53)	
ALB(g/L)	40.36 ± 1.28	40.42 ± 1.26	0.1493
T-BIL(mmol/L)	20.32 ± 4.81	20.34 ± 4.69	0.3765
ALT(u/L)	51.12 ± 15.42	50.89 ± 15.02	0.1282
AST(u/L)	42.78 ± 12.23	42.20 ± 11.92	0.2233

all outcomes are based on recurrence within 3 years after surgery. ALB, T-BIL, ALT, and AST represent albumin, the total bilirubin, alanine aminotransferase and aspartate aminotransferase, respectively. Each indicator is represented by the mean of the sample and the corresponding 95% confidence interval (CI).

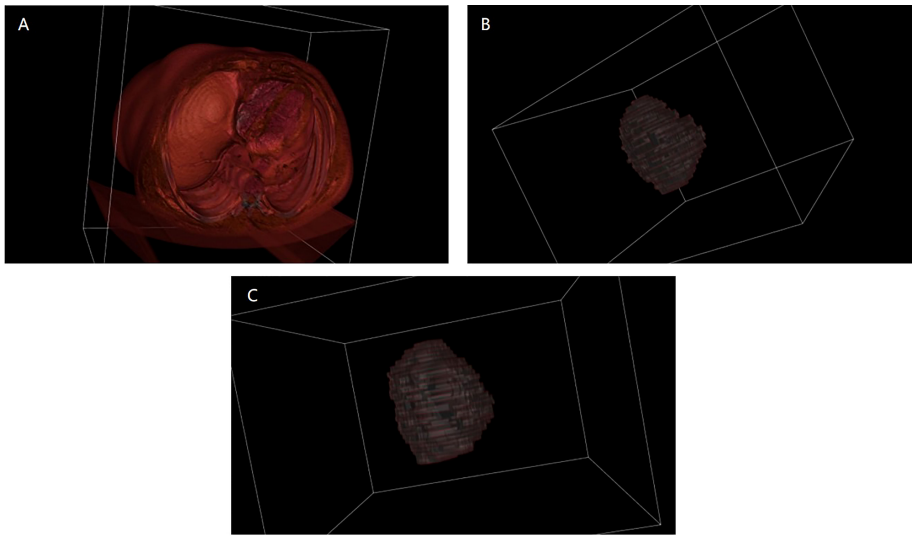


FIGURE 3
3D reconstruction visualization images before and after segmentation. **(A)** is the 3D visualization of the original CT image before segmentation; **(B)** is the 3D visualization after manually segmenting the tumor; **(C)** is the 3D visualization after segmenting the tumor using deep learning.

radiomics features of DP, radiomics features of AP combined with clinical indicators, radiomics features of PVP combined with clinical indicators, and radiomics features of DP combined with clinical indicators. In order to explore the most excellent feature representation, we separately input the above features into the ensemble learning algorithms and optimized the training. Considering the randomness of the results based on the small sample size, the training process adopted 5-fold cross-validation, that is, the dataset was randomly divided into 5 equal parts, 4 of which were used

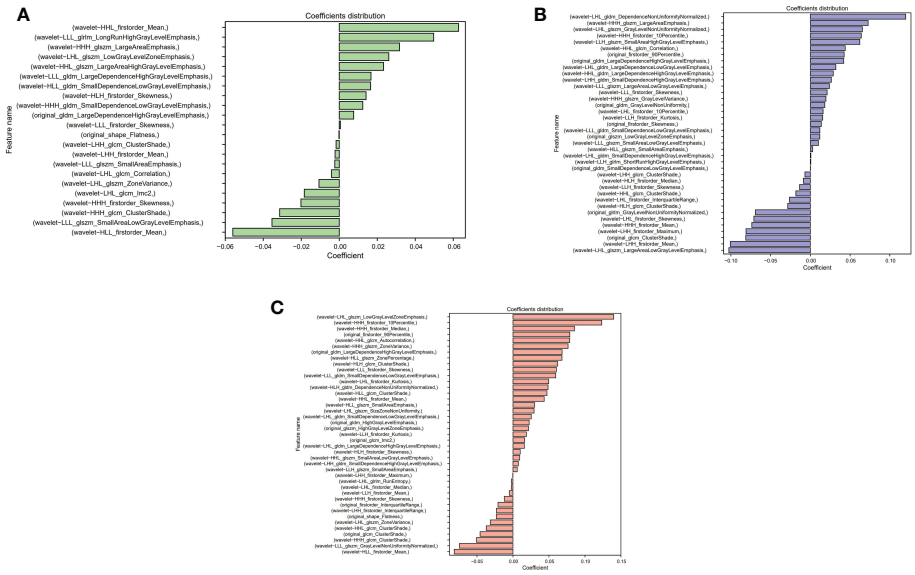


FIGURE 4
Distribution of selected radiomics feature coefficients. **(A–C)** show the features and their distributions during arterial phase (AP), portal venous phase (PVP) and delay period (DP), respectively.

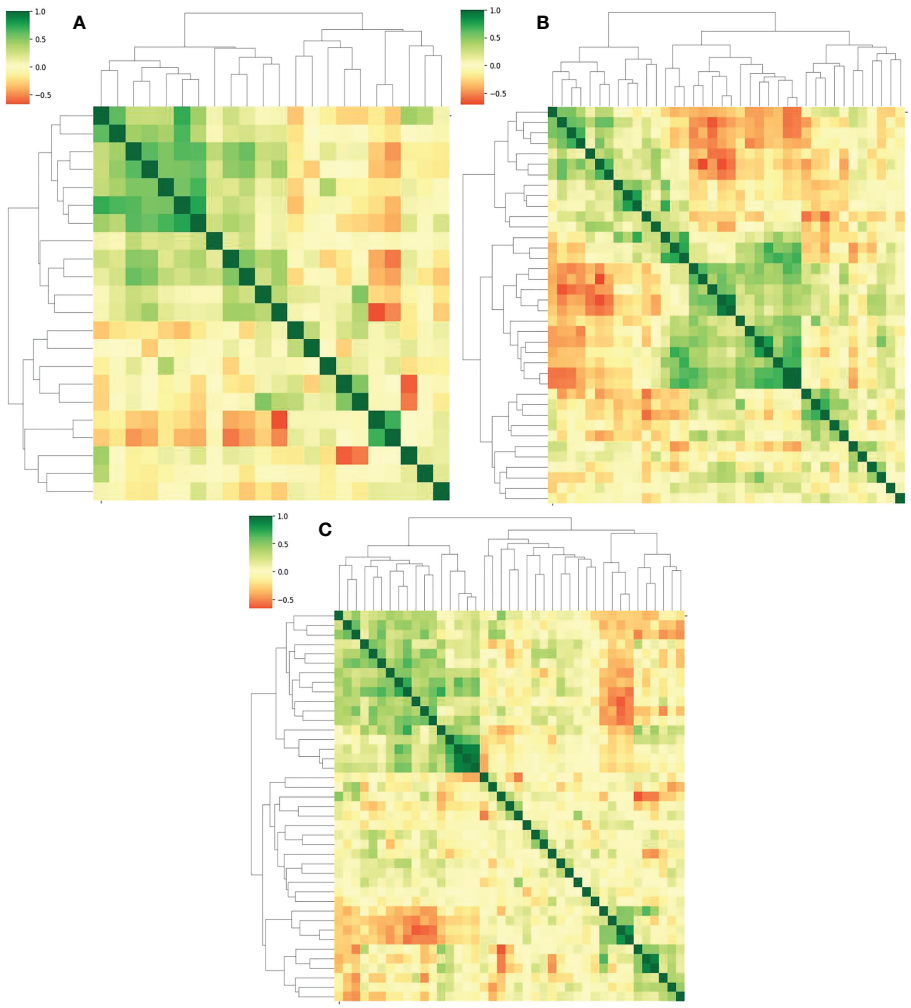


FIGURE 5
(A–C) represent the correlation and the clustering heatmaps between features during the AP, PVP and DP, respectively.

for training and the remaining 1 was used for testing. This step was repeated 5 times. The average value of 5 experiments and the corresponding 95% CI were regarded as the evaluation standard. Meanwhile, the ROC curves and their AUC values reflected the generalization ability of the models. The ROC curves of the models with different features were drawn and their AUC values were calculated. Due to space limitations, we only show the results using the LightGBM algorithm in Table 2 and Figure 6, and the rest of the results are in the Appendix. It can be seen that the effect of combining radiomics features with clinical baseline indicators was better than inputting radiomics features or clinical indicators alone, with AP combining obtaining the best effect, followed by DP combining and PVP combining. The effect of only inputting clinical indicators was the least satisfactory, which might be caused by too little information represented by the features.

3.4.2 Comparison of prediction models

Seven feature representations were employed to compare the performance of ensemble learning models. Likewise, the study performed five-fold cross-validation on each model and calculated the associated evaluation metrics. During training, GridSearchCV method was adopted to adjust the model parameters and no overfitting occurred for each model. Due to space limitations, we only show the results inputting the most effective feature representations in this section, and the rest of the results are in the Appendix. It is found that for the four ensemble learning algorithms, different feature expressions input have similar laws, so the following only analyzes the models when radiomics features during AP and clinical indicators are input. Table 3 shows certain key parameters of each model. The test results of the Boosting ensemble models are shown in Table 4. It can be found that the performance of LightGBM

TABLE 2 Comparison of recurrence prediction results of ensemble learning models using different feature representations.

Feature representation	Acc	Recall	Prec	F_1 score
Personal and	0.6062	0.6164	0.6019	0.6039
clinical indicators	± 0.0877	± 0.1429	± 0.0930	± 0.1027
AP	0.7224	0.6946	0.7528	0.7156
	± 0.0834	± 0.0944	± 0.1231	± 0.0804
PVP	0.6438	0.6782	0.6409	0.6570
	± 0.1117	± 0.1122	± 0.1019	± 0.1014
DP	0.6343	0.6909	0.6331	0.6560
	± 0.0690	± 0.0413	± 0.0872	± 0.0520
AP+	0.7495	0.7673	0.7402	0.7502
other indicators	± 0.0629	± 0.1051	± 0.0525	± 0.0710
PVP+	0.6824	0.6927	0.6844	0.6846
other indicators	± 0.0783	± 0.0941	± 0.0780	± 0.0771
DP+	0.6819	0.6309	0.7168	0.6630
other indicators	± 0.0659	± 0.0912	± 0.1273	± 0.0756

Each result is represented by the mean of 5 experiments and 95% CI.

was the most excellent, with an average Acc of 0.7600, recall of 0.7673, Prec of 0.7733, and F_1 score of 0.7553, which indicated that this algorithm can accurately predict recurrence outcome within 3 years after surgery. It is worth noting that XGBoost performed well in previous similar studies, but not as good as the former in this task. It had an Acc of 0.7224 and an F_1 score of 0.6936, which was not as superior to LightGBM. Additionally, as the baseline model, GBDT only obtained an average Acc of 0.6543, recall of 0.6382, Prec of 0.6600 and F_1 score of 0.6387. The per-fold and averaged ROC curves and corresponding AUC values are shown in Figure 7. LightGBM had the strongest generalization, and its AUC reached 0.8338 (CI: ± 0.0680), followed by CatBoost (0.8084 ± 0.0650), XGBoost (0.7441 ± 0.0946), and GBDT (0.7343 ± 0.0214).

4 Discussion

In this study, the LightGBM model was constructed for the first time to accurately predict the recurrence outcome of HCC within three years after surgery. An efficient feature representation was explored, that is, the combination of radiomics features of tumor during AP, patient personal information, and clinical indicators. We trained the deep learning automatic segmentation model to make the process efficient. The results show that the proposed method was the most effective, achieving an accuracy of 0.7600 and an AUC of 0.8338.

Compared with manual segmentation, although the effect of deep learning segmentation is not as good as the former, it has higher efficiency and lower labor cost (34). In this paper, the mIoU of 3D U-Net reached 0.8874, which indicated that this algorithm can accurately segment the liver tumor region. It only took 1.22-1.85s to execute each sample on the local device, which was much faster than the manual way. It is undeniable that deep learning with excellent performance is the future trend of lesion segmentation methods (35).

This work selected 22 radiomics features during AP combined with 8 clinical baseline features from the seven feature representations and validated superiority. This feature combination eliminated dimensional redundancy, including tumor features with large contribution coefficients and clinical factors that affect prognosis. Notably, the present study found that the radiomics features during AP were superior to during PVP and DP, suggesting that AP might better capture features affecting recurrence. In addition, combined representations outperformed individual clinical or radiomics feature representations. Possibly the combination increased the amount of available information, making the model more likely to learn complex preoperative-prognostic associations (36).

Four novel Boosting ensemble models were adopted for comparison, among which LightGBM achieves the best performance (AUC=0.8338), outperforming CatBoost (AUC=0.8084), XGBoost (AUC=0.7441) and GBDT (AUC=0.7343) when inputting radiomics features during AP and clinical baseline indicators. Previous studies have confirmed the state-of-the-art of the XGBoost algorithm in the HCC prognosis prediction task (19). XGBoost belongs to the boosting family and is an engineering implementation of the GBDT algorithm. It focuses the residuals during training, uses a second-order Taylor expansion in the objective function and adds regularization. Meanwhile, the exact greedy idea is adopted in the generation process of the decision tree. When looking for the best split point, a pre-sort algorithm is adopted, that is, all features are pre-sorted according to the value of the feature, and then all the split points on all the features are traversed, and the

10.3389/fonc.2022.1019009

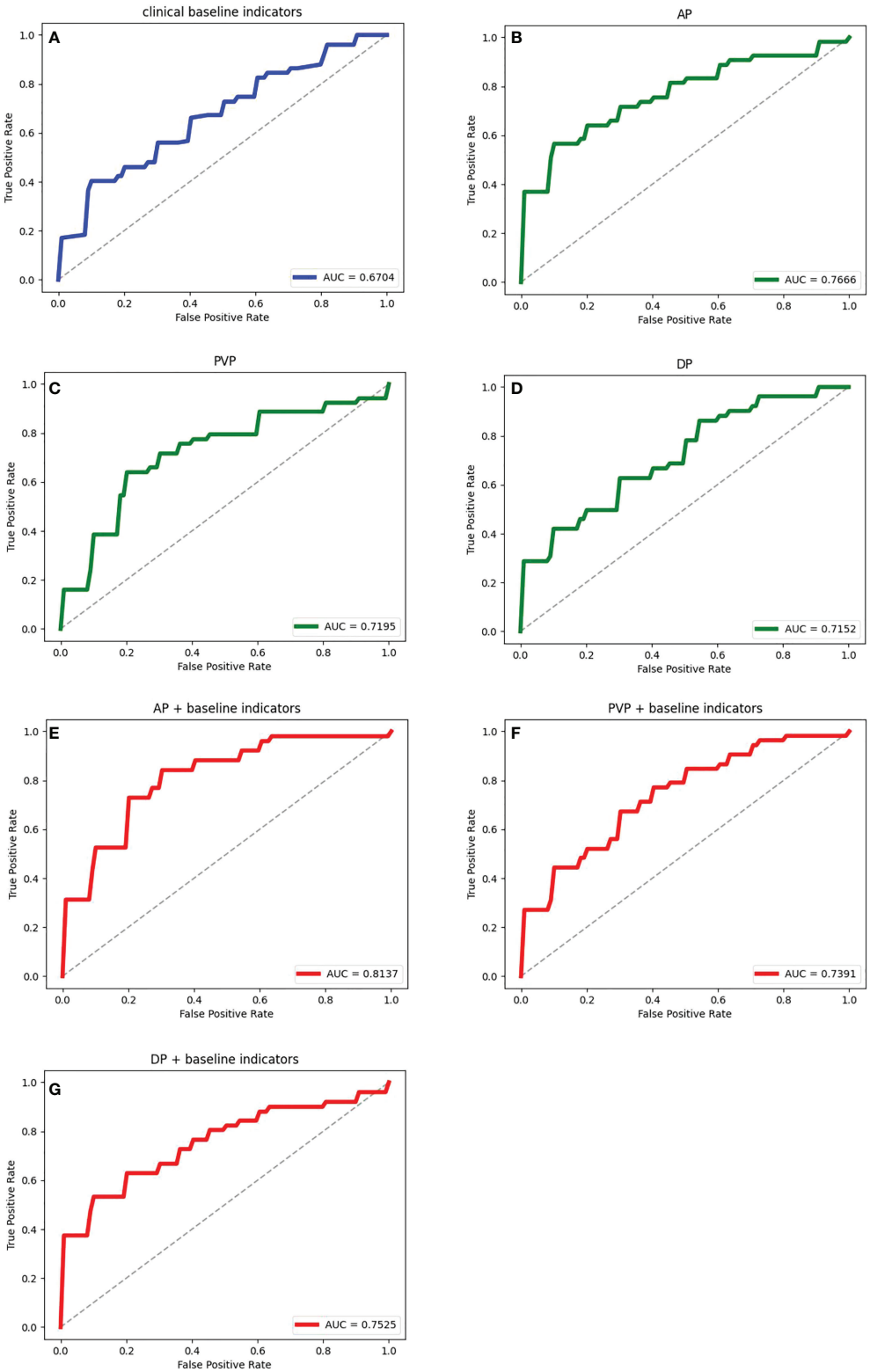


FIGURE 6
The ROC curves and the corresponding AUCs of the ensemble learning model with different feature representations. **(A)** is the result of inputting personal information and clinical indicators; **(B–D)** are the results of inputting radiomic features during AP, PVP and DP respectively; **(E–G)** are the results of inputting radiomics features during AP, PVP and DP combined with clinical data respectively.

TABLE 3 Key parameter settings for each ensemble learning model.

Model	Parameter name	Parameter settings
LightGBM	n_jobs	-1
	n_estimators	600
	learning_rate	0.01
	max_depth	5
	num_leaves	32
	colsample_bytree	0.51
	subsample	0.6
CatBoost	iterations	5000
	learning_rate	0.01
	l2_leaf_reg	3
	bagging_temperature	1
	subsample	0.6
	random_strength	1
	depth	6
	border_count	128
XGBoost	learning_rate	0.001
	n_estimators	1000
	max_depth	5
	min_child_weight	1
	gamma	0
	subsample	0.6
	colsample_bytree	0.8
	seed	27
GBDT	n_estimators	1000
	learning_rate	0.01
	max_depth	5
	random_state	4

total number of samples split according to these candidate split points is calculated. The objective function gain is to find the feature and candidate splitting point corresponding to the maximum gain, so as to split. XGBoost training is performed by addition, that is, each time a tree is trained by focusing residuals, and the final prediction result is the sum of all trees. However, XGBoost performs pre-sorting in the selection of optimal split points, and then calculates the objective function gain of all samples for all split points of all features. The space and time complexity of this process is very large, and to a certain extent affects the accuracy (31).

To address this issue, we adopted LightGBM for predicting recurrence. Based on XGBoost, LightGBM employs histogram algorithm to solve the problem of excessive number of split points. This method takes up less memory and reduces computation time. Secondly, it introduces the GOSS algorithm, which extracts according to the weight information of the samples to reduce a large number of samples with small gradients, and at the same time does not change the distribution of the dataset too much. Moreover, LightGBM also proposes the EFB mode, which reduces dimensionality by bundling features. Therefore, LightGBM can improve the model accuracy while reducing the computational effort (37), which leads to its better performance in the prognosis prediction task. In the future, it is necessary to further validate the applicability of the proposed method on larger datasets.

It should be emphasized that this study aimed to predict the postoperative recurrence risk of patients only through preoperative factors, including preoperative imaging examination and clinical indicators detection. Because only in this way can it help the doctor's clinical decision-making. Although postoperative pathological examinations, such as microvascular invasion (MVI) are very meaningful for recurrence prediction (38), they were not considered in this study. The feasibility and effectiveness of this method have been demonstrated in reference (39, 40).

There are some other studies to predict the recurrence of HCC after surgery. Shen et al. (41) used the TCGA database and machine learning method to build a prediction model for recurrence of HCC patients, and optimized the recurrence prediction model. After the model was optimized, the prediction accuracy was 74.19%. Lee et al. (20) employed genetic algorithm to predict early recurrence of HCC, and extracted a total of 143 features, including 26 preoperative clinical features, 5 postoperative pathological features, and 112 imaging features. After training, the AUC of the preoperative and postoperative

TABLE 4 5-fold cross-validation results for recurrence prediction using different ensemble learning models.

Model	Acc	Recall	Prec	F_1 score
LightGBM	0.7600	0.7673	0.7733	0.7553
	± 0.0579	± 0.1311	± 0.0771	± 0.0716
CatBoost	0.6833	0.6164	0.7107	0.6511
	± 0.0543	± 0.1313	± 0.0397	± 0.0879
XGBoost	0.7224	0.6346	0.8032	0.6936
	± 0.0834	± 0.1154	± 0.1521	± 0.0978
GBDT	0.6543	0.6382	0.6600	0.6387
	± 0.0463	± 0.1328	± 0.0286	± 0.0828

Each model was evaluated employing the mean of each fold result and corresponding 95% CI.

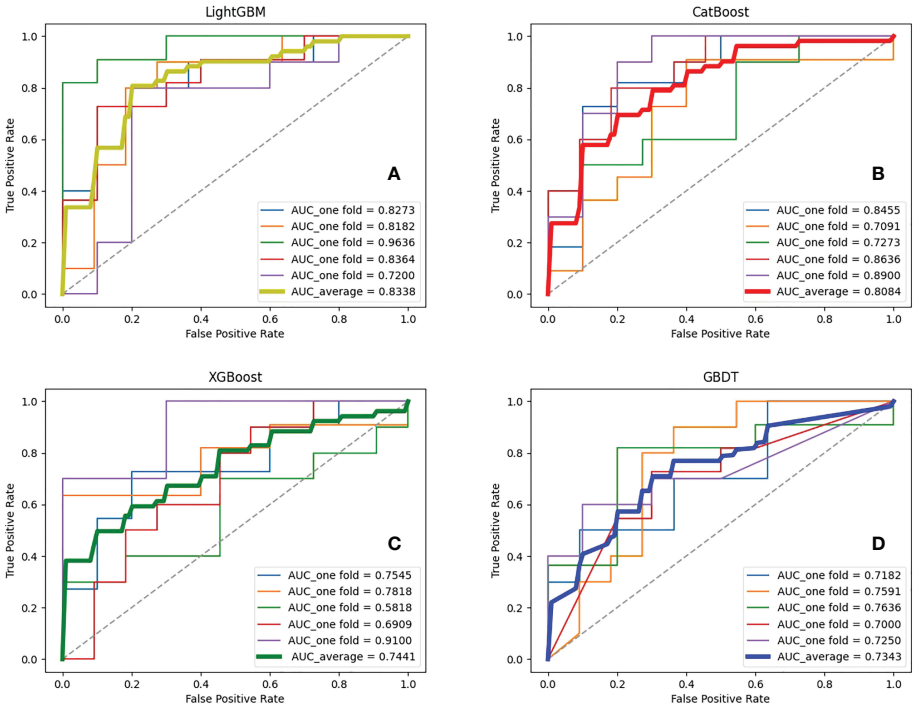


FIGURE 7
ROC curves and corresponding AUCs of various ensemble learning models. **(A–D)** represent the results of ensemble learning models - Light Gradient Boosting Machine (LightGBM), Categorical Boosting (CatBoost), eXtreme Gradient Boosting (XGBoost) and Gradient Boosting Decision Tree (GBDT), respectively.

models were 0.781 and 0.767 on the training set, and 0.739 and 0.741 on the test set, respectively. Saito et al. (42) adopted support vector machine (SVM) to predict the recurrence outcome of HCC patients based on the postoperative pathological results. The patients were grouped according to the criteria of recurrence within 1 year, 1-2 years, and 4 years after resection. The final accuracy of ROI prediction in HCC and non-HCC regions was 80.6% and 68.1%, respectively. It must be emphasized that our work only collected 105 patients, but still obtained relatively remarkable performance, suggesting that the proposed method had more potential for predicting recurrence outcomes.

It is undeniable that the present study still has some shortcomings. For example, the small sample size from a single center challenges the applicability of the models. This work only focuses on the prediction of recurrence outcomes within 3 years, and further follow-up is required to predict at different times in the future. Moreover, the proposed method has not been tested in real clinical practice, which needs to be validated in the future. Zeng et al. (43) developed a machine learning method to predict the early recurrence of radical HCC hepatectomy using the data from two centers, and the effect was relatively significant. While we have mined the key features that influence the model, the interpretability issues of machine learning still need to be addressed.

5 Conclusion

This study aims to help physicians to evaluate the effectiveness of surgery and thus facilitate rational clinical decision-making. An ensemble learning strategy based on efficient feature representation was proposed for the recurrence outcome in HCC patients within three years after surgery. The 3D U-Net was used to automatically segment the lesions. Radiomics features during AP and clinical baseline features were selected as input and four ensemble models were trained. The results showed that LighGBM outperformed other ensemble algorithms, suggesting that it may be a novel model for predicting recurrence. In the future, the dataset will be expanded for early and late recurrence prediction and external clinical validation will be performed to validate the applicability of the method. When the generalization ability of the method is successfully verified, the relevant software (or web program) will be designed and applied to clinical practice.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: Since the dataset involved in this study involves

patient privacy and has signed a non-disclosure agreement, it cannot be made public. Requests to access these datasets should be directed to Jiahong Dong, dongjiahong@mail.tsinghua.edu.cn.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

LW: Study concept, image preprocessing, experimental design, data analysis, writing of manuscript. MW: Experimental design, editing the manuscript, and data collection. CZ: Data analysis and data collection. RL: Data collection. SB: Experimental design. SY and JD: Study concept and funding. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by National Natural Science Foundation of China (grant number: 82090052, 82090050, 81930119); CAMS Innovation Fund for Medical Sciences (grant number: 2019-I2M-5-056).

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.1019009/full#supplementary-material

References

1. Chen E, Xu X, Liu R, Liu T. Small but heavy role: MicroRNAs in hepatocellular carcinoma progression. *BioMed Res Int* (2018) 2018:6784607. doi: 10.1155/2018/6784607

2. Lafaro KJ, Demirjian AN, Pawlik TM. Epidemiology of hepatocellular carcinoma. *Surg Oncol Clin N Am* (2015) 24(1):1–17. doi: 10.1016/j.soc.2014.09.001

3. Huang TE, Deng YN, Hsu JL, Leu W-J, Marchesi E, Capobianco ML, et al. Evaluation of the anticancer activity of a bile acid-dihydroartemisinin hybrid ursodeoxycholic-dihydroartemisinin in hepatocellular carcinoma cells. *Front Pharmacol* (2020) 11:599067. doi: 10.3389/fphar.2020.599067

4. Feng J, Dai W, Mao Y, Wu L, Li J, Chen K, et al. Simvastatin re-sensitizes hepatocellular carcinoma cells to sorafenib by inhibiting HIF-1α/PPAR-γ/PKM2-mediated glycolysis. *J Exp Clin Cancer Res* (2020) 39(1):24. doi: 10.1186/s13046-020-1528-x

5. Choi J, Jo C, Lim YS. Tenofovir versus entecavir on recurrence of hepatitis b virus-related hepatocellular carcinoma after surgical resection. *Hepatology* (2021) 73(2):661–73. doi: 10.1002/hep.31289

6. Lee HA, Lee YS, Kim BK, Jung YK, Kim SU, Park JY, et al. Change in the recurrence pattern and predictors over time after complete cure of hepatocellular carcinoma. *Gut Liver* (2021) 15(3):420–9. doi: 10.5009/gnl20101

7. Rattanasupar A, Chartleeraha S, Akarapatima K, Chang A. Factors that affect the surveillance and late-stage detection of a newly diagnosed hepatocellular carcinoma. *Asian Pac J Cancer Prev* (2021) 22(10):3293–8. doi: 10.31557/APJCP.2021.22.10.3293

8. Loi M, Comito T, Franzese C, Dominici L, Franceschini D, Mancosu P, et al. Stereotactic body radiotherapy in hepatocellular carcinoma: patient selection and predictors of outcome and toxicity. *J Cancer Res Clin Oncol* (2021) 147(3):927–36. doi: 10.1007/s00432-020-03389-2

9. Zhang W, Zhang B, Chen XP. Adjuvant treatment strategy after curative resection for hepatocellular carcinoma. *Front Med* (2021) 15(2):155–69. doi: 10.1007/s11684-021-0848-3

10. Gentile D, Donadon M, Lleo A, Aghemo A, Roncalli M, di Tommaso L, et al. Surgical treatment of hepatocholangiocarcinoma: A systematic review. *Liver Cancer* (2020) 9(1):15–27. doi: 10.1159/000503719

11. Jia J, Zhang J, Shao Q, Wang Y, Qian B, Hu T, et al. Efficacy of surgical treatment on different sizes of hepatitis b virus-related hepatocellular carcinoma and prognostic analysis. *J BUON* (2020) 25(4):1866–74.

12. Yoon YI, Kim KH, Cho HD, Kwon JH, Jung DH, Park GC, et al. Long-term perioperative outcomes of pure laparoscopic liver resection versus open liver resection for hepatocellular carcinoma: a retrospective study. *Surg Endosc* (2020) 34(2):796–805. doi: 10.1007/s00464-019-06831-w

13. Beumer BR, Takagi K, Vervoort B, Buettner S, Umeda Y, Yagi T, et al. Prediction of early recurrence after surgery for liver tumor (ERASL): An international validation of the ERASL risk models. *Ann Surg Oncol* (2021) 28 (13):8211–20. doi: 10.1245/s10434-021-10235-3

14. Lee IC, Lei HJ, Chau GY, Yeh YC, Wu CJ, Su CW, et al. Predictors of long-term recurrence and survival after resection of HBV-related hepatocellular carcinoma: the role of HBsAg. *Am J Cancer Res* (2021) 11(7):3711–25.

15. He Q, Jiang JJ, Jiang YX, Wang WT, Yang L. Liver surgery group. health-related quality of life comparisons after radical therapy for early-stage hepatocellular carcinoma. *Transplant Proc* (2018) 50(5):1470–4. doi: 10.1016/j.transproceed.2018.04.041

16. Harding-Theobald E, Louissaint J, Maraj B, Cuaresma E, Townsend W, Mendiratta-Lala M, et al. Systematic review: radiomics for the diagnosis and prognosis of hepatocellular carcinoma. *Aliment Pharmacol Ther* (2021) 54 (7):890–901. doi: 10.1111/apt.16563

17. El Jabbour T, Lagana SM, Lee H. Update on hepatocellular carcinoma: Pathologists' review. *World J Gastroenterol* (2019) 25(14):1653–65. doi: 10.3748/wjg.v25.i14.1653

18. Ji GW, Zhu FP, Xu Q, Wang K, Wu MY, Tang WW, et al. Machine-learning analysis of contrast-enhanced CT radiomics predicts recurrence of hepatocellular carcinoma after resection: A multi-institutional study. *EBioMedicine* (2019) 50:156–65. doi: 10.1016/j.ebiom.2019.10.057

19. Huang Y, Chen H, Zeng Y, Liu Z, Ma H, Liu J. Development and validation of a machine learning prognostic model for hepatocellular carcinoma recurrence after surgical resection. *Front Oncol* (2021) 10:593741. doi: 10.3389/fonc.2020.593741

20. Lee IC, Huang JY, Chen TC, Yen CH, Chiu NC, Hwang HE, et al. Evolutionary learning-derived clinical-radiomic models for predicting early recurrence of hepatocellular carcinoma after resection. *Liver Cancer* (2021) 10 (6):572–82. doi: 10.1159/000518728

21. Chaitanya K, Karani N, Baumgartner CF, Erdil E, Becker A, Donati O, et al. Semi-supervised task-driven data augmentation for medical image segmentation. *Med Image Anal* (2021) 68:101934. doi: 10.1016/j.media.2020.101934

22. Kim J, Kim Y, Lee EK, Chae CK, Lee K, Kim WJ, et al. Rotational variance-based data augmentation in 3D graph convolutional network. *Chem Asian J* (2021) 16(18):2610–3. doi: 10.1002/asia.202100789

23. Hsu LM, Wang S, Walton L, Wang TW, Lee SH, Shih YI. 3D U-net improves automatic brain extraction for isotropic rat brain magnetic resonance imaging data. *Front Neurosci* (2021) 15:801008. doi: 10.3389/fnins.2021.801008

24. El Khoury K, Fockedey M, Brion E, Macq B. Improved 3D U-net robustness against JPEG 2000 compression for male pelvic organ segmentation in radiotherapy. *J Med Imaging (Bellingham)* (2021) 8(4):041207. doi: 10.1117/1.JMI.8.4.041207

25. Yan J, Xu Y, Cheng Q, Jiang S, Wang Q, Xiao Y, et al. LightGBM: accelerated genomically designed crop breeding through ensemble learning. *Genome Biol* (2021) 22(1):271. doi: 10.1186/s13059-021-02492-y

26. Zhang C, Lei X, Liu L. Predicting metabolite-disease associations based on LightGBM model. *Front Genet* (2021) 12:660275. doi: 10.3389/fgene.2021.660275

27. Rufo DD, Debelee TG, Ibenthal A, Negera WG. Diagnosis of diabetes mellitus using gradient boosting machine (LightGBM). *Diagnostics (Basel)* (2021) 11(9):1714. doi: 10.3390/diagnostics11091714

28. Zheng C, Tian J, Wang K, Han L, Yang H, Ren J, et al. Time-to-event prediction analysis of patients with chronic heart failure comorbid with atrial fibrillation: a LightGBM model. *BMC Cardiovasc Disord* (2021) 21(1):379. doi: 10.1186/s12872-021-02188-y

29. Ambe K, Suzuki M, Ashikaga T, Tohkin M. Development of quantitative model of a local lymph node assay for evaluating skin sensitization potency applying machine learning CatBoost. *Regul Toxicol Pharmacol* (2021) 125:105019. doi: 10.1016/j.yrtph.2021.105019

30. Zhao QY, Wang H, Luo JC, Lou MH, Liu LP, Yu SJ, et al. Development and validation of a machine-learning model for prediction of extubation failure in intensive care units. *Front Med (Lausanne)* (2021) 8:676343. doi: 10.3389/fmed.2021.676343

31. Hou N, Li M, He L, Xie B, Wang L, Zhang R, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. *J Transl Med* (2020) 18(1):462. doi: 10.1186/s12967-020-02620-5

32. Davagdorj K, Pham VH, Theera-Umpon N, Ryu KH. XGBoost-based framework for smoking-induced noncommunicable disease prediction. *Int J Environ Res Public Health* (2020) 17(18):6513. doi: 10.3390/ijerph17186513

33. Zhang Y, Kuang S, Shan Q, Rong D, Zhang Z, Yang H, et al. Can IVIM help predict HCC recurrence after hepatectomy? *Eur Radiol* (2019) 29(11):5791–803. doi: 10.1007/s00330-019-06180-1

34. Tang P, Liang Q, Yan X, Xiang S, Sun W, Zhang D, et al. Efficient skin lesion segmentation using separable-unet with stochastic weight averaging. *Comput Methods Programs Biomed* (2019) 178:289–301. doi: 10.1016/j.cmpb.2019.07.005

35. Feng B, Ma XH, Wang S, Cai W, Liu XB, Zhao XM. Application of artificial intelligence in preoperative imaging of hepatocellular carcinoma: Current status and future perspectives. *World J Gastroenterol* (2021) 27(32):5341–50. doi: 10.3748/wjg.v27.i32.5341

36. Lewis S, Hectors S, Taouli B. Radiomics of hepatocellular carcinoma. *Abdom Radiol (NY)* (2021) 46(1):111–23. doi: 10.1007/s00261-019-02378-5

37. Zhu J, Su Y, Liu Z, Liu B, Sun Y, Gao W, et al. Real-time biomechanical modelling of the liver using LightGBM model. *Int J Med Robot* (2022) 18:e2433. doi: 10.1002/rcs.2433

38. Yanhan W, Lianfang L, Hao L, Yunfeng D, Nannan S, Fanfan L, et al. Effect of microvascular invasion on the prognosis in hepatocellular carcinoma and analysis of related risk factors: A two-center study. *Front Surg* (2021) 8:733343. doi: 10.3389/fsurg.2021.733343

39. Wei H, Jiang H, Qin Y, Wu Y, Lee JM, Yuan F, et al. Comparison of a preoperative MR-based recurrence risk score versus the postoperative score and four clinical staging systems in hepatocellular carcinoma: a retrospective cohort study. *Eur Radiol* (2022). doi: 10.1007/s00330-022-08811-6

40. Ding DY, Liu L, Li HL, Gan XJ, Ding WB, Gu FM, et al. Development of preoperative prognostic models including radiological features for survival of singular nodular HCC patients. *Hepatobiliary Pancreat Dis Int* (2022) S1499-3872(22):00052–2. doi: 10.1016/j.hbpd.2022.04.002

41. Shen J, Qi L, Zou Z, Du J, Kong W, Zhao L, et al. Identification of a novel gene signature for the prediction of recurrence in HCC patients by machine learning of genome-wide databases. *Sci Rep* (2020) 10(1):4435. doi: 10.1038/s41598-020-61298-3

42. Saito A, Toyoda H, Kobayashi M, Yoshinori K, Fujii H, Fojita K, et al. Prediction of early recurrence of hepatocellular carcinoma after resection using digital pathology images assessed by machine learning. *Mod Pathol* (2021) 34 (2):417–25. doi: 10.1038/s41379-020-00671-z

43. Zeng J, Zeng J, Lin K, Lin H, Wu Q, Guo P, et al. Development of a machine learning model to predict early recurrence for hepatocellular carcinoma after curative resection. *Hepatobiliary Surg Nutr* (2022) 11(2):176–87. doi: 10.21037/hbsn-20-466

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fonc.2022.1049627

OPEN ACCESS

EDITED BY
Alireza Mansouri,
The Pennsylvania State University (PSU), United States

REVIEWED BY
Mohammed A. Azab,
Boise State University, United States
Julius Höhne,
University Medical centre Regensburg, Germany

*CORRESPONDENCE
Hani J. Marcus
h.marcus@ucl.ac.uk

SPECIALTY SECTION
This article was submitted to Neuro-Oncology and Neurosurgical Oncology, a section of the journal Frontiers in Oncology

RECEIVED 20 September 2022
ACCEPTED 24 October 2022
PUBLISHED 04 January 2023

CITATION
CRANIAL Consortium (2023) CSF rhinorrhoea after endonasal intervention to the skull base (CRANIAL): A multicentre prospective observational study.
Front. Oncol. 12:1049627.
doi: 10.3389/fonc.2022.1049627

CSF rhinorrhoea after endonasal intervention to the skull base (CRANIAL): A multicentre prospective observational study

CRANIAL Consortium

Objective: Despite progress in endonasal skull-base neurosurgery, cerebrospinal fluid (CSF) rhinorrhoea remains common and significant. The CRANIAL study sought to determine 1) the scope of skull-base repair methods used, and 2) corresponding rates of postoperative CSF rhinorrhoea in the endonasal transsphenoidal approach (TSA) and the expanded endonasal approach (EEA) for skull-base tumors.

Methods: A prospective observational cohort study of 30 centres performing endonasal skull-base neurosurgery in the UK and Ireland (representing 91% of adult units). Patients were identified for 6 months and followed up for 6 months. Data collection and analysis was guided by our published protocol and pilot studies. Descriptive statistics, univariate and multivariable logistic regression models were used for analysis.

Results: A total of 866 patients were included - 726 TSA (84%) and 140 EEA (16%). There was significant heterogeneity in repair protocols across centres. In TSA cases, nasal packing (519/726, 72%), tissue glues (474/726, 65%) and hemostatic agents (439/726, 61%) were the most common skull base repair techniques. Comparatively, pedicled flaps (90/140, 64%), CSF diversion (38/140, 27%), buttresses (17/140, 12%) and gasket sealing (11/140, 9%) were more commonly used in EEA cases. CSF rhinorrhoea (biochemically confirmed or requiring re-operation) occurred in 3.9% of TSA (28/726) and 7.1% of EEA (10/140) cases. A significant number of patients with CSF rhinorrhoea (15/38, 39%) occurred when no intraoperative CSF leak was reported. On multivariate analysis, there may be marginal benefits with using tissue glues in TSA (OR: 0.2, CI: 0.1-0.7, p<0.01), but no other technique reached significance. There was evidence that certain characteristics make CSF rhinorrhoea more likely – such as previous endonasal surgery and the presence of intraoperative CSF leak.

Conclusions: There is a wide range of skull base repair techniques used across centres. Overall, CSF rhinorrhoea rates across the UK and Ireland are lower than generally reported in the literature. A large proportion of postoperative leaks occurred in the context of occult intraoperative CSF leaks, and decisions for universal sellar repairs should consider the risks and cost-effectiveness of repair

strategies. Future work could include longer-term, higher-volume studies, such as a registry; and high-quality interventional studies.

KEYWORDS

cerebrospinal fluid rhinorrhoea, CSF, EEA, endoscopic endonasal, cerebrospinal fluid leak, skull base surgery

Introduction

Endonasal approaches have revolutionized skull-base neurosurgery (1, 2). The most commonly utilized approach is the transsphenoidal approach (TSA), frequently used for sellar lesions. More recently, the development of the expanded endonasal approach (EEA) has allowed access to pathologies extending beyond the sella, with growing indications as this technique evolves (3, 4).

An international expert consensus on TSA workflow highlighted the potential for practice variations, particularly in closure, due to a variety of skull-base repair options (5). Previous systematic reviews examining skull-base repair techniques across endonasal skull-base neurosurgery found absolute heterogeneity across studies and centres, likely due to a paucity of high-level comparative evidence (6). Similarly, there is variance in postoperative cerebrospinal fluid (CSF) rhinorrhoea rates, one of the commonest postoperative complications – generally up to 5% in TSA and 20% in EEA (4, 7–12). CSF rhinorrhoea has potentially serious consequences including pneumocephalus, meningitis, and prolonged hospital admission or re-admission (9, 13, 14).

CRANIAL (CSF Rhinorrhoea After Endonasal Intervention to the Skull Base) was a prospective, multicentre observational study seeking to determine the: (1) scope of the methods of skull-base repair; and (2) corresponding rates of postoperative CSF rhinorrhoea in the UK and Ireland (15–17). CRANIAL was a collaboration between three bodies: students and junior doctors *via* NANSIG (The Neurology and Neurosurgery Interest Group), neurosurgical trainees *via* BNTRC (British Neurosurgical Trainee Research Collaborative) and skull-base consultants (neurosurgery and otorhinolaryngology) *via* the CRANIAL Steering Committee.

After piloting at 12 centres, preliminary results suggested practice heterogeneity (15, 16). Thus, the study was expanded UK and Ireland wide, and herein, we present the results.

Abbreviations: BMI, Body mass index; CRANIAL, CSF Rhinorrhoea After Endonasal Intervention to the Skull Base; CSF, Cerebrospinal fluid; EEA, Expanded endonasal approach; TSA, Transsphenoidal approach; CT, Computed Tomography; VPS, Ventriculoperitoneal shunt.

Methods

The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement guided this methodology and report (18).

Study design

A multicentre, prospective, observational cohort study design was conducted across tertiary neurosurgical units with 2 pilot phases (Phase 1, 4 centres, 01/11/2019-22/03/2020; Phase 2, 12 centres, 23/03/2020-31/07/2020) and a full study period (15–17). The full study included 30 centres, representing 91% (29/32, of adult neurosurgical centres performing endonasal skull-base neurosurgery in the UK and Ireland). One pediatric centre was included, whilst others provided both adult and pediatric services. The study period included 6 months of consecutive case recruitment (10/08/20–10/02/21) and 6 months of follow-up (10/02/21–10/08/21) (19).

Cases included patients of all ages undergoing TSA for sellar tumors and EEA for skull base tumors (17). TSA was defined as surgical access to the sella alone (transsphenoidal) whilst EEA was defined as acquiring surgical access to an area not limited to the sella (e.g., transplanum or transcribriform) (17, 20). Exclusion criteria were patients undergoing transcranial surgery and those with preoperative CSF rhinorrhoea.

Data collection

Each centre registered the project as a service evaluation with appropriate approvals. Following the BNTRC model (21), the local team consisted of consultant lead(s) with overall project responsibility, with trainee lead(s) and student lead(s) for data collection *via* a secure web-based central database (Castor Electronic Data Capture). NANSIG and the BNTRC provided project support, overseen by the CRANIAL consultant steering committee.

Data were collected as per protocol (15–17). The Esposito-Kelly system graded intraoperative CSF leak if present (22).

Local teams aimed to collect data within 30 days of operation for admission data, and at the end of the 6-month follow-up window for follow-up data (17). Primary outcomes were: (1) methods of intraoperative skull-base reconstruction, and (2) postoperative CSF rhinorrhoea biochemically confirmed or requiring intervention (CSF diversion and/or operative repair) (17).

Data validation

Data were confirmed with operating surgeons or senior team members before final submission. An independent local data validator screened a random 10% of submitted cases at each centre. The primary validation target was >95% accuracy across audited data (17). Finally, each local team reviewed their final validated dataset before analysis.

Data analysis

Pre-processing included re-categorizing free-text entries. Descriptive statistics summarized baseline characteristics (demographic, tumour, and operative characteristics) and surgical outcomes, using Microsoft Excel (Version 16.54). The incidence density of repair methods and combinations within TSA/EEA and CSF leak grade subgroups were calculated. Corresponding postoperative CSF rhinorrhoea rates were summarized as incidence percentages per TSA/EEA subgroups and repair method used. Univariate and multivariable logistic regression models assessed the impact of baseline characteristics (from the literature) on skull-base repair methods, and the influence of baseline characteristics and skull-base repair methods on CSF rhinorrhoea incidence, with odds ratios and 95% confidence intervals reported (Stata, Version 16, StataCorp, USA) (17). Fisher's exact test was used to compare repair methods used with and without intraoperative CSF leak.

Results

866 patients (726 TSA, 140 EEA) were included across 30 centres. All centres completed data validation, with >95% record accuracy in audited cases and no duplicates included.

Patient characteristics

The median patient age was 53 years (range: 5–84), 23% (198/866) were older than 65. There were 416 male patients and 450 female patients; 238 (TSA: 210/726; EEA: 28/140) patients were obese (body mass index >30) (Tables 1, 2). Pre-operative visual deficits (acuity and/or field) were present in 464 patients (TSA: 374/726; EEA: 91/140); 6 were blind with binocular <6/60 acuity (TSA: 9/374; EEA: 3/91) (Supplementary Material 6). Pre-operative anterior hypopituitarism (requiring hydrocortisone supplementation) was present in 215 cases (TSA: 184/726; EEA: 31/140), and posterior hypopituitarism (requiring desmopressin supplementation) in 36 cases (TSA: 28/726; EEA: 8/140). The commonest TSA pathologies were non-functioning pituitary adenoma (410/726), functioning pituitary adenoma (249/726), and Rathke's cleft cyst (26/726) (Supplementary Material 3). For EEA, craniopharyngioma (38/140), meningioma (25/140) and non-functioning pituitary adenoma (23/140) were the commonest. Most tumors were >1cm in maximum diameter (TSA: 607/726; EEA: 131/140).

Operation characteristics

Of TSA cases, endoscopic was most prevalent (615/726), followed by microscopic (80/726), and a combined approach (32/726) method. Revision surgery was infrequent (TSA 98/726; EEA 21/140). On multivariate logistic regression, TSA was less likely to be used for larger tumors (maximum diameter >1cm) compared to EEA, aligning with indications for these approaches (OR: 0.4, CI: 0.2-0.9, p=0.03). Most TSA surgeries were performed by neurosurgeons alone (458/726), whereas most EEA cases were performed with both neurosurgery and otorhinolaryngology specialists (90/140). Infrequently cases were performed by otorhinolaryngologists alone (TSA: 22/726; EEA: 3/140). The median operation duration was 110 minutes for TSA (range: 29–540 minutes) and 220 minutes for EEA (range: 30–795 minutes).

Intraoperative CSF leak was reported in 214 TSA cases (214/726) and 79 EEA cases (79/140). Intraoperative CSF leaks were most commonly low-flow in TSA (131/214 grade 1) and high-flow in EEA (39/79 grade 3) (Tables 1, 2).

Skull-base reconstruction overview

A taxonomy for skull-base repair was adapted from a systematic review of the literature (Supplementary Material 2) (20, 21). Heterogeneity of repair technique choice across both approaches was evident (Figures 1, 2).

In TSA, the commonest techniques were nasal packing (519/726), tissue glues (474/726) and hemostatic agents (439/726) (Table 1; Supplementary Material 4). The most prevalent nasal packing was Nasopore® (369/519), Merocel® (94/519) and Rapid Rhinos® (33/519). Tissue glues most frequently used were Adherus® (146/474), Duraseal® (137/474) and Tisseel® (126/474); whilst common hemostatic agents included Surgicel® (189/439), Surgiflo® (141/439) and Floseal® (91/439). Tissue

TABLE 1 Incidence of repair technique categories across surgical approaches, intraoperative CSF leak presence/grade, tumour diameter, BMI and age.

Category	Dural closure	Dural replacement	Tissue graft	Synthetic graft	Button technique	Pedicled flap	Tissue glue	Haemostatic agent	Buttress	Gasket sealing	Nasal packing	CSF diversion	CSF rhinorrhoea
Approach													
TSA (N = 726), n (n/N%)	0 (0%)	196 (27%)	221 (30.4%)	204 (28.1%)	20 (2.8%)	116 (16%)	474 (65.3%)	439 (60.5%)	31 (4.3%)	15 (2.1%)	519 (71.5%)	29 (4%)	28 (3.9%)
EEA (N = 140), n (n/N%)	0 (0%)	66 (47.1%)	65 (46.4%)	47 (33.6%)	7 (5%)	90 (64.3%)	114 (81.4%)	93 (66.4%)	17 (12.1%)	11 (7.9%)	116 (82.9%)	38 (27.1%)	10 (7.1%)
Intraoperative CSF leak grade													
Grade 0 (N = 573), n (n/N%)	0 (0%)	136 (23.7%)	106 (18.5%)	163 (28.4%)	9 (1.6%)	88 (15.4%)	335 (58.5%)	358 (62.5%)	19 (3.3%)	11 (1.9%)	403 (70.3%)	19 (3.3%)	15 (2.6%)
Grade 1 (N = 143), n (n/N%)	0 (0%)	54 (37.8%)	89 (62.2%)	45 (31.5%)	7 (4.9%)	37 (25.9%)	124 (86.7%)	82 (57.3%)	7 (4.9%)	3 (2.1%)	114 (79.7%)	13 (9.1%)	4 (2.8%)
Grade 2 (N = 67), n (n/N%)	0 (0%)	27 (40.3%)	41 (61.2%)	18 (26.9%)	7 (10.4%)	33 (49.3%)	55 (82.1%)	33 (49.3%)	10 (14.9%)	4 (6%)	52 (77.6%)	8 (11.9%)	10 (14.9%)
Grade 3 (N = 44), n (n/N%)	0 (0%)	23 (52.3%)	33 (75%)	15 (34.1%)	3 (6.8%)	30 (68.2%)	44 (100%)	28 (63.6%)	9 (20.5%)	6 (13.6%)	31 (70.5%)	16 (36.4%)	2 (4.5%)
Grade unknown (N = 39), n (n/N%)	0 (0%)	22 (56.4%)	17 (43.6%)	10 (25.6%)	1 (2.6%)	18 (46.2%)	30 (76.9%)	31 (79.5%)	1 (2.6%)	2 (5.1%)	18 (46.2%)	46.2 (30%)	7 (17.9%)
Specialty													
Neurosurgery only (N=505), n (n/N%)	0 (0%)	154 (30.5%)	219 (43.4%)	164 (32.5%)	24 (4.8%)	63 (12.5%)	361 (71.5%)	274 (54.3%)	33 (6.5%)	21 (4.2%)	297 (58.8%)	40 (7.9%)	21 (4.2%)
Otorhinolaryngology only (N=25), n (n/N%)	0 (0%)	17 (68%)	2 (8%)	14 (56%)	0 (0%)	5 (20%)	25 (100%)	25 (100%)	0 (0%)	0 (0%)	25 (100%)	0 (0%)	4 (16%)
Multidisciplinary (N=336), n (n/N%)	0 (0%)	91 (27.1%)	65 (19.3%)	73 (21.7%)	3 (0.9%)	138 (41.1%)	202 (60.1%)	233 (69.3%)	15 (4.5%)	5 (1.5%)	313 (93.2%)	27 (8%)	13 (3.9%)
Tumour diameter													
>1cm (N=738), n (n/N%)	0 (0%)	238 (32.2%)	243 (32.9%)	218 (29.5%)	26 (3.5%)	190 (25.7%)	510 (69.1%)	456 (61.8%)	44 (6%)	24 (3.3%)	546 (74%)	61 (8.3%)	31 (4.2%)
<1cm (N=128), n (n/N%)	0 (0%)	24 (18.8%)	43 (33.6%)	33 (25.8%)	1 (0.8%)	16 (12.5%)	78 (60.9%)	76 (59.4%)	4 (3.1%)	2 (1.6%)	89 (69.5%)	6 (4.7%)	7 (5.5%)
BMI													
<30 (N=628), n (n/N%)	0 (0%)	190 (30.3%)	211 (33.6%)	181 (28.8%)	20 (3.2%)	148 (23.6%)	416 (66.2%)	378 (60.2%)	41 (6.5%)	24 (3.8%)	456 (72.6%)	51 (8.1%)	25 (4%)
>30 (N=238), n (n/N%)	0 (0%)	72 (30.3%)	75 (31.5%)	70 (29.4%)	7 (2.9%)	58 (24.4%)	172 (72.3%)	154 (64.7%)	7 (2.9%)	2 (0.8%)	179 (75.2%)	16 (6.7%)	13 (5.5%)
Age													
<65 (N=668), n (n/N%)	0 (0%)	201 (30.1%)	216 (32.3%)	197 (29.5%)	19 (2.8%)	168 (25.1%)	462 (69.2%)	419 (62.7%)	35 (5.2%)	17 (2.5%)	493 (73.8%)	54 (8.1%)	35 (5.2%)
>65 (N=198), n (n/N%)	0 (0%)	61 (30.8%)	70 (35.4%)	54 (27.3%)	8 (4%)	38 (19.2%)	126 (63.6%)	113 (57.1%)	13 (6.6%)	9 (4.5%)	142 (71.7%)	13 (6.6%)	3 (1.5%)

CSF, cerebrospinal fluid; BMI, body mass index.

10.3389/fonc.2022.1049627

TABLE 2 Summary of CSF rhinorrhoea incidence per baseline and operative risk factor subgroups – incidence and statistical analysis *via* multivariate logistic regression.

	Transsphenoidal approach		*Expanded Endonasal Approach*	
	CSF Rhinorrhoea rate	**Multivariate Analyses (OR, CI, p-value)**	**CSF Rhinorrhoea rate**	**Multivariate Analyses (OR, CI, p-value)**
Approach				
TSA	28/726 (3.9%)	–	–	–
EEA	–	–	10/140 (7.1%)	–
Baseline characteristics				
Age >65	0/172 (0.0%)	–	3/27 (11.1%)	OR: 3.8, CI: 0.6–23.7, p =0.16
Age <65	28/553 (5.1%)	Reference	7/113 (6.2%)	Reference
BMI >30	11/210 (5.2%)	OR: 1.7, CI: 0.7-4.4, p=0.26	2/28 (7.1%)	OR: 0.7, CI: 0.1-6.1, p=0.7
BMI<30	17/516 (3.3%)	Reference	8/112 (7.1%)	Reference
Tumour diameter >1cm	21/607 (3.5%)	OR:0.5; CI: 0.2 – 1.5, p=0.22	10/131 (7.6%)	–
Tumour diameter <1cm	7/119 (6.0%)	Reference	0/9 (0%)	Reference
Primary surgery	8/98 (8.2%)	OR:0.4, CI: 0.1-0.9, p=0.05	1/21 (4.8%)	OR: 0.6, CI; 0.1-8.4, p=0.71
Revision surgery	19/573 (3.3%)	Reference	7/113 (6.2%)	Reference
Presence of Otorhinolaryngologist	9/268 (3.4%)	OR: 0.4, CI: 0.1-1.6, p=0.2	8/93 (8.6%)	OR: 0.6, CI: 0.1-7.4, p=0.72
Presence of Neurosurgeon	25/704 (3.6%)	OR: 0.2, CI: 0.1-1.9, p=0.17	9/137 (6.6%)	OR: 0.1, CI: 0-1.8, p=0.1
Intra-operative CSF leak grade				
Grade 0	11/512 (2.1%)	Reference	4/61 (6.6%)	Reference
Grade 1	3/131 (2.3%)	OR: 1.5, CI: 0.4-6.6, p=0.56	1/12 (8.3%)	OR: 2.2, CI: 0.1-39.9, p= 0.61
Grade 2	9/54 (16.7%)	OR: 16.1, CI: 4.6-56.3, p<0.01	1/13 (7.7%)	OR: 1.8, CI: 0.1-24.2, p=0.67
Grade 3	0/5 (0%)	-	2/39 (5.6%)	OR: 1.2, CI: 0.1-11.5, p=0.87
Leak present, grade unknown	5/24 (20.8%)	OR: 7.6, CI: 1.8-33.4, p<0.01	2/15 (13.3%)	OR: 12, CI: 0.4-356.3, p=0.15
Repair methods				
Dural closure	–	–	–	–
Dural replacement	11/196 (5.6%)	OR: 2.6, CI: 0.8-8.8, p=0.13	5/66 (7.6%)	OR: 0.9, CI: 0.1-5.1, p=0.85
Tissue graft	13/221 (5.9%)	OR: 1.8, CI: 0.6-5.3, p=0.29	3/65 (4.6%)	OR: 0.3, CI: 0.1-2.2, p=0.21
Synthetic graft	7/204 (3.4%)	OR: 1.2, CI: 0.4-3.6, p=0.79	6/47 (12.8%)	OR: 5.2, CI: 0.7-39.1, p=0.11
Button Technique	0/20 (0%)	–	0/7 (0%)	–
Pedicled Flap	5/116 (4.3%)	OR: 0.9, CI: 0.3-3.2, p=0.87	8/90 (8.9%)	–
Tissue Glue	15/474 (3.2%)	OR: 0.2, CI: 0.1-0.7, p<0.01	8/114 (7.0%)	OR: 4.4, CI: 0.3-78.6, p=0.31
Haemostatic agent	18/439 (4.1%)	OR: 1.3, CI: 0.5-3.4, p=0.63	5/93 (5.4%)	OR: 0.3, CI: 0.1-2.5, p=0.27
Buttress	0/31 (0%)	–	1/17 (5.9%)	OR: 2.8, CI: 0.1-63.1, p=0.53
Gasket sealing	0/15 (0%)	–	0/11 (0%)	–
Nasal packing	22/519 (4.2%)	OR: 1.9, CI: 0.6-5.8, p=0.29	10/116 (8.6%)	–
CSF diversion	1/29 (3.4%)	OR: 0.9, CI: 0.1-8.3, p=0.96	1/38 (2.6%)	OR: 0.2, CI: 0-5.3, p =0.298

grafts were used in 221 cases (221/726), usually fat (189/221, most commonly abdominal), fascia (27/221, most often fascia lata) and mucosa (28/221, usually middle turbinate). Synthetic grafts (204/726) included Spongostan™ (181/204), Tachosil® (21/204) and Gelfoam® (2/204). The button technique was used with these grafts in 20 cases (20/726). There was overlap between these graft materials and dural replacement (or reconstruction *via* layering) strategies (196/726) which usually consisted of Duragen® (136/196), fascia lata (18/196) or Lyoplant® (17/196). Pedicled flaps were used in 116 cases (116/726), most frequently nasoseptal flaps (105/116). Rigid buttresses were used in 31 cases (31/726), commonly Medpor® (15/31), autologous bone (14/31, usually septal) and autologous cartilage (1/31). These buttresses were used with a gasket seal technique in 15 cases (15/726), usually with fascia lata.

Comparatively, pedicled flaps (90/140), CSF diversion (38/140), buttresses (17/140), and gasket sealing (11/140) were more commonly used in EEA cases (Table 1; Supplementary Material 4). Nasoseptal flaps (87/90) were again the most frequent pedicled flaps. Like TSA, supportive buttresses were often Medpor® (10/17) or autologous bone (5/17), the majority of these being used with the gasket seal technique (11/17).

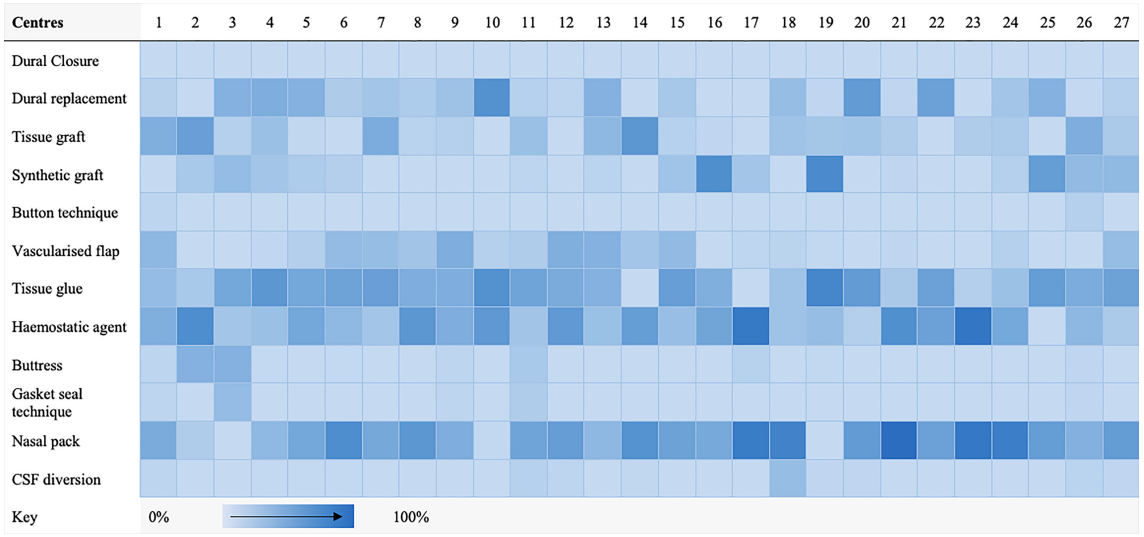


FIGURE 1
Heat map highlighting frequency of repair technique category use per centre for transsphenoidal cases.

Additionally, nasal packs (116/140), tissue glue (114/140) and hemostatic agents (93/140) were prevalent. The commonest nasal packs were Nasopore® (86/116), Merocel® (20/116) and Bismuth-Soaked Ribbon Gauze (11/116). Again, Tisseel® (32/114), Adherus® (22/114) and Duraseal® (22/114) were the most used tissue glues; whilst Surgicel® (51/93), Surgiflo® (24/93) and Floseal® (13/93) were common hemostatic agents. Tissue grafts (65/140), were frequently fat (45/65), fascia (36/65) and mucosa (8/65), akin to TSA. Similarly, synthetic grafts (47/140) included Spongostan™ (39/47) and Tachosil® (5/47). The button technique was sometimes used with these grafts (47/140). Finally, common dural replacement (66/140) strategies included Duragen® (43/66), fascia lata (12/66) and Tutoplast® (6/66).

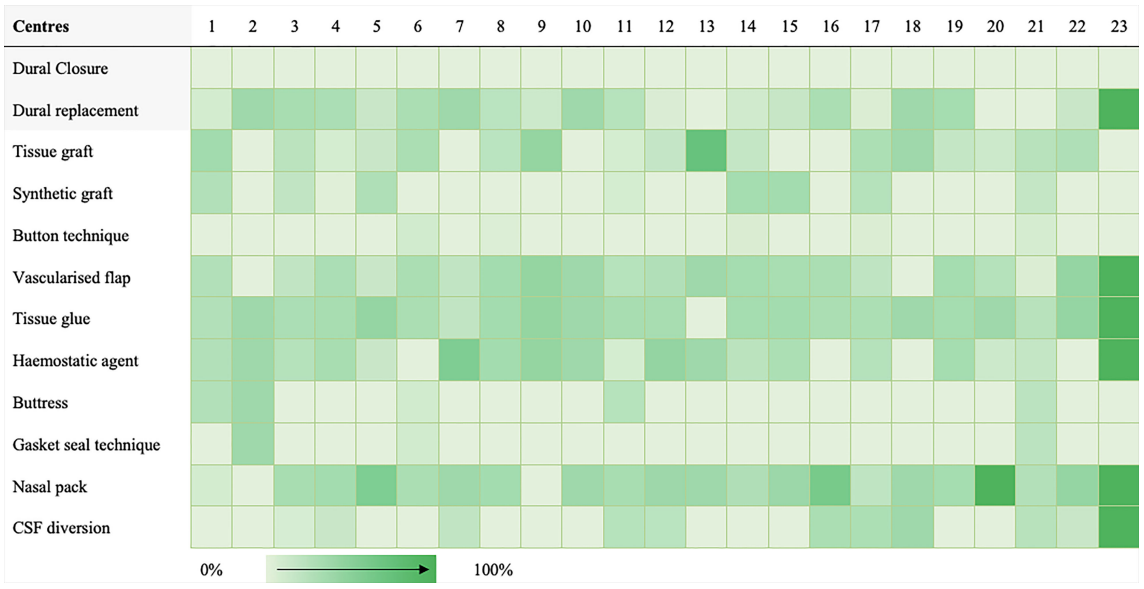


FIGURE 2
Heat map highlighting frequency of repair technique category use per centre for expanded endonasal cases.

Factors affecting repair technique choice

Repair methods appeared to be tailored according to postoperative CSF rhinorrhoea risk (Table 1 for descriptive analyses, Supplementary Material 5 for further statistical analyses). In cases with intraoperative CSF leak, there was a statistically significant (via Fisher's exact test) increased use of tissue grafts ($p<0.01$), pedicled flaps ($p<0.01$), tissue glues ($p<0.01$) and CSF diversion (TSA $p<0.01$; EEA $p<0.05$) for both TSA and EEA on univariate analysis. Additionally, dural replacements ($p<0.01$), hemostatic agents ($p=0.01$) and buttresses ($p<0.01$) were also used more in EEA (but not TSA) with intraoperative CSF leak. Similarly, the use of pedicled flaps (OR: 2.3, CI: 1.3-4.2, $p=0.01$), dural replacement (OR: 2.1, CI: 1.3-3.4, $p<0.01$) and tissue glues (OR: 1.36, CI: 1.1-2.5, $p=0.02$) were statistically associated with operations for larger tumors (maximum diameter >1cm) on multivariate logistic regression. Regarding surgical specialty, the use of pedicled flaps (OR: 4.5, CI: 3.1-6.3, $p<0.01$) and hemostatic agents (OR: 1.9, CI: 1.5-2.7, $p<0.01$) were statistically associated with otorhinolaryngology involvement, whilst the use of tissue grafts (OR: 0.3, CI: 0.2-0.5, $p<0.01$) and tissue glues (OR: 0.6, CI: 0.4-0.8, $p<0.01$) was reduced on multivariate logistic regression.

CSF diversion

67 cases used CSF diversion (TSA: 29/726; EEA: 38/140). In TSA, lumbar drainage was most common (27/29) with one of these patients subsequently requiring a ventriculoperitoneal shunt (VPS). The remainder underwent lumbar puncture (1/29), or external ventricular drain (EVD) placement (1/29). Lumbar drains were usually placed under the same anesthetic (pre-procedure, 15/29; post-procedure, 7/29), with regimes (if specified) volume-led (14/29, usually 5-10mls/hr), pressure-led (6/29) or using a LiquoGuard® system (1/29). Three drains inserted pre-procedure were removed before any effective postoperative CSF drainage (used for intraoperative saline injection or inserted prophylactically in case of subsequent CSF rhinorrhoea). Excluding these, the median length of drainage *via* lumbar drain was five days (range: 2-11).

Regarding EEA surgeries, all CSF diversion was performed *via* lumbar drain with most placed under the same anesthetic (immediately pre-procedure: 22/38; or immediately post-procedure: 8/38). The most common drainage regime was volume-led (21/22), with 5-10mls/hr the commonest protocol. One case also had an EVD placed one week before tumour resection for acute hydrocephalus. Three pre-procedure drains inserted were removed before any effective postoperative CSF drainage. Excluding these, the median length of drainage was five days (range: 1-7).

Postoperative management

The median patient hospital stay was four days (range: 1–37) for TSA and seven days (range: 1–35) for EEA. Regarding conservative measures, bed rest was advised in 20% (147/726) TSA cases (head elevated: 72/147; head flat: 5/147; unspecified height: 70/152) and 40% (52/140) EEA cases (head elevated: 37/52; head flat: 3/52; unspecified height: 12/52). Avoiding straining (e.g., lifting, sneezing, etc.) was advised in most TSA (502/726) and EEA (91/140) cases. Stool softeners were prescribed in 191 TSA cases (191/726) and 30 EEA cases (30/140). Rarely, acetazolamide (TSA: 1/726; EEA 1/140) was offered. Visual outcomes, endocrine outcomes and complications at 6 months follow-up are summarized in Supplementary Material 6.

Postoperative CSF rhinorrhoea

CSF rhinorrhoea (biochemically confirmed or requiring re-operation) occurred in 3.9% of TSA (28/726) and 7.1% of EEA (10/140) cases.

In TSA, most cases occurred during the index admission (21/28), presenting a median of 2 days postoperatively (range: 1-17), whereas those presenting during follow-up (7/28) a median of 10 days postoperatively (range: 2-84). Almost all cases were managed operatively (index: 18/21; follow-up: 6/7). Initial surgical treatment included lumbar drain alone (8/24), lumbar drain & direct endonasal repair (8/24), direct endonasal repair alone (6/24), or VPS alone (2/24). Five cases required further operations for recurrent CSF rhinorrhoea. Regarding EEA, CSF rhinorrhoea occurred during the index admission for 8 cases, and during follow-up for 2 cases. All cases were managed operatively (lumbar drain & endonasal repair: 6/10; lumbar drain alone 3/10; endonasal repair alone: 1/10). Two cases required further operations for recurrent CSF rhinorrhoea. Cases presenting during index admission were detected at a median of 2 days postoperatively (range: 1-11), whilst those detected during follow-up were found at a median of 19 days postoperatively (range: 8-54).

On univariate logistic regression analysis, displayed in Figure 3, the following variables were associated with CSF rhinorrhoea: revision surgery (TSA), presence of intraoperative CSF leak (TSA), and the absence of neurosurgery involvement (TSA) (Table 2; Figure 3; Supplementary Material 5). On multivariate analysis, revision surgery and the presence of intraoperative CSF leak remained a predictor of CSF rhinorrhoea in TSA (Table 2; Figure 3; Supplementary Material 5). No specific technique category (including CSF diversion) considerably impacted the odds of CSF rhinorrhoea for EEA. However, tissue glues in TSA (OR: 0.2, CI: 0.1-0.7, $p<0.01$) may be related to a slight decrease in CSF rhinorrhoea

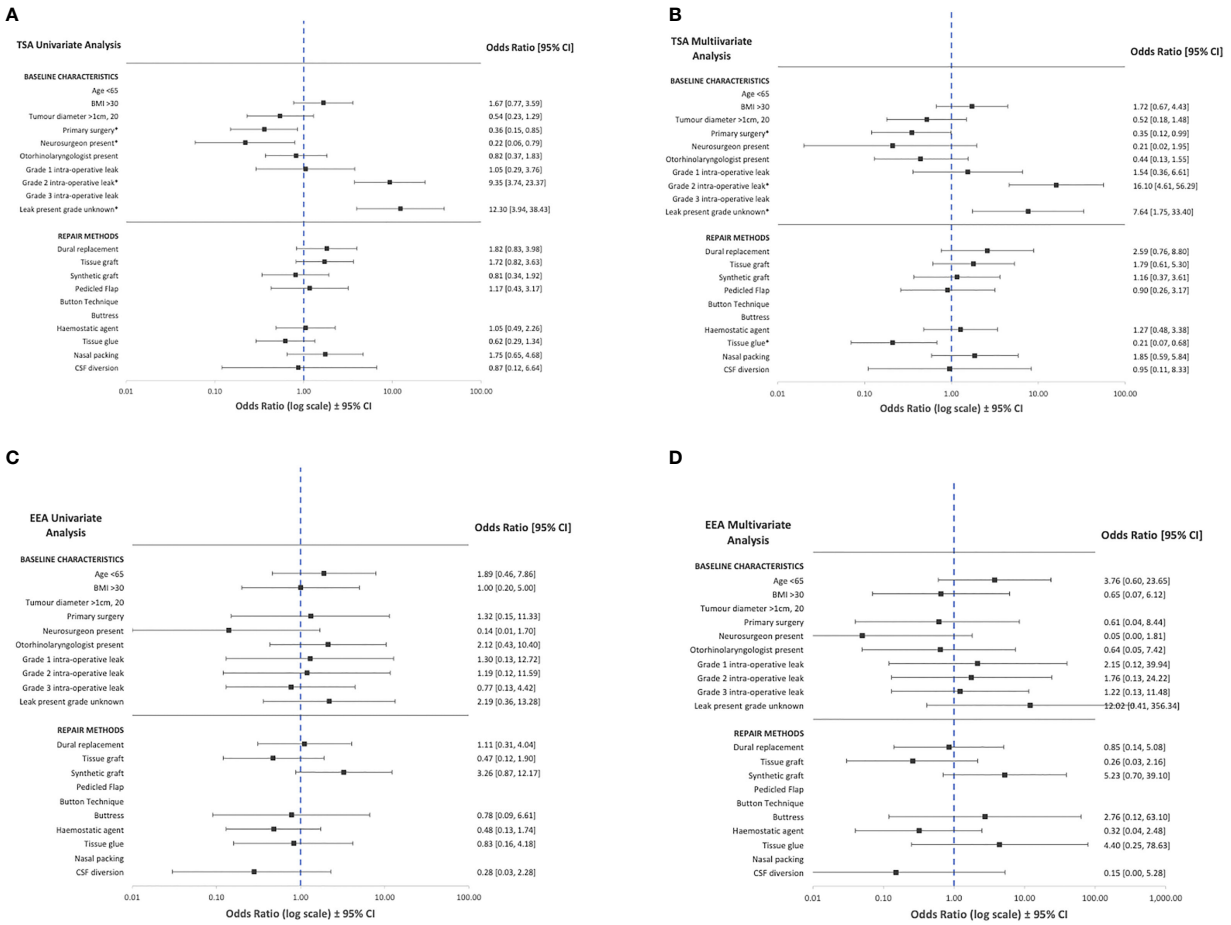


FIGURE 3
Summary of univariate and multivariate logistic regression of baseline characteristics and operative technique against CSF rhinorrhoea across transsphenoidal **(A, B)** and expanded endonasal **(C, D)** appraoches. CSF, cerebrospinal fluid, BMI=body mass index, TSA=transsphenoidal approach; EEA, expanded endonasal approach. *Statistically significant relationships (p<0.05, see Table 2 and Supplementary Information 3).

rates on multivariate analyses (Table 2; Figure 3; Supplementary Material 5).

Discussion

Principal findings

This multicentre, prospective, observational study represents the first study of its kind, exploring skull base repair techniques and CSF rhinorrhoea rates in a collaborative project involving almost all neurosurgical centres in the UK and Ireland.

There is clear heterogeneity in skull-base repair regimes across centres, with no two sharing the same protocol. Additionally, no specific type of repair technique made a significant difference in postoperative CSF rates, although there may be marginal benefits with tissue glue in TSA. Certain characteristics appear to make CSF rhinorrhoea more likely – previous endonasal surgery and intraoperative CSF leak. This translates into the tailoring of repair strategies. For example, in EEA, multilayer regimes using pedicled flaps, rigid buttresses (often with gasket sealing) and CSF diversion were frequent. Similarly, in the context of intraoperative CSF leak, tissue grafts, tissue glues, pedicled flaps and CSF diversion were used more often. Larger tumors (maximum diameter >1cm) were associated with the use of pedicled flaps, dural replacement and tissue glues. Surgeon preference or training may also factor in, with pedicled flaps and hemostatic agents used less in the absence of otorhinolaryngologists. Tissue grafts, tissue glues, and construct support strategies (e.g., rigid buttresses and CSF diversion) were less frequent in the absence of neurosurgical involvement.

CSF rhinorrhoea for both TSA (28/726, 3.9%) and EEA (10/140, 7.1%) are lower than generally reported in the literature (6, 7, 9, 10, 12, 23). This may reflect the ongoing improvement in endonasal skull-base repair and CSF rhinorrhoea rates, demonstrated by recent meta-analyses over time (24). Additionally, the UK and Ireland are consolidating pituitary services into dedicated "centres of excellence", which may influence surgical outcomes (25). Furthermore, as a prospective series, surgeons were aware of the monitoring of this outcome, perhaps influencing their management *via* the *Hawthorne effect* (26). Importantly, a significant proportion of postoperative CSF rhinorrhoea cases had no recorded intraoperative CSF leak (Total: 15/38; TSA: 11/28; EEA: 4/10), suggesting occult intraoperative leak, or possibly a thinned and vulnerable arachnoid dome which tears postoperatively in the absence of support. In our series, this subgroup had the lowest frequency of almost every repair method category (except synthetic grafts and hemostatic agents). This phenomenon is described in other case series, with many authors advocating for universal sellar repair for this reason, and some recommending routine use of intrathecal fluorescein (27, 28). However, these strategies should be balanced against the increased operative time, cost-effectiveness, and additional repair-related morbidity (e.g., donor site injuries or scars) (27, 28).

Findings in the context of literature

The development of endonasal techniques have revolutionized skull base surgery by allowing direct access to the skull base regions *via* a natural working channel which although long and narrow, accommodates specialized long instruments (1, 29–31). Gravity and the pressure gradient across the surgical bony/dural opening often assists surgical resection, for example in the descent of softer pituitary tumours, and is sometimes manipulated through pressure modulation (e.g., Valsalva and intrathecal saline injection) (1, 29–31). However, these advantages also contribute to the endonasal approach's inherent susceptibility of CSF rhinorrhoea – repairing the skull base using long rigid instruments *via* a narrow surgical corridor, with restricted motion and dexterity, against gravity and CSF pressure, creating a significant surgical challenge (1, 6, 29–31). This challenge has been met by refinements in endoscopic and microscope techniques, however, large variations in CSF rhinorrhoea rates still exist in both transsphenoidal and expanded endonasal surgery (6, 24, 32). An important component of this refinement has been the development of new closure strategies (24, 33–35).

However, recent systematic reviews of skull-base repair techniques have highlighted the variations across surgeons and centres, likely related to the lack of high-level comparative evidence (6, 36–38). There is an ever-expanding list of repair options, from autologous grafts to synthetic glues and even 3D-printed custom implants, without a complimentary expansion in the evidence base (6, 36, 39). These repair materials are sometimes supported by CSF diversion to reduce the pressure across the surgical repair. In fact, the only high-level evidence within the field of endonasal skull base repair is a randomized controlled trial investigating perioperative lumbar drainage (combined with nasoseptal flap repair) in EEA with high-flow intraoperative CSF leak (35). Lumbar drains were inserted immediately postoperatively (under the same anaesthetic), draining 10 ml/h for 3 days, resulting in a decrease in CSF rhinorrhoea rates (8.2% with lumbar drainage vs. 21.2% without; $p = 0.03$) (35).

Furthermore, most modern protocols adapt the extent of skull base reconstruction to postoperative CSF rhinorrhoea risk, balancing the risks of the former against the latter (4, 6, 10, 40–46). Numerous factors weigh into this decision-making, from demographics, co-morbidities, tumour characteristics, and operative factors (e.g., CSF leak), although the exact contribution of each potential factor in surgical decision-making remains poorly defined (6, 14, 22, 32, 44, 47, 48). Techniques reported commonly for low-risk cases include fat grafts, fascia lata grafts and synthetic grafts; whereas multilayer regimes with vascularized flaps, gasket-sealing, and lumbar drains are commoner in higher-risk cases (6, 37, 49, 50). Future studies would benefit from multimodal datasets which encompass these risk factors (e.g. combination of clinical metadata, imaging and operative video) and advanced analysis techniques (e.g. machine learning) to explore the interactions between risk factors, repair techniques and CSF rhinorrhoea rates.

Strengths and limitations

The strengths of this study are its prospective, consecutive recruitment (despite COVID-19), and the creation of a collaborative network of neurosurgeons and otorhinolaryngologists with a specialist interest in skull-base and pituitary, spanning almost every adult neurosurgical centre in the UK and Ireland. There are several limitations. Firstly, the study involved only two countries, limiting the generalizability of the findings. Furthermore, the study is observational and occurred during a pandemic wave, possibly hampering case recruitment. Due to pandemic-related pressures and redeployments, several centres uploaded data in retrospect but submitted cases were reviewed in detail by supervising consultants. Only one dedicated pediatric centre was included, although 6 centres (joint adult and pediatric) included patients less than 16 years old. CSF rhinorrhoea was infrequent, whilst there was a wide array of combinations for relevant variables (particularly skull-base repair methods) making statistical analysis challenging.

Conclusions

Heterogeneity of skull-base repair techniques exists across centres. Multilayer regimes with vascularized flaps, CSF diversion and rigid buttresses appear commoner in higher-risk cases, such as in EEAs. Overall, corresponding CSF rhinorrhoea

rates across the UK and Ireland are lower than generally reported in the literature. A large proportion of postoperative leaks occurred in the context of occult intraoperative CSF leaks, and decisions for universal sellar repairs should consider the risks and cost-effectiveness of repair methods used. Future work could include longer-term, higher-volume studies, such as a registry; and high-quality interventional studies.

Data availability statement

Data is available upon reasonable request. Requests to access the datasets should be directed to h.marcus@ucl.ac.uk.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

CRANIAL Consortium

Danyal Z Khan, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; **Hani J Marcus**, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; **Soham Bandyopadhyay**, Oxford University Global Surgery Group, Nuffield Department of Surgical Sciences, University of Oxford, Oxford; **Benjamin E Schroeder**, Department of Neurology, University Hospital of Wales, Cardiff University, Cardiff; **Vikesh Patel**, Division of Neurosurgery, Cambridge University Hospitals Trust, Cambridge**; Alice O'Donnell**, Birmingham Medical School, University of Birmingham, Birmingham; **Neurology and Neurosurgery Interest Group**, NANSIG; **British Neurosurgical Trainee Research Collaborative**, BNTRC; **Anastasios Giamouriadis**, Department of Neurosurgery, Aberdeen Royal Infirmary, Aberdeen; **Pragnesh Bhatt**, Department of Neurosurgery, Aberdeen Royal Infirmary, Aberdeen; **Bhaskar Ram**, Department of Otorhinolaryngology, Aberdeen Royal Infirmary, Aberdeen; **Adithya Varma**, Department of Neurosurgery, Aberdeen Royal Infirmary, Aberdeen; **Philip Weir**, Department of Neurosurgery, Royal Victoria Hospital, Belfast; **Brendan Hanna**, Department of Otorhinolaryngology, Royal Victoria Hospital, Belfast; **Theodore C Hirst**, Department of Neurosurgery, Royal Victoria Hospital, Belfast; **Patrick McAleavey**, Department of Neurosurgery, Royal Victoria Hospital, Belfast; **Alessandro Paluzzi**, Department of Neurosurgery, Queen Elizabeth Hospital Birmingham, Birmingham; **Georgios Tsermoulas**, Department of Neurosurgery, Queen Elizabeth Hospital Birmingham, Birmingham; **Shahzada Ahmed**, Department of Otorhinolaryngology, Queen Elizabeth Hospital Birmingham, Birmingham; **Wai Cheong Soon**, Department of Neurosurgery, Queen Elizabeth Hospital Birmingham, Birmingham; **Yasir Arafat Chowdhury**, Department of Neurosurgery, Queen Elizabeth Hospital Birmingham, Birmingham; **Suhaib Abualsaud**, Department of Neurosurgery, Queen Elizabeth Hospital Birmingham, Birmingham; **Shumail Mahmood**, Department of Neurosurgery, Queen Elizabeth Hospital Birmingham, Birmingham; **Paresh Naik**, Department of Otorhinolaryngology, Queen Elizabeth Hospital Birmingham, Birmingham; **Zohra Haiderkhan**, Department of Neurosurgery, Queen Elizabeth Hospital Birmingham, Birmingham; **Rafid Al-Mahfoudh**, Department of Neurosurgery, Hurstwood Park Neurosciences Centre and Royal Sussex County Hospital, Brighton; **Andrea Perera**, Department of Neurosurgery, Hurstwood Park Neurosciences Centre and Royal Sussex County Hospital, Brighton; **Mircea Rus**, Department of Neurosurgery, Hurstwood Park Neurosciences Centre and Royal Sussex County Hospital, Brighton; **Adam Williams**, Department of Neurosurgery, Southmead Hospital Bristol, Bristol; **Charles Hand**, Department of Neurosurgery, Southmead Hospital Bristol, Bristol; **Kumar Abhinav**, Department of Neurosurgery, Southmead Hospital Bristol, Bristol; **Cristina Cernei**, Department of Neurosurgery, Southmead Hospital Bristol, Bristol; **Aiman Dilnawaz**, Department of Neurosurgery, Southmead Hospital Bristol, Bristol; **Richard Mannion**, Division of Neurosurgery, Cambridge University Hospitals Trust, Cambridge; **Thomas Santarius**, Division of Neurosurgery, Cambridge University Hospitals Trust, Cambridge; **James Tysome**, Division of Otorhinolaryngology, Cambridge University Hospitals Trust, Cambridge; **Rishi Sharma**, Division of Otorhinolaryngology, Cambridge University Hospitals Trust, Cambridge; **Angelos G Kolias**, Division of Neurosurgery, Cambridge University Hospitals Trust, Cambridge; **Neil Donnelly**, Division of Otorhinolaryngology, Cambridge University Hospitals Trust, Cambridge; **Vikesh Patel**, Division of Neurosurgery, Cambridge University Hospitals Trust, Cambridge; **Ashwin Venkatesh**, Division of Neurosurgery, Cambridge University Hospitals Trust, Cambridge; **Caroline Hayhurst**, Department of Neurosurgery, University Hospital of Wales, Cardiff; **Amr Mohamed**, Department of Neurosurgery, University Hospital of Wales, Cardiff; **Benjamin Stew**, Department of Otorhinolaryngology, University Hospital of Wales, Cardiff; **Joseph Merola**, Department of Neurosurgery, University Hospital of Wales, Cardiff; **Setthasorn Zhi Yang Ooi**, Department of Neurosurgery, University Hospital of Wales, Cardiff; **Mahmoud Kamel**, Department of Neurosurgery, Cork University Hospitals, Ireland; **Mohammad Habibullah Khan**, Department of Otorhinolaryngology, Cork University Hospitals,

Ireland; **Sahibzada Abrar**, Department of Neurosurgery, Cork University Hospitals, Ireland; **Christopher Mckeon**, Department of Neurosurgery, Cork University Hospitals, Ireland; **Daniel McSweeney**, Department of Neurosurgery, Cork University Hospitals, Ireland; **Mohsen Javadpour**, Department of Neurosurgery, National Neurosurgical Centre, Beaumont Hospital, Ireland; **Peter Lacy**, Department of Otorhinolaryngology, National Neurosurgical Centre, Beaumont Hospital, Ireland; **Daniel Murray**, Department of Neurosurgery, National Neurosurgical Centre, Beaumont Hospital, Ireland; **Elena Roman**, Department of Neurosurgery, National Neurosurgical Centre, Beaumont Hospital, Ireland; **Kismet Hossain-Ibrahim**, Department of Neurosurgery, Ninewells Hospital, Dundee; **Peter Ross**, Department of Otorhinolaryngology, Ninewells Hospital, Dundee; **David Bennett**, Department of Neurosurgery, Ninewells Hospital, Dundee; **Nathan McSorley**, Department of Neurosurgery, Ninewells Hospital, Dundee; **Adam Hounat**, Department of Neurosurgery, Ninewells Hospital, Dundee; **Patrick Statham**, Department of Clinical Neurosciences, BioQuarter, Edinburgh; **Mark Hughes**, Department of Clinical Neurosciences, BioQuarter, Edinburgh; **Alhafidz Hamdan**, Department of Clinical Neurosciences, BioQuarter, Edinburgh; **Caroline Scott**, Department of Clinical Neurosciences, BioQuarter, Edinburgh; **Jigi Moudgil-Joshi**, Department of Clinical Neurosciences, BioQuarter, Edinburgh; **Anuj Bahl**, Department of Neurosurgery, Hull University Teaching Hospitals, Hull; **Anna Bjornson**, Department of Neurosurgery, Hull University Teaching Hospitals, Hull; **Daniel Gatt**, Department of Neurosurgery, Hull University Teaching Hospitals, Hull; **Nick Phillips**, Department of Neurosurgery, Leeds Teaching Hospitals, Leeds; **Neeraj Kalra**, Department of Neurosurgery, Leeds Teaching Hospitals, Leeds; **Melissa Bautista**, Department of Neurosurgery, Leeds Teaching Hospitals, Leeds; **Seerat Shirazi**, Department of Neurosurgery, Leeds Teaching Hospitals, Leeds; **Catherine E Gilkes**, Department of Neurosurgery, The Walton Centre, Liverpool; **Christopher P Millward**, Department of Neurosurgery, The Walton Centre, Liverpool; **Ahmad MS Ali**, Department of Neurosurgery, The Walton Centre, Liverpool; **Dimitris Paraskevopoulos**, Department of Neurosurgery, Barts and The Royal London Hospital, London; **Jarnail Bal**, Department of Neurosurgery, Barts and The Royal London Hospital, London; **Samir Matloob**, Department of Neurosurgery, Barts and The Royal London Hospital, London; **Rhannon Lobo**, Department of Neurosurgery, Barts and The Royal London Hospital, London; **Nigel Mendoza**, Department of Neurosurgery, Charing Cross Hospital, London; **Ramesh Nair**, Department of Neurosurgery, Charing Cross Hospital, London; **Arthur Dalton**, Department of Neurosurgery, Charing Cross Hospital, London; **Adarsh Nadig**, Department of Neurosurgery, Charing Cross Hospital, London; **Lucas Hernandez**, Department of Neurosurgery, Charing Cross Hospital, London; **Nick Thomas**, Department of Neurosurgery, King's College Hospital, London; **Eleni Maratos**, Department of Neurosurgery, King's College Hospital, London; **Jonathan Shapey**, Department of Neurosurgery, King's College Hospital, London; **Sinan Al-Barazi**, Department of Neurosurgery, King's College Hospital, London; **Asfand Baig Mirza**, Department of Neurosurgery, King's College Hospital, London; **Mohamed Okasha**, Department of Neurosurgery, King's College Hospital, London; **Prabhjot Singh Malhotra**, Department of Neurosurgery, King's College Hospital, London; **Razna Ahmed**, Department of Neurosurgery, King's College Hospital, London; **Neil L Dorward**, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; **Joan Grieve**, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; **Parag Sayal**, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; **David Choi**, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; **Ivan Cabrilo**, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; **Hugo Layard Horsfall**, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; **Jonathan Pollock**, Department of Neurosurgery, Barking, Havering & Redbridge University Hospitals, London; **Alireza Shoakazemi**, Department of Neurosurgery, Barking, Havering & Redbridge University Hospitals, London; **Oscar Maccormac**, Department of Neurosurgery, Barking, Havering & Redbridge University Hospitals, London; **Guru N K Amirthalingam**, Department of Neurosurgery, Barking, Havering & Redbridge University Hospitals, London; **Andrew Martin**, Department of Neurosurgery, St George's University Hospitals Trust, London; **Simon Stapleton**, Department of Neurosurgery, St George's University Hospitals Trust, London; **Florence Hogg**, Department of Neurosurgery, St George's University Hospitals Trust, London; **Daniel Richardson**, Department of Neurosurgery, St George's University Hospitals Trust, London; **Kanna Gnanalingham**, Department of Neurosurgery, Salford Royal Trust, Manchester; **Omar Pathmanaban**, Department of Neurosurgery, Salford Royal Trust, Manchester; **Daniel M Fountain**, Department of Neurosurgery, Salford Royal Trust, Manchester; **Raj Bhalla**, Department of Otorhinolaryngology, Salford Royal Trust, Manchester; **Cathal J Hannan**, Department of Neurosurgery, Salford Royal Trust, Manchester; **Annabel Chadwick**, Department of Neurosurgery, Salford Royal Trust, Manchester; **Alistair Jenkins**, Department of Neurosurgery, Royal Victoria Infirmary, Newcastle; **Claire Nicholson**, Department of Neurosurgery, Royal Victoria Infirmary, Newcastle; **Syed Shumon**, Department of Neurosurgery, Royal Victoria Infirmary, Newcastle; **Mohamed Youssef**, Department of Neurosurgery, Royal Victoria Infirmary, Newcastle; **Callum Allison**, Department of Neurosurgery, Royal Victoria Infirmary, Newcastle; **Graham Dow**, Department of Neurosurgery,

Queen's Medical Centre Nottingham, Nottingham; **Iain Robertson**, Department of Neurosurgery, Queen's Medical Centre Nottingham, Nottingham**; Laurence Johann Glancz**, Department of Neurosurgery, Queen's Medical Centre Nottingham, Nottingham; **Murugan Sitaraman**, Department of Neurosurgery, Queen's Medical Centre Nottingham, Nottingham; **Ashwin Kumaria**, Department of Neurosurgery, Queen's Medical Centre Nottingham, Nottingham; **Ananyo Bagchi**, Department of Neurosurgery, Queen's Medical Centre Nottingham, Nottingham; **Simon Cudlip**, Department of Neurosurgery, John Radcliffe Hospital, Oxford University Hospitals, Oxford; **Jane Halliday**, Department of Neurosurgery, John Radcliffe Hospital, Oxford University Hospitals, Oxford; **Rory J Piper**, Department of Neurosurgery, John Radcliffe Hospital, Oxford University Hospitals, Oxford; **Alexandros Boukas**, Department of Neurosurgery, John Radcliffe Hospital, Oxford University Hospitals, Oxford; **Meriem Amarouche**, Department of Neurosurgery, John Radcliffe Hospital, Oxford University Hospitals, Oxford; **Damjan Veljanoski**, Department of Neurosurgery, John Radcliffe Hospital, Oxford University Hospitals, Oxford; **Samiul Muquit**, Department of Neurosurgery, University Hospitals Plymouth, Plymouth; **Ellie Edlmann**, Department of Neurosurgery, University Hospitals Plymouth, Plymouth; **Haritha Maripi**, Department of Neurosurgery, University Hospitals Plymouth, Plymouth; **Yi Wang**, Department of Neurosurgery, University Hospitals Plymouth, Plymouth; **Mehnaz Hossain**, Department of Neurosurgery, University Hospitals Plymouth, Plymouth; **Andrew Alalade**, Department of Neurosurgery, Lancashire Teaching Hospitals NHS Foundation Trust, Preston; **Syed Maroof**, Department of Neurosurgery, Lancashire Teaching Hospitals NHS Foundation Trust, Preston; **Pradnya Patkar**, Department of Neurosurgery, Lancashire Teaching Hospitals NHS Foundation Trust, Preston; **Saurabh Sinha**, Department of Neurosurgery, Royal Hallamshire Hospital & Sheffield Children's Hospital, Sheffield; **Showkat Mirza**, Department of Otorhinolaryngology, Royal Hallamshire Hospital & Sheffield Children's Hospital, Sheffield; **Duncan Henderson**, Department of Neurosurgery, Royal Hallamshire Hospital & Sheffield Children's Hospital, Sheffield; **Mohammad Saud Khan**, Department of Neurosurgery, Royal Hallamshire Hospital & Sheffield Children's Hospital, Sheffield; **Nijaguna Mathad**, Department of Neurosurgery, University Hospital Southampton, Southampton; **Jonathan Hempenstall**, Department of Neurosurgery, University Hospital Southampton, Southampton; **Difei Wang,** Department of Neurosurgery, University Hospital Southampton, Southampton; **Pavan Marwaha**, Department of Neurosurgery, University Hospital Southampton, Southampton; **Simon Shaw**, Department of Neurosurgery, Royal Stoke University Hospital, Stoke; **Georgios Solomou**, Department of Neurosurgery, Royal Stoke University Hospital, Stoke; **Alina Shrestha**, Department of Neurosurgery, Royal Stoke University Hospital, Stoke. List of Collaborators (data validators): **Andrew Fraser**, Department of Neurosurgery, Aberdeen Royal Infirmary, Aberdeen**; Theodore Hirst**, Department of Neurosurgery, Royal Victoria Hospital, Belfast; **Yasir Chowdhury,** Department of Neurosurgery, Queen Elizabeth Hospital Birmingham, Birmingham; **Sobiya Bilal**, Department of Neurosurgery, Hurstwood Park Neurosciences Centre and Royal Sussex County Hospital, Brighton; **Jack Wildman**, Department of Neurosurgery, Southmead Hospital Bristol, Bristol; **Ashwin Venkatesh**, Division of Neurosurgery, Cambridge University Hospitals Trust, Cambridge; **Priya Babu**, Department of Neurosurgery, University Hospital of Wales, Cardiff; **Cian Carey**, Department of Neurosurgery, Cork University Hospitals, Ireland; **Renitha Reddi Bathuni**, Department of Neurosurgery, National Neurosurgical Centre, Beaumont Hospital, Ireland; **Kismet Hossain-Ibrahim**, Department of Neurosurgery, Ninewells Hospital, Dundee; **Joseph Nathaniel Brennan**, Department of Neurosurgery, The Western General Hospital, Edinburgh; **Anna Bjornson**, Department of Neurosurgery, Hull University Teaching Hospitals, Hull **Howra Ktayen**, Department of Neurosurgery, Leeds Teaching Hospitals, Leeds; **Sandhya T Trichinopoly**, Department of Neurosurgery, The Walton Centre, Liverpool; **Samir Matloob**, Department of Neurosurgery, Barts and The Royal London Hospital, London; **Adarsh Nadig**, Department of Neurosurgery, Charing Cross Hospital, London; **Mohamed Okasha,** Department of Neurosurgery, King's College Hospital, London; **Danyal Khan**, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; **Alireza Shoakazemi**, Department of Neurosurgery, Barking, Havering & Redbridge University Hospitals, London; **Florence Hogg**, Department of Neurosurgery, St George's University Hospitals Trust, London; **Seun Sobawale**, Department of Neurosurgery, Salford Royal Trust, Manchester; **Amir Suliman,** Department of Neurosurgery, Royal Victoria Infirmary, Newcastle; **Ashwin Kumaria**, Department of Neurosurgery, Queen's Medical Centre Nottingham, Nottingham; **Rory Piper**, Department of Neurosurgery, John Radcliffe Hospital, Oxford University Hospitals, Oxford; **Will Owen**, Department of Neurosurgery, John Radcliffe Hospital, Oxford University Hospitals, Oxford; **Ellie Edlmann**, Department of Neurosurgery, University Hospitals Plymouth, Plymouth; **Afaq Sartaj**, Department of Neurosurgery, Lancashire Teaching Hospitals NHS Foundation Trust, Preston; **Edward Goacher**, Department of Neurosurgery, Royal Hallamshire Hospital & Sheffield Children's Hospital, Sheffield; **Euan Strachan**, Department of Neurosurgery, University Hospital Southampton, Southampton; **Giorgios Solomou**, Department of Neurosurgery, Royal Stoke University Hospital, Stoke.

Author contributions

This is a group authorship model paper where all authors contributed to data collection and approved the final manuscript. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgments

The authors would like to thank the Neurology and Neurosurgery Interest Group (NANSIG) and the British Neurosurgical Trainee Research Collaborative (BNTRC) without which this study would not have been possible. A special thanks to the data validation team for ensuring data accuracy (Supplementary Material 1).

Conflict of interest

HM is supported by the Wellcome (203145Z/16/Z) EPSRC (NS/A000050/1) Centre for Interventional and Surgical Sciences, University College London. HM is also funded by the NIHR Biomedical Research Centre at University College London. DK is supported by an NIHR Academic Clinical Fellowship. DK is also supported by a Cancer Research UK Predoctoral Fellowship. For the purpose of Open Access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.1049627/full#supplementary-material

References

1. Cappabianca P, Cavallo LM, de Divitiis E. Endoscopic endonasal transsphenoidal surgery. *Neurosurgery* (2004) 55(4):933–41. doi: 10.1227/01.NEU.0000137330.02549.0D

2. Liu JK, Das K, Weiss MH, Laws ERJr., Couldwell WT. The history and evolution of transsphenoidal surgery. *J Neurosurg* (2001) 95(6):1083–96. doi: 10.3171/jns.2001.95.6.1083

3. Kassam A, Carrau RL, Snyderman CH, Gardner P, Mintz A. Evolution of reconstructive techniques following endoscopic expanded endonasal approaches. *Neurosurg Focus* (2005) 19(1):1–7. doi: 10.3171/foc.2005.19.1.9

4. Dehdashti AR, Ganna A, Witterick I, Gentili F. Expanded endoscopic endonasal approach for anterior cranial base and suprasellar lesions: Indications and limitations. *Neurosurgery* (2009) 64(4):677–89. doi: 10.1227/01.NEU.0000339121.20101.85

5. Marcus HJ, Khan DZ, Borg A, Buchfelder M, Cetas JS, Collins JW, et al. Pituitary society expert Delphi consensus: Operative workflow in endoscopic transsphenoidal pituitary adenoma resection. *Pituitary* (2021) 24:1–15. doi: 10.1007/s11102-021-01162-3

6. Khan DZ, Ali AM, Koh CH, Dorward NL, Grieve J, Layard Horsfall H, et al. Skull base repair following endonasal pituitary and skull base tumour resection: A systematic review. *Pituitary* (2021) 24:1–16. doi: 10.1007/s11102-021-01145-4

7. Esquenazi Y, Essayed WI, Singh H, Mauer E, Ahmed M, Christos PJ, et al. Endoscopic endonasal versus microscopic transsphenoidal surgery for recurrent and/or residual pituitary adenomas. *World Neurosurg* (2017) 101:186–95. doi: 10.1016/j.wneu.2017.01.110

8. Fraser S, Gardner PA, Koutourousiou M, Kubik M, Fernandez-Miranda JC, Snyderman CH, et al. Risk factors associated with postoperative cerebrospinal fluid leak after endoscopic endonasal skull base surgery. *J Neurosurg* (2018) 128(4):1066–71. doi: 10.3171/2016.12.JNS1694

9. Ivan C, Ann R, Craig B, Debi P. Complications of transsphenoidal surgery: Results of a national survey, review of the literature, and personal experience. *Neurosurgery* (1997) 40(2):225–37. doi: 10.1097/00006123-199702000-00001

10. Nishioka H, Haraoka J, Ikeda Y. Risk factors of cerebrospinal fluid rhinorrhea following transsphenoidal surgery. *Acta Neurochir (Wien)* (2005) 147(11):1163–6. doi: 10.1007/s00701-005-0586-3

11. Patel MR, Stadler ME, Snyderman CH, Carrau RL, Kassam AB, Germanwala AV, et al. How to choose? endoscopic skull base reconstructive options and limitations. *Skull Base* (2010) 20(06):397–404. doi: 10.1055/s-0030-1253573

12. Strickland BA, Lucas J, Harris B, Kulubya E, Bakhsheshian J, Liu C, et al. Identification and repair of intraoperative cerebrospinal fluid leaks in endonasal transsphenoidal pituitary surgery: Surgical experience in a series of 1002 patients. *J Neurosurg* (2017) 129(2):425–9. doi: 10.3171/2017.4.JNS162451

13. Kono Y, Prevedello DM, Snyderman CH, Gardner PA, Kassam AB, Carrau RL, et al. One thousand endoscopic skull base surgical procedures demystifying the infection potential: Incidence and description of postoperative meningitis and brain abscesses. *Infect Control Hosp Epidemiol* (2011) 32(1):77–83. doi: 10.1086/657635

14. Liu P, Wu S, Li Z, Wang B. Surgical strategy for cerebrospinal fluid rhinorrhea repair. *Operative Neurosurg* (2010) 66(suppl_2):ons281–6. doi: 10.1227/01.NEU.0000369660.30126.02

15. CRANIAL-Consortium. Csf rhinorrhea after endonasal intervention to the skull base (Cranial) — Part 2: Impact of covid-19. *World Neurosurg* (2021) 149:e1090–7. doi: 10.1016/j.wneu.2020.12.169

16. CRANIAL-Consortium. Csf rhinorrhoea after endonasal intervention to the skull base (Cranial) - part 1: Multicenter pilot study. *World Neurosurg* (2021) 149: e1077–89. doi: 10.1016/j.wneu.2020.12.171

17. Khan DZ, Bandyopadhyay S, Patel V, Schroeder BE, Cabrilo I, Choi D, et al. Csf rhinorrhoea after endonasal intervention to the anterior skull base (Cranial): Proposal for a prospective multicentre observational cohort study. *Br J Neurosurg* (2020) 35:1–10. doi: 10.1080/02688697.2020.1795622

18. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (Strobe) statement: Guidelines for reporting observational studies. *Ann Intern Med* (2007) 147(8):573–7. doi: 10.7326/0003-4819-147-8-200710160-00010

19. Lee JJ, Kim HY, Dhong HJ, Chung SK, Kong DS, Nam DH, et al. Delayed cerebrospinal fluid leakage after treatment of skull base tumors: Case series of 9 patients. *World Neurosurg* (2019) 132:e591–8. doi: 10.1016/j.wneu.2019.08.067

20. Kassam AB, Gardner P, Snyderman C, Mintz A, Carrau R. Expanded endonasal approach: Fully endoscopic, completely transnasal approach to the middle third of the clivus, petrous bone, middle cranial fossa, and infratemporal fossa. *Neurosurg Focus* (2005) 19(1):1–10. doi: 10.3171/foc.2005.19.1.7

21. Chari A, Jamjoom AA, Edlmann E, Ahmed AI, Coulter IC, Ma R, et al. The British neurosurgical trainee research collaborative: Five years on. *Acta Neurochir (Wien)* (2018) 160(1):23–8. doi: 10.1007/s00701-017-3351-5

22. Esposito F, Dusick JR, Fatemi N, Kelly DF. Graded repair of cranial base defects and cerebrospinal fluid leaks in transsphenoidal surgery. *Operative Neurosurg* (2007) 60 (suppl_4):ONS–295. doi: 10.1227/01.NEU.0000255354.64077.66

23. Li A, Liu W, Cao P, Zheng Y, Bu Z, Zhou T. Endoscopic versus microscopic transsphenoidal surgery in the treatment of pituitary adenoma: A systematic review and meta-analysis. *World Neurosurg* (2017) 101:236–46. doi: 10.1016/j.wneu.2017.01.022

24. Zamanipoor Najafabadi AH, Khan DZ, Muskens IS, Broekman MLD, Dorward NL, van Furth WR, et al. Trends in cerebrospinal fluid leak rates following the extended endoscopic endonasal approach for anterior skull base meningioma: A meta-analysis over the last 20 years. *Acta Neurochir (Wien)* (2020) 89:1567–76. doi: 10.1007/s00701-020-04641-x

25. Casanueva FF, Barkan AL, Buchfelder M, Klibanski A, Laws ER, Loeffler JS, et al. Criteria for the definition of pituitary tumor centres of excellence (Ptcoe): A pituitary society statement. *Pituitary* (2017) 20(5):489–98. doi: 10.1007/s11102-017-0838-2

26. Demetriou C, Hu L, Smith TO, Hing CB. Hawthorne Effect on surgical studies. *ANZ J Surg* (2019) 89(12):1567–76. doi: 10.1111/ans.15475

27. Jakimovski D, Bonci G, Attia M, Shao H, Hofstetter C, Tsiouris AJ, et al. Incidence and significance of intraoperative cerebrospinal fluid leak in endoscopic pituitary surgery using intrathecal fluorescein. *World Neurosurg* (2014) 82(3-4): e513–23. doi: 10.1016/j.wneu.2013.06.005

28. Sanders-Taylor C, Anaizi A, Kosty J, Zimmer LA, Theodosopoulos PV. Sellar reconstruction and rates of delayed cerebrospinal fluid leak after endoscopic pituitary surgery. *J Neurol Surg B Skull Base* (2015) 76(4):281. doi: 10.1055/s-0034-1544118

29. Cavallo LM, Messina A, Esposito F, de Divitiis O, Dal Fabbro M, de Divitiis E, et al. Skull base reconstruction in the extended endoscopic transsphenoidal approach for suprasellar lesions. *J Neurosurg* (2007) 107(4):713–20. doi: 10.3171/JNS-07/10/0713

30. Castaño-Leon AM, Paredes I, Munarriz PM, Jiménez-Roldán L, Hilario A, Calatayud M, et al. Endoscopic transnasal trans-sphenoidal approach for pituitary adenomas: A comparison to the microscopic approach cohort by propensity score analysis. *Neurosurgery* (2020) 86(3):348–56. doi: 10.1093/neuros/nyz201

31. Ogiwara T, Nagm A, Hasegawa T, Hanaoka Y, Ichinose S, Goto T, et al. Pitfalls of skull base reconstruction in endoscopic endonasal approach. *Neurosurg Rev* (2019) 42(3):683–9. doi: 10.1007/s10143-018-1006-5

32. Conger A, Zhao F, Wang X, Eisenberg A, Griffiths C, Esposito F, et al. Evolution of the graded repair of csf leaks and skull base defects in endonasal endoscopic tumor surgery: Trends in repair failure and meningitis rates in 509 patients. *J Neurosurg* (2018) 130(3):861–75. doi: 10.3171/2017.11.JNS172141

33. Hadad G, Bassagasteguy L, Carrau RL, Mataza JC, Kassam A, Snyderman CH, et al. A novel reconstructive technique after endoscopic expanded endonasal approaches: Vascular pedicle nasoseptal flap. *Laryngoscope* (2006) 116(10):1882–6. doi: 10.1097/01.mlg.0000234933.37779.e4

34. Leng LZ, Brown S, Anand VK, Schwartz TH. "Gasket-seal" watertight closure in minimal-access endoscopic cranial base surgery. *Operative Neurosurg* (2008) 62(suppl_5):ONS342–3. doi: 10.1227/01.neu.0000326017.84315.1f

35. Zwagerman NT, Wang EW, Shin SS, Chang YF, Fernandez-Miranda JC, Snyderman CH, et al. Does lumbar drainage reduce postoperative cerebrospinal fluid leak after endoscopic endonasal skull base surgery? a prospective, randomized controlled trial. *J Neurosurg* (2018) 1:1–7. doi: 10.3171/2018.4.JNS172447

36. Hannan CJ, Kelleher E, Javadpour M. Methods of skull base repair following endoscopic endonasal tumor resection: A review. *Front Oncol* (2020) 10. doi: 10.3389/fonc.2020.01614

37. Harvey RJ, Parmar P, Sacks R, Zanation AM. Endoscopic skull base reconstruction of Large dural defects: A systematic review of published evidence. *Laryngoscope* (2012) 122(2):452–9. doi: 10.1002/lary.22475

38. Oakley GM, Orlandi RR, Woodworth BA, Batra PS, Alt JA. Management of cerebrospinal fluid rhinorrhea: An evidence-based review with recommendations. *Int Forum Allergy Rhinol* (2016) 6(1):17–24. doi: 10.1002/alr.21627

39. Essayed WI, Unadkat P, Hosny A, Frisken S, Rassi MS, Mukundan S, et al. 3d printing and intraoperative neuronavigation tailoring for skull base reconstruction after extended endoscopic endonasal surgery: Proof of concept. *J Neurosurg JNS.* (2018) 130(1):248–55. doi: 10.3171/2017.9.JNS171253

40. Zhou Q, Yang Z, Wang X, Wang Z, Zhao C, Zhang S, et al. Risk factors and management of intraoperative cerebrospinal fluid leaks in endoscopic treatment of pituitary adenoma: Analysis of 492 patients. *World Neurosurg* (2017) 101:390–5. doi: 10.1016/j.wneu.2017.01.119

41. Zaidi HA, Awad A-W, Bohl MA, Chapple K, Knecht L, Jahnke H, et al. Comparison of outcomes between a less experienced surgeon using a fully endoscopic technique and a very experienced surgeon using a microscopic transsphenoidal technique for pituitary adenoma. *J Neurosurg* (2016) 124 (3):596–604. doi: 10.3171/2015.4.JNS15102

42. Rabadán AT, Hernández D, Ruggeri CS. Pituitary tumors: Our experience in the prevention of postoperative cerebrospinal fluid leaks after transsphenoidal surgery. *J Neurooncol* (2009) 93(1):127–31. doi: 10.1007/s11060-009-9858-8

43. Nix P, Tyagi A, Phillips N. Retrospective analysis of anterior skull base csf leaks and endoscopic repairs at Leeds. *Br J Neurosurg* (2016) 30(4):422–6. doi: 10.3109/02688697.2016.1161176

44. Hannan CJ, Almhanedi H, Al-Mahfoudh R, Bhojak M, Looby S, Javadpour M. Predicting post-operative cerebrospinal fluid (Csf) leak following endoscopic transnasal pituitary and anterior skull base surgery: A multivariate analysis. *Acta Neurochir (Wien)* (2020) 162(6):1309–15. doi: 10.1007/s00701-020-04334-5

45. Han Z-L, He D-S, Mao Z-G, Wang H-J. Cerebrospinal fluid rhinorrhea following trans-sphenoidal pituitary macroadenoma surgery: Experience from 592 patients. *Clin Neurol Neurosurg* (2008) 110(6):570–9. doi: 10.1016/j.clineuro.2008.02.017

46. Dlouhy BJ, Madhavan K, Clinger JD, Reddy A, Dawson JD, O'Brien EK, et al. Elevated body mass index and risk of postoperative csf leak following transsphenoidal surgery. *J Neurosurg* (2012) 116(6):1311–7. doi: 10.3171/2012.2.JNS111837

47. Park J-H, Choi JH, Kim Y-I, Kim SW, Hong Y-K. Modified graded repair of cerebrospinal fluid leaks in endoscopic endonasal transsphenoidal surgery. *J Korean Neurosurgical Society* (2015) 58(1):36. doi: 10.3340/jkns.2015.58.1.36

48. Cappabianca P, Cavallo LM, Esposito F, Valente V, de Divitiis E. Sellar repair in endoscopic endonasal transsphenoidal surgery: Results of 170 cases. *Neurosurgery* (2002) 51(6):1365–72. doi: 10.1097/00006123-200212000-00006

49. Iavarone A, Luparello P, Lazio MS, Comini LV, Martelli F, De Luca O, et al. The surgical treatment of cerebrospinal fistula: Qualitative and quantitative analysis of indications and results. *Head Neck* (2020) 42(2):344–56. doi: 10.1002/hed.25981

50. Soudry E, Turner JH, Nayak JV, Hwang PH. Endoscopic reconstruction of surgically created skull base defects: A systematic review. *Otolaryngology–Head Neck Surg* (2014) 150(5):730–8. doi: 10.1177/0194599814520685

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 23 March 2023
DOI 10.3389/fonc.2023.1046519

OPEN ACCESS

EDITED BY
Alireza Mansouri,
The Pennsylvania State University (PSU), United States

REVIEWED BY
Julius Höhne,
University Medical Center Regensburg, Germany
Ramez Kirollos,
National Neuroscience Institute (NNI), Singapore

*CORRESPONDENCE
Adrito Das
adrito.das.20@ucl.ac.uk
Danyal Z. Khan
d.khan@ucl.ac.uk

SPECIALTY SECTION
This article was submitted to Neuro-Oncology and Neurosurgical Oncology, a section of the journal Frontiers in Oncology

RECEIVED 16 September 2022
ACCEPTED 27 January 2023
PUBLISHED 23 March 2023

CITATION
CRANIAL Consortium (2023) Machine learning driven prediction of cerebrospinal fluid rhinorrhoea following endonasal skull base surgery: A multicentre prospective observational study.
Front. Oncol. 13:1046519.
doi: 10.3389/fonc.2023.1046519

Machine learning driven prediction of cerebrospinal fluid rhinorrhoea following endonasal skull base surgery: A multicentre prospective observational study

CRANIAL Consortium

Background: Cerebrospinal fluid rhinorrhoea (CSFR) is a common complication following endonasal skull base surgery, a technique that is fundamental to the treatment of pituitary adenomas and many other skull base tumours. The CRANIAL study explored CSFR incidence and related risk factors, particularly skull base repair techniques, *via* a multicentre prospective observational study. We sought to use machine learning to leverage this complex multicentre dataset for CSFR prediction and risk factor analysis.

Methods: A dataset of 865 cases - 725 transsphenoidal approach (TSA) and 140 expanded endonasal approach (EEA) - with cerebrospinal fluid rhinorrhoea as the primary outcome, was used. Relevant variables were extracted from the data, and prediction variables were divided into two categories, preoperative risk factors; and repair techniques, with 6 and 11 variables respectively. Three types of machine learning models were developed in order to predict CSFR: logistic regression (LR); decision tree (DT); and neural network (NN). Models were validated using 5-fold cross-validation, compared *via* their area under the curve (AUC) evaluation metric, and key prediction variables were identified using their Shapley additive explanations (SHAP) score.

Results: CSFR rates were 3.9% (28/725) for the transsphenoidal approach and 7.1% (10/140) for the expanded endonasal approach. NNs outperformed LR and DT for CSFR prediction, with a mean AUC of 0.80 (0.70-0.90) for TSA and 0.78 (0.60-0.96) for EEA, when all risk factor and intraoperative repair data were integrated into the model. The presence of intraoperative CSF leak was the most prominent risk factor for CSFR. Elevated BMI and revision surgery were also associated with CSFR for the transsphenoidal approach. CSF diversion and gasket sealing appear to be strong predictors of the absence of CSFR for both approaches.

Conclusion: Neural networks are effective at predicting CSFR and uncovering key CSFR predictors in patients following endonasal skull base surgery, outperforming traditional statistical methods. These models will be improved further with larger and more granular datasets, improved NN architecture, and external validation. In

the future, such predictive models could be used to assist surgical decision-making and support more individualised patient counselling.

KEYWORDS

cerebrospinal fluid leak, cerebrospinal fluid rhinorrhoea, CSF, endoscopic endonasal, skull base surgery, machine learning - ML, neural network, outcome prediction

1 Introduction

Endonasal operative approaches, including the transsphenoidal approach (TSA) and the expanded endonasal approach (EEA), have become workhorses in skull base neurosurgery (1, 2). They are predominately used in the treatment of pituitary adenomas and other sella-region neoplastic pathologies, with growing indications as these techniques evolve (3, 4). Despite the benefits the approaches offer in terms of access, the most common surgical complication remains cerebrospinal fluid rhinorrhoea (CSFR) – generally up to 5% in TSA and 20% in EEA, although these rates vary significantly across the literature (3, 5–18). CSFR has potentially serious sequelae, including meningitis; severe headache, pneumocephalus; increased length of hospital stays; re-admission; and need for further surgery (9, 12, 13).

Numerous risk factors have been identified for CSFR, including the presence of intraoperative cerebrospinal fluid (CSF) leak; revision surgery; and high body mass index (BMI) (19). A particularly important factor is the choice of skull base repair technique used intraoperatively (7, 10, 13, 16, 20). A recent expert consensus conducted *via* The Pituitary Society highlighted the practice variations across TSA, particularly during the skull base closure phase (21). A systematic review of the literature has found absolute heterogeneity across studies and centres in terms of skull base repair techniques, likely due to a lack of high-level comparative evidence (10).

CRANIAL (CSF rhinorrhoea after endonasal intervention to the skull base) was a prospective, multicentre observational study seeking to determine the: (1) scope of the methods of skull base repair; and (2) corresponding rates of CSFR (22–25). It represents the largest dataset of its kind, seeking to audit practice across the UK and Ireland. It revealed a CSFR incidence rate of 3.9% for TSA and 7.1% for EEA, lower than the literature standard, with minimal influence of particular repair regimes on CSFR incidence *via* traditional statistical analysis (25). In neurosurgery, machine learning models (MLs), or more specifically neural network models (NNs), have been shown to outperform these traditional statistical methods by leveraging their ability to utilise complex non-linear relationships between the various prediction variables (26–28). For example, NNs were able to identify the risk factors associated with a high risk of intraoperative CSF leak where traditional statistical analysis failed (29).

In this study, we use NNs on the granular multicentre CRANIAL dataset for analysis of CSFR, its risk factors, and the comparative effectiveness of skull base repair techniques in both TSA and EEA.

Abbreviations: AUC, area under the receiver operating characteristic; BMI, body mass index; CART, classification and regression trees; CRANIAL, CSF rhinorrhoea after endonasal intervention to the skull base; CSF, cerebrospinal fluid; CSFR, cerebrospinal fluid rhinorrhoea; C-value, inverse of regularization; DT(s), decision tree (models); EEA, expanded endonasal approach; IQR, inter-quartile range; LR(s), logistic regression (models); ML(s), machine learning (models); NN(s), neural network (models); ReLu, rectified linear activation unit; SGD, stochastic gradient decent; SHAP, Shapley additive explanations; TRIPOD, transparent reporting of a multivariable prediction model for individual prognosis or diagnosis; TSA, transsphenoidal approach.

2 Methods

The transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guided this methodology and report (30).

2.1 Data

2.1.1 Collection

A detailed description of the generation of the CRANIAL dataset is described in (25). In brief, it is a multicentre dataset (30 centres in the UK and Ireland), collected *via* a prospective observational study in 3 phases encompassing November 2019 – July 2020 (22–25). All TSA (defined as transsphenoidal access to the sella alone) and EEA [defined as acquiring surgical access to an area beyond the sella (17, 19)] were included. The dataset is composed of baseline characteristic data (e.g., age; sex; tumour diameter), operative data (e.g., intraoperative CSF leak presence; skull base repair method) and postoperative outcomes (e.g., CSFR) (22–25). A taxonomy for skull base repair was adapted from a systematic review of the literature (10, 24). Postoperative CSFR was confirmed biochemically and/or required intervention (CSF diversion and/or operative repair) (22–25).

2.1.2 Processing

The dataset contained 866 participants (726 TSA, 140 EEA). Variables relevant to CSFR (as guided by consensus-derived protocol and literature review) were extracted from the dataset (24, 25). The primary outcome was CSFR. Prediction variables (predictors) were divided into two prediction categories: 'preoperative risk factors for CSFR' (risk factors) and 'repair techniques used' (repair techniques), with 6 and 11 predictors respectively, as shown in Table 1. Tumour type has been excluded as a risk factor predictor in this study, as many

of the tumour types are too few in number for internal validation. Ultimately, this results in three prediction categories: 1) risk factors; 2) repair technique; 3) risk factors and repair technique.

The participants were divided into three approach categories: TSA; EEA; TSA or EEA. This, therefore, leads to nine total subcategories for each method: a separate model for the three approach categories multiplied by the three prediction categories. One additional model was created using surgical approach as a predictor, and hence the final number of subcategories for each method is 10.

Binary values (1 for used, 0 for not used) were set for all 11 repair technique predictors, and if missing, assumed not to be used and hence set to 0. Binary values were also set for the risk factor predictors: sex (male set to 1, female set to 0); BMI (>30 set to 1, ≤ 30 set to 0); tumour size (tumour diameter ≥ 1cm set to 1, tumour diameter < 1cm set to 0); intraoperative CSF leak (grade 1, 2, 3, or present but unknown grade set to 1, not present set to 0). Intraoperative CSF leak grade was not set as a categoric variable as conversion to a nominal variable would split each grade into its own prediction variable, leading to poorer correlations; and conversion to an ordinal variable would require the loss of the present but unknown grade category, representing an 18% loss of positive cases. Age was left as a continuous predictor but normalised to a Gaussian distribution with mean 0 and standard deviation 1. If any risk factor predictor was missing, the participant was excluded. Binary values were also assigned to the surgical approach (TSA set to 0, EEA set to 1) and CSFR (1 for present, 0 for not present), and if either was missing, the participant was excluded.

2.2 Model development

2.2.1 Machine learning

Three ML methods have been used in this study: logistic regression models (LRs); decision tree models (DTs); and neural network models. These have been chosen as they represent the increasing complexity of ML methods as measured by the number of adjustable parameters present in each model. The code is written in Python 3.8 (31, 32).

2.2.2 Validation

For validation, 5-fold cross-validation was used, with an 80:20 training to validation split for each fold. This was achieved by first separating the participants by the two surgical approaches, and then further separating the participants by the two CSFR outcomes, leading to four subgroups of participants (TSA with CSFR; TSA without CSFR; EEA with CSFR; EEA without CSFR). For each of these subgroups, the participants were randomly split into 5-folds, and assigned an appropriate fold number (1 to 5). Next, the participants from each output subgroup were combined by fold

TABLE 1 Distribution details of variables (predictors, approach, outcome) split by approach categories.

Category	Parameter	Distribution		
Approach	*Surgical Approach*	TSA 725 (83.4%)	EEA 140 (16.2%)	TSA or EEA (866)
Risk Factors	*Median Age (IQR)*	53 (41-64) years	51 (34-62) years	53 (40-63) years
	Male Sex	355 (49.0%)	61 (43.5%)	416 (48.0%)
	BMI > 30	210 (29.0%)	28 (20.0%)	238 (27.5%)
	Tumour Diameter ≥ 1cm	606 (83.6%)	131 (93.6%)	737 (85.2%)
	Revision Surgery	98 (13.5%)	21 (15.0%)	119 (13.8%)
	Presence of Intraoperative CSF Leak	214 (29.5%)	79 (56.4%)	293 (33.9%)
Repair Techniques	*CSF Diversion*	29 (4.0%)	38 (27.1%)	67 (7.8%)
	Dural Closure	0 (0.0%)	0 (0.0%)	0 (0.0%)
	Dural Replacement	196 (27.0%)	66 (47.1%)	262 (30.3%)
	Vascularised Flap	116 (16.0%)	90 (64.3%)	206 (23.8%)
	Tissue Graft	221 (30.5%)	65 (46.4%)	286 (33.1%)
	Synthetic Graft	203 (28.0%)	47 (33.6%)	251 (28.9%)
	Tissue Glue	473 (65.2%)	114 (81.4%)	587 (67.7%)
	Haemostatic Agent	439 (60.6%)	93 (66.4%)	532 (61.5%)
	Rigid Buttress	31 (4.3%)	17 (12.1%)	48 (5.5%)
	Gasket Seal	15 (2.1%)	11 (7.9%)	26 (3.0%)
	Nasal Packing	518 (71.4%)	116 (82.9%)	635 (73.3%)
Outcome	*CSFR*	28 (3.9%)	10 (7.1%)	38 (4.4%)

All variables are binary, excluding age which is continuous. For the binary variables the number of entries where the variable is present (represented as a 1) is given, with the round brackets giving the percentage (%) proportion. For the singular continuous parameter (age), median; and inter-quartile range (IQR) are given instead.

number, producing two groups separated by surgical approach. Finally, these two approach groups were combined by fold number. This means there are three groups separated by approach (TSA; EEA; TSA or EEA), where the ratio between the two CSFR binary outputs remains approximately the same for each fold as found in the data. Moreover, the ratio between TSA and EEA in the 'TSA or EEA' approach group remains the same as found in the data. This group methodology is displayed in Figure 1 and variable (predictors, surgical approach, outcome) distributions for each of the 5-folds can be found in Supplementary Material Table 3.

For each fold, after a model was trained on the other folds' participants (training dataset), it was then evaluated on the fold participants (validation dataset), and the evaluation metrics recorded. After repeating this for all folds, the evaluation metrics for both the mean-average and standard deviations were calculated across the 5-folds. Hyperparameter tuning of all MLs were performed through multiple runs on the validation dataset *via* grid search, and for NNs this was done at the epoch level.

Given the number of participants with CSFR represents just 4.4% of the data, for the training dataset, these participants were oversampled randomly such that they represent 50% of the data. This prevents overfitting to the entries without CSFR, where the models would simply always predict CSFR not occurring, leading to an effectively useless model. For evaluation metric calculations of both the training and validation datasets, no such oversampling was done.

2.3 Evaluation

2.3.1 AUC

The primary evaluation metric to compare MLs is the 'area under the receiver operating characteristic' (AUC) curve, which ensures a balance of both the sensitivity (true positive rate) and specificity (true negative rate), and these two are also given as secondary evaluation metrics.

2.3.2 SHAP

To compare a specific predictor's contribution to a NN predicting CSFR, 'Shapley additive explanations' (SHAP) scores were calculated. The SHAP method does this by calculating all possible combinations of the predictors, inputting each predictor combination into the model, and evaluating the combination's contribution to the model on the validation dataset. By doing this, each predictor's contribution to the model is calculated in isolation of the other predictors while also accounting for the non-linear relationships (33).

The magnitude (independent of score sign) of a SHAP score determines how large of a contribution that predictor has to the NN's outcome prediction. The sign of a predictor's score determines whether the NN has an increased (if positive) or decreased (if negative) probability of predicting a CSFR. A red dot means this probability is due to the predictor being present, a blue dot means it is due to the predictor not being present. If the red and blue dots have a clear boundary about a score of 0.0 and are not overlapping, this is interpreted as the predictor's value being highly correlated with the NN's outcome prediction. Similarly, the greater the

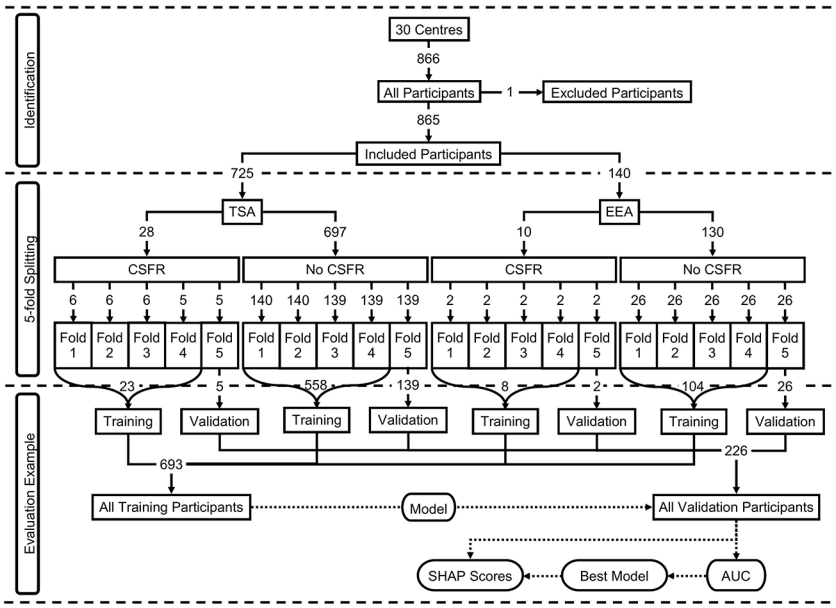


FIGURE 1
Participants breakdown displayed as a flowchart. The top section (identification) displays the included and excluded participants. The middle section (5-fold splitting) displays how the 5-folds were created, including the breakdown by surgical approach and outcome. The predictor distributions of the overall participants can be seen in Table 1, and the predicter distributions for each of the 5-folds can be seen in Supplementary Material Table 3 The bottom section (evaluation example) displays an example of a model training on one fold's training dataset, and then evaluated on the same fold's validation dataset.

overlap, the weaker the correlation. (Note purple dots are seen for age as it is a continuous variable: here red represents the oldest participant; blue the youngest participant; and purple for the ages in between.)

3 Results

3.1 Data

Out of the initial 866 participants, one case was removed due to missing age, resulting in 855 cases (725 TSA, 140 EEA). Full distribution details of all included variables (predictors, surgical approach, outcome) are given in Table 1, and the distribution across each of the 5-folds is given in Supplementary Material Table 3.

3.2 Machine learning

The trained models, and a guide on how to use them, can be found in an open-access code repository (32).

3.2.1 Logistic regression

The LRs were created using scikit-learn 0.23.2 (34), and liblinear was chosen as the optimisation algorithm. The inverse of regularisation strength (C-value) was chosen as a hyperparameter to be tuned, and found to have an optimal value of 0.1, with the remaining parameters set as default values as stated in (35).

3.2.2 Decision tree

The DTs were created using scikit-learn 0.23.2 (34), and 'classification and regression trees' (CART) was chosen as the tree algorithm. The maximum tree depth was chosen as a hyperparameter to be tuned, and found to have an optimal value of 4, with the remaining parameters set as default values as stated in (36).

3.2.3 Neural network

The NNs were created using PyTorch 1.8.1 (37) and run on an Nvidia 2070 Super GPU using CUDA 11.2. A feedforward network was created with a linear input layer of 8 neurons, 3 linear hidden layers with 12 neurons each, and a final linear output layer with one neuron, followed by a sigmoid activation function with a 0.5 threshold for CSFR classification. For the non-output layers, the 'rectified linear activation unit' (ReLu) was used as the activation function, with a 0.35 dropout. Binary cross-entropy was used as the loss function and 'stochastic gradient descent' (SGD) was used as the optimiser, with learning rate; momentum; batch size; and number of epochs hyperparameters to be tuned. A learning rate of 0.001; momentum of 0.9; batch size of 100; and number of epochs equalling 100 were found to be optimal.

3.3 Evaluation

3.3.1 AUC

From Figure 2 and Table 2, it can be seen that the NNs were able to predict the existence of CSFR across all prediction categories and approach categories with an AUC > 0.50 (an AUC of 0.50 is equivalent to a model that randomly predicts CSFR). Both LRs and DTs performances are outperformed by NNs, and for a few instances have an AUC < 0.5.

Comparing approach categories, it can be seen all three categories have similar NNs performances, but EEA performs worse than TSA for LRs. After mean-averaging across approach categories, and then comparing NNs performance across prediction categories, it can be seen risk factors slightly outperform repair techniques, which are in turn outperformed when all predictors (excluding surgical approach) are used. The inclusion of surgical approach as a predictor does not improve NN performance.

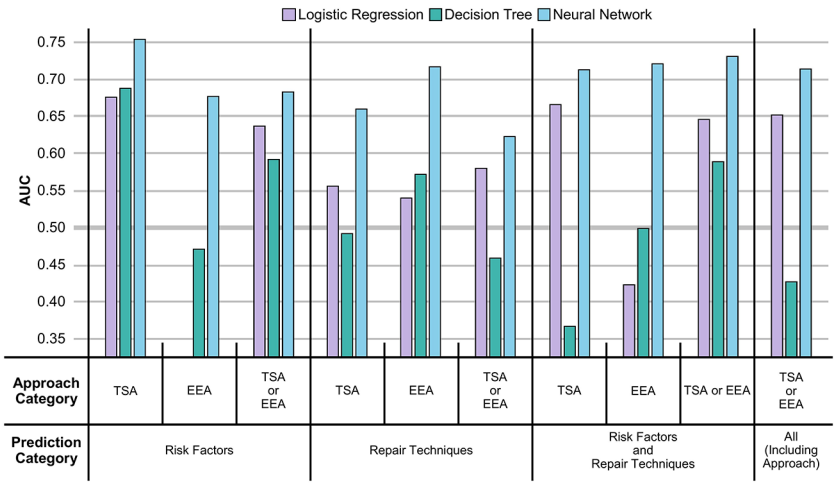


FIGURE 2
AUC of MLs displayed as a vertical bar chart. The AUC scale ranges from 0.35 to 0.75, with a thicker line at 0.50. Error bars representing the standard deviation across the 5-folds are not given. The AUC for LRs in the risk factors EEA case is not displayed as the AUC (0.22) is too low. The full values, including the standard deviation error bars, can be seen in Table 2.

TABLE 2 Performance of MLs.

Predictor Category	Surgical Category	ML Method	Training			Validation		
			AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
Risk Factors	TSA	LR	0.74±0.02	0.75±0.02	0.64±0.02	0.68±0.10	0.64±0.11	0.63±0.04
		DT	0.79±0.02	0.63±0.04	**0.79±0.01**	0.69±0.18	0.56±0.21	**0.79±0.02**
		NN	**0.83±0.02**	**0.78±0.12**	0.71±0.19	**0.75±0.08**	**0.69±0.17**	0.70±0.18
	EEA	LR	0.62±0.02	0.75±0.08	0.38±0.05	0.22±0.09	0.30±0.24	0.40±0.15
		DT	**0.87±0.05**	**0.93±0.10**	**0.68±0.12**	0.47±0.18	0.30±0.40	**0.62±0.06**
		NN	0.59±0.10	0.80±0.40	0.30±0.37	**0.68±0.08**	**0.80±0.40**	0.28±0.37
	TSA or EEA	LR	0.69±0.02	**0.69±0.05**	0.61±0.02	0.64±0.11	**0.65±0.19**	0.61±0.04
		DT	**0.83±0.01**	0.65±0.04	**0.83±0.04**	0.59±0.12	0.36±0.17	**0.81±0.03**
		NN	0.79±0.03	0.63±0.17	0.78±0.12	**0.68±0.08**	0.45±0.19	0.76±0.12
Repair Techniques	TSA	LR	0.68±0.04	0.62±0.07	0.61±0.09	0.56±0.14	0.43±0.26	0.59±0.09
		DT	**0.91±0.05**	**0.93±0.15**	**0.78±0.15**	0.49±0.22	0.10±0.20	**0.74±0.13**
		NN	0.74±0.08	0.73±0.20	0.61±0.23	**0.66±0.08**	**0.60±0.21**	0.61±0.23
	EEA	LR	**0.81±0.05**	**0.80±0.10**	0.64±0.07	0.54±0.16	0.40±0.37	0.56±0.12
		DT	0.79±0.04	0.75±0.03	**0.69±0.05**	0.57±0.06	0.53±0.13	**0.67±0.05**
		NN	0.76±0.11	0.75±0.39	0.59±0.29	**0.72±0.14**	**0.70±0.40**	0.50±0.37
	TSA or EEA	LR	0.69±0.01	0.70±0.04	0.59±0.04	0.58±0.06	**0.50±0.16**	0.59±0.04
		DT	**0.77±0.04**	**0.73±0.15**	0.68±0.08	0.46±0.07	0.35±0.25	0.68±0.12
		NN	0.77±0.05	0.72±0.20	**0.70±0.17**	**0.62±0.05**	0.49±0.22	**0.69±0.17**
Risk Factors and Repair Techniques	TSA	LR	0.79±0.01	0.73±0.04	0.68±0.01	0.67±0.09	**0.49±0.25**	0.67±0.06
		DT	0.86±0.04	0.88±0.08	0.72±0.05	0.37±0.16	0.20±0.24	0.67±0.11
		NN	**0.90±0.05**	**0.89±0.16**	**0.80±0.13**	**0.71±0.09**	0.49±0.29	**0.80±0.10**
	EEA	LR	**0.81±0.03**	**0.85±0.09**	0.59±0.05	0.42±0.20	0.40±0.37	0.47±0.07
		DT	0.75±0.02	0.76±0.07	0.61±0.06	0.50±0.07	0.41±0.10	0.59±0.05
		NN	0.79±0.10	0.58±0.38	**0.80±0.18**	**0.72±0.09**	**0.50±0.45**	**0.78±0.18**
	TSA or EEA	LR	0.75±0.01	0.70±0.04	0.66±0.01	0.65±0.10	0.57±0.21	0.64±0.04
		DT	0.84±0.01	0.78±0.15	**0.73±0.15**	0.59±0.13	0.47±0.17	**0.70±0.16**
		NN	**0.88±0.05**	**0.91±0.14**	0.67±0.17	**0.73±0.03**	**0.63±0.31**	0.64±0.17
All (Including Approach)	TSA or EEA	LR	0.76±0.01	0.73±0.03	0.67±0.01	0.65±0.09	**0.59±0.24**	0.65±0.03
		DT	0.91±0.03	**1.00±0.00**	**0.74±0.07**	0.43±0.19	0.20±0.40	0.67±0.07
		NN	**0.91±0.02**	0.96±0.06	0.69±0.06	**0.71±0.06**	0.57±0.13	**0.68±0.06**

Values are given to two decimal places in the form 'mean ± standard deviation' calculated over the 5-fold cross-validation. Bolded values highlight the best performing metric in the (subset, approach) category for that column's performance metric.

As seen in Table 2, a high AUC in the training dataset does not necessarily correspond to a high AUC in the validation dataset. In particular,
for DTs, the issue is exacerbated. For example, for the TSA repair techniques, a 0.86 training AUC translates to a 0.37 validation AUC for DT, compared to a 0.79 training AUC to 0.67 validation AUC translation for LR, or 0.90 training AUC to 0.71 validation AUC translation for NN.

3.3.2 SHAP

Figure 3 displays the SHAP scores for each predictor for two NNs (TSA risk factors and repair techniques; EEA risk factors and repair techniques). Supplementary Figures 1, 2 display the SHAP scores for the remaining eight NNs and Supplementary Table 2 shows the SHAP correlation coefficients for all ten NNs - consistent with the trends shown in Figure 3. Comparing approach categories, the SHAP scores are larger in magnitude for TSA than EEA. Comparing prediction

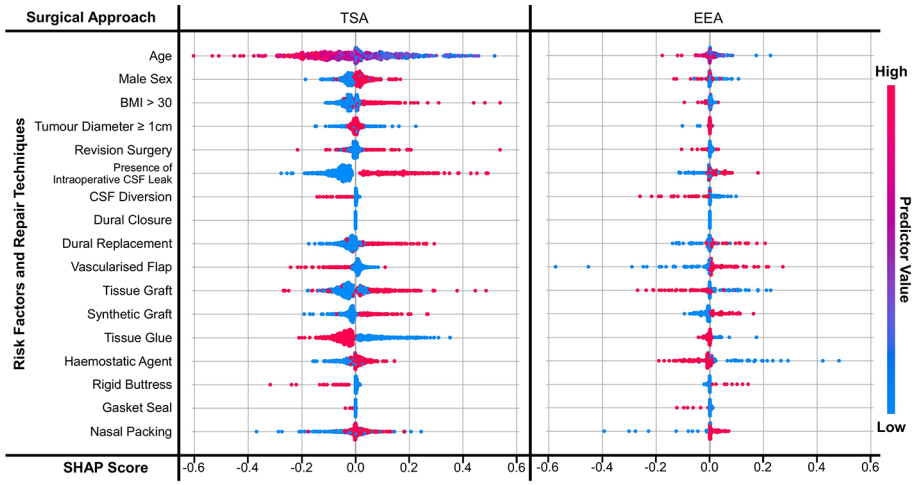


FIGURE 3
SHAP scores for predictors displayed as a bee diagram for the predictor category 'risk factors and repair techniques', where the NNs are split by approach. Scores are shown for each predictor across all 5-folds. As shown in the 'predictor value' legend – a high value is indicated in red, and a low value is indicated by blue; for binary variables this means red indicates a value of 1 (i.e. present) and blue indicates a value of 0 (i.e. not present).

categories, the SHAP scores for risk factors have a clearer boundary between not present and present than repair techniques.

Focusing on TSA risk factors, the presence of intraoperative CSF leak appears to be the strongest predictor of CSFR within the NN (Figure 3 and Supplementary Table 4). This is followed by younger age, elevated BMI, revision surgery, and male sex seem to also increase the probability of CSFR, albeit with a weaker correlation. EEA risk factors have a much smaller magnitude and weaker correlation, with intraoperative CSF leak having the strongest relative relationship with CSFR incidence (Figure 3 and Supplementary Table 4).

The impact of repair techniques on CSFR is less clear. In TSA, the use of CSF diversion, vascularised flaps, rigid buttresses +/- gasket sealing, and tissue glues appear to be protective against CSFR (Figure 3 and Supplementary Table 4). However, synthetic grafts, and to a lesser extent, dural replacement and tissue grafts appear to be associated with CSFR occurrence. For EEA, CSF diversion, gasket sealing, and to a lesser extent tissue grafts and haemostatic agents appear to reduce CSFR incidence. Synthetic grafts, vascularised flaps and dural replacement appear to be associated with CSFR occurrence.

4 Discussion

4.1 Principal findings

In this study, three ML methods were applied to a complex, multicentre, prospective skull base neurosurgery database encompassing CSFR and relevant predictor data (risk factors and intraoperative repair techniques).

Firstly, NNs outperformed LR and DT for CSFR prediction, with a mean AUC of 0.80 (0.70-0.90) for TSA and 0.78 (0.60-0.96) in EEA, when all risk factor and intraoperative repair data were integrated into the model. This is likely explained by NNs' known ability to learn complex non-linear relationships, even in the context of a large number of variables (27, 28). In this dataset, this likely reflects the use of multiple repair techniques synergistically and in layers, tailored to risk factors encountered on a case-by-case basis (10). NNs achieved this despite the class imbalance caused by a CSFR rate lower than the literature standard, with oversampling 5-fold validation (25, 28). Furthermore, there was an iterative improvement in NN performance with larger datasets, with TSA models (725 cases) generally outperforming EEA models (140 cases), and the use of risk factor data with intraoperative repair technique data improved CSFR prediction when compared with using a single data category.

Using SHAP scores, the relationship between predictor variables (risk factors and intraoperative repair techniques) was explored for their relative predictive value within NN models. The presence of intraoperative CSF leak was the most prominent risk factor for CSFR in TSA and EEA, which is in line with existing studies (7, 10, 20, 38, 39). The presence of elevated BMI and revision surgery were also associated with CSFR for the larger TSA dataset, again reflected in the literature (16). Modern repair regimes are tailored to risk factors, and this analysis consolidates pertinent factors to guide surgeons in repair technique decision-making (10).

When compared with traditional statistical models (e.g., multivariate logistic regression models), which suggested tissue glues alone may have a benefit in TSA, NN SHAP analysis has highlighted new potential relationships within the dataset, as well as reproducing the potential impact of tissue glues on CSFR rates (25). Specifically, CSF diversion and gasket sealing appear to be strong predictors of the absence of CSFR in both TSA and EEA – in line with RCT evidence (lumbar drainage) and numerous institutional series (gasket sealing) (10, 20, 40–42). Synthetic grafts and dural replacements (which often have overlapping materials) were associated with the development of CSFR in both TSA and EEA. Whilst autologous tissue repair had contradictory results depending on approach nasoseptal flaps (associated with CSFR in EEA, but protective against CSFR in TSA) and tissue grafts (associated with

CSFR in TSA, but protective against CSFR in EEA). The reasons for this are difficult to further ascertain within the NN structure, but theoretically may be due to differences in the groups of patients undergoing these repairs (for example, patients deemed at higher risk of CSFR at a baseline in EEA undergo nasoseptal flap) (25).

4.2 Comparison to literature

To our knowledge, only one other study has applied ML to a similar research question. However, this study examined intraoperative CSF leak (rather than postoperative CSFR), had a more imaging-centric dataset, was single centre (rather than 30 centres), and resultantly smaller volume (154 vs 855 cases) (29). Using a NN, Staartjes et al. were able to identify risk factors (higher Hardy grade, revision surgery, older age) whereas conventional statistical methods were unable to do so, echoing our experience in this study (29). There are however numerous studies utilising traditional statistical techniques in institutional case series in this field. Patel et al. use logistic regression models in a large volume single centre series, finding elevated BMI and hydrocephalus as significant risk factors for CSFR (43). Hannan et al. used similar methods and found that surgical experience, intraoperative CSF leak, Cushing's disease and the absence of nasoseptal flap use as CSFR risk factors (38). Similarly, Xue et al. highlighted intraoperative CSF leak as a key CSFR predictor and recommend nasoseptal flaps and lumbar drainage to decrease its incidence (39). Finally, Cai used a Least Absolute Shrinkage and Selection Operator (LASSO) model with multivariate logistic regression in a single centre moderate volume data set in the context of intraoperative CSF leak prediction, suggesting tumour size and preoperative albumin as key determinants (44).

4.3 Strengths and limitations

One of the strengths of this study is the large number of centres the data has come from, leading to data diversity, and hence improving the generalisability of the models. Overfitting was mitigated against in NNs using drop-out between layers, whilst evidence of this remained in LR and DT models (mismatch between training and validation datasets). More data (with more CSFR cases), from more countries, and an external validation dataset would be useful to improve model performance and generalisability further. Moreover, although our study is prospective with an internally validated dataset, observational studies inherently contain various types of bias, and so the correlations made may not be reflective of the overall population.

Another strength of the study is the large number and variety of predictors used, which improves model performance. On the other hand, the choice of predictors is also a limit, as other predictors, (such as type of tumour); or more granular versions of the predictors (such as intraoperative CSF leak grade rather than binary presence), have not been used. Furthermore, the range of ML models trialled, and the use of SHAP analysis, showing how and why NNs outperform LRs and DTs is a relative study strength. Nevertheless, the choice of NNs is limited to one simple architecture, and it is unknown whether more sophisticated architectures will improve performance in the future. Finally, this study shares the common issue of interpretability that many ML studies have, particularly the SHAP analysis, which may affect model usability and uptake by clinicians.

5 Conclusion

Three ML methods were applied to a complex, multicentre, prospective skull base neurosurgery database to predict CSFR following endonasal skull base surgery, and prediction variables that are most important for its development. NNs outperformed traditional statistical models and other ML models in CSFR prediction. NNs also uncovered relationships between risk factors and repair techniques on CSFR, which were otherwise not detected using traditional statistical approaches. These models will be improved further with larger and more granular datasets, improved NN architecture, and external validation. In the future, the next generation of these predictive models could be used to assist surgical decision-making and to support more individualised patient counselling.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Ethics statement

Formal institutional ethical board review and informed consent from human participants were not required owing to the nature of the study (seeking to evaluate local services as an observational study) and this was confirmed by the Health Research Authority, UK.

CRANIAL Consortium

Adrito Das*†, Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London, London; **Danyal Z Khan***†, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; **Danail Stoyanov**‡, Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London, London; **Hani J Marcus**‡, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; **Soham Bandyopadhyay**, Oxford University Global Surgery Group, Nuffield Department of Surgical Sciences, University of Oxford, Oxford; **Benjamin E Schroeder**, Department of Neurology, University Hospital of Wales, Cardiff University, Cardiff; **Vikesh Patel**, Division of Neurosurgery, Cambridge University Hospitals Trust, Cambridge; **Alice O'Donnell**, Birmingham Medical School, University of Birmingham, Birmingham; **Neurology and Neurosurgery Interest Group**, NANSIG; **British Neurosurgical Trainee Research Collaborative**, BNTRC; **Anastasios Giamouriadis**, Department of Neurosurgery, Aberdeen Royal Infirmary, Aberdeen; **Pragnesh Bhatt**, Department of Neurosurgery, Aberdeen Royal Infirmary, Aberdeen; **Bhaskar Ram**,

Department of Otorhinolaryngology, Aberdeen Royal Infirmary, Aberdeen; **Adithya Varma**, Department of Neurosurgery, Aberdeen Royal Infirmary, Aberdeen; **Philip Weir**, Department of Neurosurgery, Royal Victoria Hospital, Belfast; **Brendan Hanna**, Department of Otorhinolaryngology, Royal Victoria Hospital, Belfast; **Theodore C Hirst**, Department of Neurosurgery, Royal Victoria Hospital, Belfast; **Patrick McAleavey**, Department of Neurosurgery, Royal Victoria Hospital, Belfast; **Alessandro Paluzzi**, Department of Neurosurgery, Queen Elizabeth Hospital Birmingham, Birmingham; **Georgios Tsermoulas**, Department of Neurosurgery, Queen Elizabeth Hospital Birmingham, Birmingham; **Shahzada Ahmed**, Department of Otorhinolaryngology, Queen Elizabeth Hospital Birmingham, Birmingham; **Wai Cheong Soon**, Department of Neurosurgery, Queen Elizabeth Hospital Birmingham, Birmingham; **Yasir Arafat Chowdhury**, Department of Neurosurgery, Queen Elizabeth Hospital Birmingham, Birmingham; **Suhaib Abualsaud**, Department of Neurosurgery, Queen Elizabeth Hospital Birmingham, Birmingham; **Shumail Mahmood**, Department of Neurosurgery, Queen Elizabeth Hospital Birmingham, Birmingham; **Paresh Naik**, Department of Otorhinolaryngology, Queen Elizabeth Hospital Birmingham, Birmingham; **Zohra Haiderkhan** Department of Neurosurgery, Queen Elizabeth Hospital Birmingham, Birmingham; **Rafid Al-Mahfoudh**, Department of Neurosurgery, Hurstwood Park Neurosciences Centre and Royal Sussex County Hospital, Brighton; **Andrea Perera**, Department of Neurosurgery, Hurstwood Park Neurosciences Centre and Royal Sussex County Hospital, Brighton; **Mircea Rus**, Department of Neurosurgery, Hurstwood Park Neurosciences Centre and Royal Sussex County Hospital, Brighton; **Adam Williams**, Department of Neurosurgery, Southmead Hospital Bristol, Bristol; **Charles Hand**, Department of Neurosurgery, Southmead Hospital Bristol, Bristol; **Kumar Abhinav**, Department of Neurosurgery, Southmead Hospital Bristol, Bristol; **Cristina Cernei**, Department of Neurosurgery, Southmead Hospital Bristol, Bristol; **Aiman Dilnawaz**, Department of Neurosurgery, Southmead Hospital Bristol, Bristol; **Richard Mannion**, Division of Neurosurgery, Cambridge University Hospitals Trust, Cambridge; **Thomas Santarius**, Division of Neurosurgery, Cambridge University Hospitals Trust, Cambridge; **James Tysome**, Division of Otorhinolaryngology, Cambridge University Hospitals Trust, Cambridge; **Rishi Sharma**, Division of Otorhinolaryngology, Cambridge University Hospitals Trust, Cambridge; **Angelos G Kolias**, Division of Neurosurgery, Cambridge University Hospitals Trust, Cambridge; **Neil Donnelly**, Division of Otorhinolaryngology, Cambridge University Hospitals Trust, Cambridge; **Vikesh Patel**, Division of Neurosurgery, Cambridge University Hospitals Trust, Cambridge; **Ashwin Venkatesh**, Division of Neurosurgery, Cambridge University Hospitals Trust, Cambridge; **Caroline Hayhurst**, Department of Neurosurgery, University Hospital of Wales, Cardiff; **Amr Mohamed**, Department of Neurosurgery, University Hospital of Wales, Cardiff; **Benjamin Stew**, Department of Otorhinolaryngology, University Hospital of Wales, Cardiff; **Joseph Merola**, Department of Neurosurgery, University Hospital of Wales, Cardiff; **Setthasorn Zhi Yang Ooi**, Department of Neurosurgery, University Hospital of Wales, Cardiff; **Mahmoud Kamel**, Department of Neurosurgery, Cork University Hospitals, Ireland; **Mohammad Habibullah Khan**, Department of Otorhinolaryngology, Cork University Hospitals, Ireland; **Sahibzada Abrar**, Department of Neurosurgery, Cork University Hospitals, Ireland; **Christopher Mckeon**, Department of Neurosurgery, Cork University Hospitals, Ireland; **Daniel McSweeney**, Department of Neurosurgery, Cork University Hospitals, Ireland; **Mohsen Javadpour**, Department of Neurosurgery, National Neurosurgical Centre, Beaumont Hospital, Ireland; **Peter Lacy**, Department of Otorhinolaryngology, National Neurosurgical Centre, Beaumont Hospital, Ireland; **Daniel Murray**, Department of Neurosurgery, National Neurosurgical Centre, Beaumont Hospital, Ireland; **Elena Roman**, Department of Neurosurgery, National Neurosurgical Centre, Beaumont Hospital, Ireland; **Kismet Hossain-Ibrahim**, Department of Neurosurgery, Ninewells Hospital, Dundee; **Peter Ross**, Department of Otorhinolaryngology, Ninewells Hospital, Dundee; **David Bennett**, Department of Neurosurgery, Ninewells Hospital, Dundee; **Nathan McSorley**, Department of Neurosurgery, Ninewells Hospital, Dundee; **Adam Hounat**, Department of Neurosurgery, Ninewells Hospital, Dundee; **Patrick Statham**, Department of Clinical Neurosciences, BioQuarter, Edinburgh; **Mark Hughes**, Department of Clinical Neurosciences, BioQuarter, Edinburgh; **Alhafidz Hamdan**, Department of Clinical Neurosciences, BioQuarter, Edinburgh; **Caroline Scott**, Department of Clinical Neurosciences, BioQuarter, Edinburgh; **Jigi Moudgil-Joshi**, Department of Clinical Neurosciences, BioQuarter, Edinburgh; **Anuj Bahl**, Department of Neurosurgery, Hull University Teaching Hospitals, Hull; **Anna Bjornson**, Department of Neurosurgery, Hull University Teaching Hospitals, Hull; **Daniel Gatt**, Department of Neurosurgery, Hull University Teaching Hospitals, Hull; **Nick Phillips**, Department of Neurosurgery, Leeds Teaching Hospitals, Leeds; **Neeraj Kalra**, Department of Neurosurgery, Leeds Teaching Hospitals, Leeds; **Melissa Bautista**, Department of Neurosurgery, Leeds Teaching Hospitals, Leeds; **Seerat Shirazi**, Department of Neurosurgery, Leeds Teaching Hospitals, Leeds; **Catherine E Gilkes**, Department of Neurosurgery, The Walton Centre, Liverpool; **Christopher P Millward**, Department of Neurosurgery, The Walton Centre, Liverpool; **Ahmad MS Ali**, Department of Neurosurgery, The Walton Centre, Liverpool; **Dimitris Paraskevopoulos**, Department of Neurosurgery, Barts and The Royal London Hospital, London; **Jarnail Bal**, Department of Neurosurgery, Barts and The Royal London Hospital, London; **Samir Matloob**, Department of Neurosurgery, Barts and The Royal London Hospital, London; **Rhannon Lobo**, Department of Neurosurgery, Barts and The Royal London Hospital, London; **Nigel Mendoza**, Department of Neurosurgery, Charing Cross Hospital, London; **Ramesh Nair**, Department of Neurosurgery, Charing Cross Hospital, London; **Arthur Dalton**, Department of Neurosurgery, Charing Cross Hospital, London; **Adarsh Nadig**, Department of Neurosurgery, Charing Cross Hospital, London; **Lucas Hernandez**, Department of Neurosurgery, Charing Cross Hospital, London; **Nick Thomas**, Department of Neurosurgery, King's College Hospital, London; **Eleni Maratos**, Department of Neurosurgery, King's College Hospital, London; **Jonathan Shapey**, Department of Neurosurgery, King's College Hospital, London; **Sinan Al-Barazi**, Department of Neurosurgery, King's College Hospital, London; **Asfand Baig Mirza**, Department of Neurosurgery, King's College Hospital, London; **Mohamed Okasha**, Department of Neurosurgery, King's College Hospital, London; **Prabhjot Singh Malhotra**, Department of Neurosurgery, King's College Hospital, London; **Razna Ahmed**, Department of Neurosurgery, King's College Hospital, London; **Neil L Dorward**, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; **Joan Grieve**, Department of Neurosurgery,

National Hospital for Neurology and Neurosurgery, London; **Hani J Marcus**, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; **Parag Sayal**, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; **David Choi**, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; **Ivan Cabrilo**, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; **Hugo Layard Horsfall**, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London; **Jonathan Pollock**, Department of Neurosurgery, Barking, Havering & Redbridge University Hospitals, London; **Alireza Shoakazemi**, Department of Neurosurgery, Barking, Havering & Redbridge University Hospitals, London; **Oscar Maccormac**, Department of Neurosurgery, Barking, Havering & Redbridge University Hospitals, London; **Guru N K Amirthalingam**, Department of Neurosurgery, Barking, Havering & Redbridge University Hospitals, London; **Andrew Martin**, Department of Neurosurgery, St George's University Hospitals Trust, London; **Simon Stapleton**, Department of Neurosurgery, St George's University Hospitals Trust, London; **Florence Hogg**, Department of Neurosurgery, St George's University Hospitals Trust, London; **Daniel Richardson**, Department of Neurosurgery, St George's University Hospitals Trust, London; **Kanna Gnanalingham**, Department of Neurosurgery, Salford Royal Trust, Manchester; **Omar Pathmanaban**, Department of Neurosurgery, Salford Royal Trust, Manchester; **Daniel M Fountain**, Department of Neurosurgery, Salford Royal Trust, Manchester; **Raj Bhalla**, Department of Otorhinolaryngology, Salford Royal Trust, Manchester; **Cathal J Hannan**, Department of Neurosurgery, Salford Royal Trust, Manchester; **Annabel Chadwick**, Department of Neurosurgery, Salford Royal Trust, Manchester; **Alistair Jenkins**, Department of Neurosurgery, Royal Victoria Infirmary, Newcastle; **Claire Nicholson**, Department of Neurosurgery, Royal Victoria Infirmary, Newcastle; **Syed Shumon**, Department of Neurosurgery, Royal Victoria Infirmary, Newcastle; **Mohamed Youssef**, Department of Neurosurgery, Royal Victoria Infirmary, Newcastle; **Callum Allison**, Department of Neurosurgery, Royal Victoria Infirmary, Newcastle; **Graham Dow**, Department of Neurosurgery, Queen's Medical Centre Nottingham, Nottingham; **Iain Robertson**, Department of Neurosurgery, Queen's Medical Centre Nottingham, Nottingham; **Laurence Johann Glancz**, Department of Neurosurgery, Queen's Medical Centre Nottingham, Nottingham; **Murugan Sitaraman**, Department of Neurosurgery, Queen's Medical Centre Nottingham, Nottingham; **Ashwin Kumaria**, Department of Neurosurgery, Queen's Medical Centre Nottingham, Nottingham; **Ananyo Bagchi**, Department of Neurosurgery, Queen's Medical Centre Nottingham, Nottingham; **Simon Cudlip**, Department of Neurosurgery, John Radcliffe Hospital, Oxford University Hospitals, Oxford; **Jane Halliday**, Department of Neurosurgery, John Radcliffe Hospital, Oxford University Hospitals, Oxford; **Rory J Piper**, Department of Neurosurgery, John Radcliffe Hospital, Oxford University Hospitals, Oxford; **Alexandros Boukas**, Department of Neurosurgery, John Radcliffe Hospital, Oxford University Hospitals, Oxford; **Meriem Amarouche**, Department of Neurosurgery, John Radcliffe Hospital, Oxford University Hospitals, Oxford; **Damjan Veljanoski**, Department of Neurosurgery, John Radcliffe Hospital, Oxford University Hospitals, Oxford; **Samiul Muquit**, Department of Neurosurgery, University Hospitals Plymouth, Plymouth; **Ellie Edlmann**, Department of Neurosurgery, University Hospitals Plymouth, Plymouth; **Haritha Maripi**, Department of Neurosurgery, University Hospitals Plymouth, Plymouth; **Yi Wang**, Department of Neurosurgery, University Hospitals Plymouth, Plymouth; **Mehnaz Hossain**, Department of Neurosurgery, University Hospitals Plymouth, Plymouth; **Andrew Alalade**, Department of Neurosurgery, Lancashire Teaching Hospitals NHS Foundation Trust, Preston; **Syed Maroof**, Department of Neurosurgery, Lancashire Teaching Hospitals NHS Foundation Trust, Preston; **Pradnya Patkar**, Department of Neurosurgery, Lancashire Teaching Hospitals NHS Foundation Trust, Preston; **Saurabh Sinha**, Department of Neurosurgery, Royal Hallamshire Hospital & Sheffield Children's Hospital, Sheffield; **Showkat Mirza**, Department of Otorhinolaryngology, Royal Hallamshire Hospital & Sheffield Children's Hospital, Sheffield; **Duncan Henderson**, Department of Neurosurgery, Royal Hallamshire Hospital & Sheffield Children's Hospital, Sheffield; **Mohammad Saud Khan**, Department of Neurosurgery, Royal Hallamshire Hospital & Sheffield Children's Hospital, Sheffield; **Nijaguna Mathad**, Department of Neurosurgery, University Hospital Southampton, Southampton; **Jonathan Hempenstall**, Department of Neurosurgery, University Hospital Southampton, Southampton; **Difei Wang**, Department of Neurosurgery, University Hospital Southampton, Southampton; **Pavan Marwaha**, Department of Neurosurgery, University Hospital Southampton, Southampton; **Simon Shaw**, Department of Neurosurgery, Royal Stoke University Hospital, Stoke; **Georgios Solomou**, Department of Neurosurgery, Royal Stoke University Hospital, Stoke; **Alina Shrestha**, Department of Neurosurgery, Royal Stoke University Hospital, Stoke. †Joint first authorship, ‡Joint senior authorship.

Author contributions

This is a group authorship model paper where all authors contributed to data collection and approved the final manuscript. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication. Study design and analysis were performed by AD and DZK. The first draft of the manuscript was written by AD and DZK. Revision of manuscript and supervision was provided by HJM and DS.

Funding

This research was funded in whole, or in part, by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) [203145/Z/16/Z]; the Engineering and Physical Sciences Research Council (EPSRC) [EP/P027938/1, EP/R004080/1, EP/P012841/1]; and the Royal Academy of Engineering Chair in Emerging Technologies Scheme. AD is supported by EPSRC [EP/S201612/1]. HJM is supported by WEISS [NS/A000050/1] and by the National Institute for Health and Care Research (NIHR) Biomedical Research Centre at University College London. DZK is supported by the NIHR Academic Clinical Fellowship and the Cancer Research UK (CRUK) Predoctoral Fellowship. For the purpose of open access, the author has applied a CC BY public copyright licence to any author-accepted manuscript version arising from this submission.

Acknowledgments

The authors would like to thank the Neurology and Neurosurgery Interest Group (NANSIG) and the British Neurosurgical Trainee Research Collaborative (BNTRC) without which this study would not have been possible. A special thanks to the data validation team for ensuring data accuracy (Supplementary Material Table 2).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1046519/full#supplementary-material

References

1. Liu JK, Das K, Weiss MH, Laws ERJr., Couldwell WT. The history and evolution of transsphenoidal surgery. *J Neurosurg* (2001) 95(6):1083–96. doi: 10.3171/jns.2001.95.6.1083

2. Cappabianca P, Cavallo LM, de Divitiis E. Endoscopic endonasal transsphenoidal surgery. *Neurosurgery* (2004) 55(4):933–41. doi: 10.1227/01.NEU.0000137330.02549.0D

3. Dehdashti AR, Ganna A, Witterick I, Gentili F. Expanded endoscopic endonasal approach for anterior cranial base and suprasellar lesions: Indications and limitations. *Neurosurgery* (2009) 64(4):677–89. doi: 10.1227/01.NEU.0000339121.20101.85

4. Kassam A, Carrau RL, Snyderman CH, Gardner P, Mintz A. Evolution of reconstructive techniques following endoscopic expanded endonasal approaches. *Neurosurg Focus* (2005) 19(1):1–7. doi: 10.3171/foc.2005.19.1.9

5. Esquenazi Y, Essayed WI, Singh H, Mauer E, Ahmed M, Christos PJ, et al. Endoscopic endonasal versus microscopic transsphenoidal surgery for recurrent and/or residual pituitary adenomas. *World Neurosurg* (2017) 101:186–95. doi: 10.1016/j.wneu.2017.01.110

6. Fraser S, Gardner PA, Koutourousiou M, Kubik M, Fernandez-Miranda JC, Snyderman CH, et al. Risk factors associated with postoperative cerebrospinal fluid leak after endoscopic endonasal skull base surgery. *J Neurosurg* (2018) 128(4):1066–71. doi: 10.3171/2016.12.JNS1694

7. Hannan CJ, Kelleher E, Javadpour M. Methods of skull base repair following endoscopic endonasal tumor resection: A review. *Front Oncol* (2020) 10. doi: 10.3389/fonc.2020.01614

8. Harvey RJ, Parmar P, Sacks R, Zanation AM. Endoscopic skull base reconstruction of large dural defects: A systematic review of published evidence. *Laryngoscope* (2012) 122 (2):452–9. doi: 10.1002/lary.22475

9. Ivan C, Ann R, Craig B, Debi P. Complications of transsphenoidal surgery: Results of a national survey, review of the literature, and personal experience. *Neurosurgery* (1997) 40(2):225–37. doi: 10.1097/00041327-199712000-00044

10. Khan DZ, Ali AM, Koh CH, Dorward NL, Grieve J, Horsfall HL, et al. Skull base repair following endonasal pituitary and skull base tumour resection: A systematic review. *Pituitary* (2021) 42(5):698–713. doi: 10.1007/s11102-021-01145-4

11. Khan DZ, Muskens IS, Mekary RA, Zamanipoor Najafabadi AH, Helmy AE, Reisch R, et al. The endoscope-assisted supraorbital "keyhole" approach for anterior skull base meningiomas: An updated meta-analysis. *Acta Neurochir (Wien)* (2020) 163(3):661–76. doi: 10.1007/s00701-020-04544-x

12. Kono Y, Prevedello DM, Snyderman CH, Gardner PA, Kassam AB, Carrau RL, et al. One thousand endoscopic skull base surgical procedures demystifying the infection potential: Incidence and description of postoperative meningitis and brain abscesses. *Infect Control Hosp Epidemiol* (2011) 32(1):77–83. doi: 10.1086/657635

13. Liu P, Wu S, Li Z, Wang B. Surgical strategy for cerebrospinal fluid rhinorrhea repair. *Operative Neurosurg* (2010) 66(6):ons281–6. doi: 10.1227/01.NEU.0000369660.30126.02

14. Nishioka H, Haraoka J, Ikeda Y. Risk factors of cerebrospinal fluid rhinorrhea following transsphenoidal surgery. *Acta Neurochir (Wien)* (2005) 147(11):1163–6. doi: 10.1007/s00701-005-0586-3

15. Oakley GM, Orlandi RR, Woodworth BA, Batra PS, Alt JA. Management of cerebrospinal fluid rhinorrhea: An evidence-based review with recommendations. *Int Forum Allergy Rhinol* (2016) 6(1):17–24. doi: 10.1002/alr.21627

16. Patel MR, Stadler ME, Snyderman CH, Carrau RL, Kassam AB, Germanwala AV, et al. How to choose? endoscopic skull base reconstructive options and limitations. *Skull Base* (2010) 20(06):397–404. doi: 10.1055/s-0030-1253573

17. Strickland BA, Lucas J, Harris B, Kulubya E, Bakhsheshian J, Liu C, et al. Identification and repair of intraoperative cerebrospinal fluid leaks in endonasal transsphenoidal pituitary surgery: Surgical experience in a series of 1002 patients. *J Neurosurg* (2017) 129(2):425–9. doi: 10.3171/2017.4.JNS162451

18. Zamanipoor Najafabadi AH, Khan DZ, Muskens IS, Broekman MLD, Dorward NL, van Furth WR, et al. Trends in cerebrospinal fluid leak rates following the extended endoscopic endonasal approach for anterior skull base meningioma: A meta-analysis over the last 20 years. *Acta Neurochir (Wien)* (2020) 163(3):711–19. doi: 10.1007/s00701-020-04641-x

19. Zhang J, Liu J, Huang L. Factors influencing cerebrospinal fluid leaking following pituitary adenoma transsphenoidal surgery: A meta-analysis and comprehensive review. *BioMed Res Int* (2022) 2022:5213744. doi: 10.1155/2022/5213744

20. Zwagerman NT, Wang EW, Shin SS, Chang Y-F, Fernandez-Miranda JC, Snyderman CH, et al. Does lumbar drainage reduce postoperative cerebrospinal fluid leak after endoscopic endonasal skull base surgery? A prospective, randomized controlled trial. *J Neurosurg* (2018) 131(4):1172–8. doi: 10.3171/2018.4.JNS172447

21. Marcus HJ, Khan DZ, Borg A, Buchfelder M, Cetas JS, Collins JW, et al. Pituitary society expert Delphi consensus: Operative workflow in endoscopic transsphenoidal pituitary adenoma resection. *Pituitary* (2021) 24(4): 839–53. doi: 10.1007/s11102-021-01162-3

22. CRANIAL-Consortium. CSF rhinorrhoea after endonasal intervention to the skull base (CRANIAL) - part 1: Multicenter pilot study. *World Neurosurg* (2021) 149:e1077–e89. doi: 10.1016/j.wneu.2020.12.171

23. CRANIAL-Consortium. CSF rhinorrhea after endonasal intervention to the skull base (CRANIAL) — part 2: Impact of COVID-19. *World Neurosurg* (2021) 149:e1090–e7. doi: 10.1016/j.wneu.2020.12.169

24. Khan DZBS, Patel V, Schroeder BE, Cabrilo I, Choi D, Cudlip SA, et al. Neurology and neurosurgery interest group, British neurosurgical trainee research collaborative, CRANIAL steering committee. CSF rhinorrhoea after endonasal intervention to the anterior skull base (CRANIAL): Proposal for a prospective multicentre observational cohort study. *Br J Neurosurg* (2020) 35(4): 408–17. doi: 10.1080/02688697.2020.1795622

25. CRANIAL-Consortium. CSF rhinorrhoea after endonasal intervention to the skull base (CRANIAL): A multicentre prospective observational study. *Front Oncol* (2023) 12. doi: 10.3389/fonc.2022.1049627

26. Buchlak QD, Esmaili N, Leveque J-C, Farrokhi F, Bennett C, Piccardi M, et al. Machine learning applications to clinical decision support in neurosurgery: An artificial intelligence augmented systematic review. *Neurosurg Rev* (2020) 43(5):1235–53. doi: 10.1007/s10143-019-01163-8

27. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: Promises and perils. *Ann Surg* (2018) 268(1):70–6. doi: 10.1097/SLA.0000000000002693

28. Lammers DT, Eckert CM, Ahmad MA, Bingham JR, Eckert MJ. A surgeon's guide to machine learning. *Ann Surg Open* (2021) 2(3):e091. doi: 10.1097/AS9.0000000000000091

29. Staartjes VE, Zattra CM, Akeret K, Maldaner N, Muscas G, Bas van Niftrik CH, et al. Neural network-based identification of patients at high risk for intraoperative cerebrospinal fluid leaks in endoscopic pituitary surgery. *J Neurosurg* (2019) 133(2):329–35. doi: 10.3171/2019.4.Jns19477

30. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement. *Circulation* (2015) 131(2):211–9. doi: 10.1161/CIRCULATIONAHA.114.014508

31. *Python*. Available at: https://www.python.org/ (Accessed 01/08/2022).

32. *Das a: CSFR model code* (2022). Available at: https://github.com/dreets/csfr-public (Accessed 01/08/22).

33. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* (2017) 30:1–10.

34. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res* (2011) 12:2825–30. Available at: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

35. *Scikit-learn: Logistic regression*. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html (Accessed 01/08/22).

36. *Scikit-learn: Decision tree*. Available at: https://scikit-learn.org/stable/modules/tree.html (Accessed 01/08/22).

37. *Pytorch* . Available at: https://pytorch.org/ (Accessed 01/08/22).

38. Hannan CJ, Almhanedi H, Al-Mahfoudh R, Bhojak M, Looby S, Javadpour M. Predicting post-operative cerebrospinal fluid (CSF) leak following endoscopic transnasal pituitary and anterior skull base surgery: A multivariate analysis. *Acta Neurochir (Wien)* (2020) 162(6):1309–15. doi: 10.1007/s00701-020-04334-5

39. Xue H, Wang X, Yang Z, Bi Z, Liu P. Risk factors and outcomes of cerebrospinal fluid leak related to endoscopic pituitary adenoma surgery. *Br J Neurosurg* (2020) 34 (4):447–52. doi: 10.1080/02688697.2020.1754336

40. Carnevale C, Tomás-Barberán M, Til-Pérez G, Ibañez-Domínguez J, Arancibia-Tagle D, Rodríguez-Villalba R, et al. Endoscopic reconstruction of large anterior skull base defects with opening of the sellar diaphragm. experience at a tertiary level university hospital. *J Laryngol Otol* (2019) 133(10):889–94. doi: 10.1017/s0022215119001841

41. Garcia-Navarro V, Anand VK, Schwartz TH. Gasket seal closure for extended endonasal endoscopic skull base surgery: Efficacy in a large case series. *World Neurosurg* (2013) 80(5):563–8. doi: 10.1016/j.wneu.2011.08.034

42. Leng LZ, Brown S, Anand VK, Schwartz TH. "Gasket-seal" watertight closure in minimal-access endoscopic cranial base surgery. *Operative Neurosurg* (2008) 62(suppl_5): ONS342–ONS3. doi: 10.1227/01.neu.0000326017.84315.1f

43. Patel PN, Stafford AM, Patrinely JR, Smith DK, Turner JH, Russell PT, et al. Risk factors for intraoperative and postoperative cerebrospinal fluid leaks in endoscopic transsphenoidal sellar surgery. *Otolaryngology–Head Neck Surg* (2018) 158(5):952–60. doi: 10.1177/0194599818756272

44. Cai X, Zhu J, Yang J, Tang C, Yuan F, Cong Z, et al. Development and validation of nomogram to preoperatively predict intraoperative cerebrospinal fluid leakage in endoscopic pituitary surgery: A retrospective cohort study. *Front Oncol* (2021) 11:719494. doi: 10.3389/fonc.2021.719494

frontiers | Frontiers in Oncology

TYPE Perspective
PUBLISHED 23 June 2023
DOI 10.3389/fonc.2023.1063937

OPEN ACCESS

EDITED BY
Marco Scarpa,
University Hospital of Padua, Italy

REVIEWED BY
Riki Kawaguchi,
University of California, Los Angeles, United States

*CORRESPONDENCE
Ciaran Scott Hill
ciaran.hill@ucl.ac.uk

RECEIVED 07 October 2022
ACCEPTED 24 April 2023
PUBLISHED 23 June 2023

CITATION
Hill CS and Pandit AS (2023) Moving towards a unified classification of glioblastomas utilizing artificial intelligence and deep machine learning integration. *Front. Oncol.* 13:1063937. doi: 10.3389/fonc.2023.1063937

Moving towards a unified classification of glioblastomas utilizing artificial intelligence and deep machine learning integration

Ciaran Scott Hill[1,2]* and Anand S. Pandit[1,2]

[1]Institute of Neurology, University College London, London, United Kingdom, [2]Victor Horsley Department of Neurosurgery, National Hospital for Neurology and Neurosurgery (NHNN), London, United Kingdom

Glioblastoma a deadly brain cancer that is nearly universally fatal. Accurate prognostication and the successful application of emerging precision medicine in glioblastoma relies upon the resolution and exactitude of classification. We discuss limitations of our current classification systems and their inability to capture the full heterogeneity of the disease. We review the various layers of data that are available to substratify glioblastoma and we discuss how artificial intelligence and machine learning tools provide the opportunity to organize and integrate this data in a nuanced way. In doing so there is the potential to generate clinically relevant disease sub-stratifications, which could help predict neuro-oncological patient outcomes with greater certainty. We discuss limitations of this approach and how these might be overcome. The development of a comprehensive unified classification of glioblastoma would be a major advance in the field. This will require the fusion of advances in understanding glioblastoma biology with technological innovation in data processing and organization.

KEYWORDS

glioma, glioblastoma, classification, artificial intelligence, machine learning

Gliomas represent the most common primary brain cancer. They have distinct biological features and clinical behavior, and account for nearly 80% of the malignant brain tumors in adults (1, 2). The commonest subtype of glioma is glioblastoma a deadly brain cancer that is nearly universally fatal. Understanding of natural history, accurate prognostication, therapeutic efficacy, and the successful application of emerging precision medicine in glioblastoma relies upon the resolution and exactitude of classification. The WHO classification of Central Nervous System tumors began in 1970 (3). The first edition was largely based on anatomical and histological findings. Many of the major shifts in neuro-oncology and glioblastoma understanding over the intervening years have been represented in the subsequent WHO classification updates and the associated cIMPACT-NOW statements (4). A major conceptual leap was made in 2012 with the recognition of key subclassification of glioblastoma based on IDH mutation status (10.1038/nature10860.). This single mutation cleaved glioblastoma into two major subtypes with differing etiology, therapeutic vulnerability, and prognosis. In 2021 the significance of this stratification became codified by separating glioblastoma (IDH wild type) fully from grade 4 diffuse astrocytoma with IDH mutation (5).

In addition to the formal WHO classification there have been a multitude of differing stratifications of glioblastoma categorizations based on transcriptional profiles. A major development was in 2006 when Phillips et al. published a transcriptional classification of high-grade glioblastoma (6). This was advanced in 2010 when Verhaak et al. used data derived from The Cancer Genome Atlas to sub-stratify into 4 subgroups; proneural, mesenchymal, classical, and neural (7, 8). These were reported to have differing prognosis and treatment vulnerabilities. Further modifications and refinements to transcriptional groups, including single-cell profiling of both the tumor cells and microenvironmental components such as the neuro-immune niche have since been proposed by several groups including Neftel et al (9) and Richards et al (10). Another layer of complexity was added by epigenetic DNA-methylation profiling. This has already had significant clinical impact in supporting diagnosis and risk stratification (11, 12). Single-cell level profiling is not limited to transcriptional RNA profiling, it can also be applied across a range of biological analytic technologies including proteomic analysis - this opens up unparalleled levels of biological data.

Technologies in spatial-omics enable a greater understanding of cellular organizations and interactions within a tissue of interest. This is particularly of interest in cancer biology and can be applied at microscopic and super-resolution levels across the full range of spectral wavelengths and including spectroscopy data (13–16).

However, histology, selective mutations, transcriptional profiles and epigenetic changes do not tell the full story of glioblastoma diversity. One of the major barriers to successful new therapies in glioblastoma is considered the intra- and inter-heterogeneity of the tumors and this extends beyond these molecular sources of variability. In addition to transcriptional and epigenetic variability, anatomical location and structural features – including presence of cysts, degree of necrosis, proliferation indices etc (17), variability in radiomic findings (18), whole genome genetic/mutational characteristics (potentially including variability in extra-chromosomal sites) (19, 20) metabolic and lipidomic (21) and proteomics (22) can all be used to codify glioblastoma.

Integrating all these variables into a unified classification which reflects the diversity of glioblastoma states and in a clinically relevant manner, represents a daunting task (see Figure 1). Without this nuanced lamination we continue to risk masking the efficacy of new therapies by disease heterogeneity leading to variability of response. Likewise, our ability to accurately provide disease prognostication will remain limited.

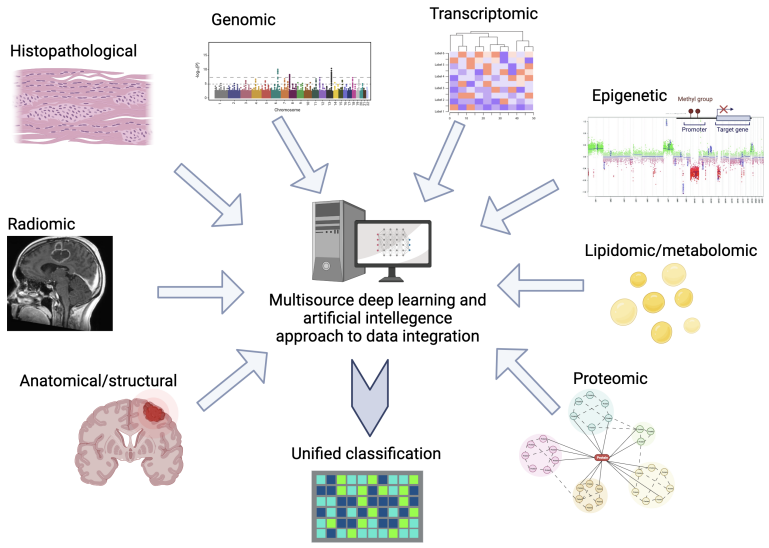


FIGURE 1
Example of multisource data input and integration for deep learning guided classification of glioblastoma.

Artificial intelligence (AI) tools provide the opportunity to organize and integrate these factors to generate clinically relevant disease sub-stratifications, which would help predict neuro-oncological patient outcomes with greater certainty. With enough data, DL (deep learning) methods based on neural networks have emerged as a leading approach for capturing highly informative features in oncology datasets. Using these tools, rapid progress has been made in each of the modalities described above. However, unanswered questions remain about how multimodal data can be integrated and a unified classification model be built.

A key requirement of multimodal integration is that each data source complements the others, enhancing information content beyond the scope of any single modality. For example, radiological data on macroscopic tumor morphology, as well as molecular and histological data, describe disease from different perspectives and scales. Each data source in a unified model should be at least partially orthogonal to the next.

While multimodal patient stratification methods have been developed for cancer patients in general, these mainly rely on *multi-omic* (multi-dimensional genomic data) in the absence of radiological or clinical information (23), and there currently exist few examples which utilize multimodal strategies for glioma patients specifically. Among these, there have been single-center studies which stratify glioma patients using multiparametric MRI, molecular and transcriptome information using kernel based learning (24), and deep learning approaches to predict survival which integrate both histological and genomic (but not radiological) information based on gliomas from The Cancer Genome Atlas (25, 26). These studies suggest that multimodal integration improves patient stratification and outcome prediction over unimodal methods.

Given its purported success, what is limiting this type of work? Many major limitations are simply related to the lack of availability of large, annotated datasets with multimodal information streams, which are sufficiently rich and class-balanced that the breadth of glioma heterogeneity can be encompassed. Other limitations pertain to how individual data modalities should be fused. It is unclear whether raw data should be concatenated from the start and used to train a single model. Or, alternatively, a composite model should be built from learned features, that are each derived from multiple single modality models (27).

In this regard, novel dimension-reduction and clustering methods (28), alongside other techniques which appropriately weigh will help in leveraging the vast amount of collected multimodal parameters for each patient and help prevent overfitting (29, 30). Finally, interpretation of deep learning models is notoriously difficult, and if clinicians want to understand how a unified model relates to the disease process, methods to make such models explainable are urgently needed.

Only by developing a comprehensive unified classification of glioblastoma can we optimize our prognostication and maximize the chance of precision therapies being successful. A system that allows integration of ever-increasing complexity and nuance will allow flexibility and adaption to new discoveries and therapies. Given the multiple layers of data involved in glioblastoma biology and their deep complexity and inter-related influence the consolidation and organization into a utilizable structure will require novel approaches. The application of artificial intelligence and deep machine learning in oncology is expanding at an explosive rate with numerous potential applications (31, 32). These technologies will be instrumental in achieving this final goal of a single unified classification of glioblastoma heterogeneity.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

CH conceived the manuscript. CH and AP co-wrote and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

Funding

CH is supported by a CRUK Pioneer Award, an AMS Starter Grant, a National Brain Appeal Innovation Award, and BRC (UCL/NIHR) funding.

Conflict of interest

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

Publisher's note

References

1. Ostrom QT, Gittleman H, Truitt G, Boscia A, Kruchko C, Barnholtz-Sloan JS. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2011–2015. *Neuro-oncology* (2018) 20:iv1–iv86. doi: 10.1093/neuonc/noy131

2. Weller M, Wick W, Aldape K, Brada M, Berger M, Pfister SM, et al. Glioma. *Nat Rev Dis Primers* (2015) 1:15017. doi: 10.1038/nrdp.2015.17

3. Scheithauer BW. Development of the WHO classification of tumors of the central nervous system: a historical perspective. *Brain Pathol* (2009) 19:551–64. doi: 10.1111/j.1750-3639.2008.00192.x

4. Gonzalez Castro LN, Wesseling P. The cIMPACT-NOW updates and their significance to current neuro-oncology practice. *Neuro-Oncol Practice* (2021) 8:4–10. doi: 10.1093/nop/npaa055

5. Komori T. The 2021 WHO classification of tumors, central nervous system tumors: the 10 basic principles. *Brain Tumor Pathol* (2022) 39(2), 47–50. doi: 10.1007/s10014-022-00428-3

6. Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* (2006) 9(3):157–73. doi: 10.1016/j.ccr.2006.02.019

7. Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* (2010) 17(1):98–110. doi: 10.1016/j.ccr.2009.12.020

8. Sidaway P. Glioblastoma subtypes revisited. *Nat Rev Clin Oncol* (2017) 14:587–7. doi: 10.1038/nrclinonc.2017.122

9. Neftel C, Laffy J, Filbin MG, Hara T, Shore ME, Rahme GJ, et al. An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* (2019) 178 (4):835–849.e21. doi: 10.1016/j.cell.2019.06.024

10. Richards LM, Whitley OKN, MacLeod G, Cavalli FMG, Coutinho FJ, Jaramillo JE, et al. Gradient of developmental and injury response transcriptional states defines functional vulnerabilities underpinning glioblastoma heterogeneity. *Nat Cancer* (2021) 2(2):157–73. doi: 10.1038/s43018-020-00154-9

11. Jaunmuktane Z, Capper D, Jones DTW, Schrimpf D, Sill M, Dutt M, et al. Methylation array profiling of adult brain tumours: diagnostic outcomes in a large, single centre. *Acta Neuropathol Commun* (2019) 7(1):24. doi: 10.1186/s40478-019-0668-8

12. Klughammer J, Kiesel B, Roetzer T, Fortelny N, Nemc A, Nenning KH, et al. The DNA methylation landscape of glioblastoma disease progression shows extensive heterogeneity in time and space. *Nat Med* (2018) 24(10):1611–24. doi: 10.1038/s41591-018-0156-x

13. Dey N, Li S, Bermond K, Heintzman R, Curcio CA, Ach T 3, et al. Multi-modal image fusion for multispectral super-resolution in microscopy. *Proc SPIE Int Soc Opt Eng* (2019) 10949:109490D. doi: 10.1117/12.2512598

14. Manni F, van der Sommen F, Fabelo H, Zinger S, Shan C, Edström E, et al. Hyperspectral imaging for glioblastoma surgery: improving tumor identification using a deep spectral-spatial approach. *Sensors (Basel)* (2020) 20(23):6955. doi: 10.3390/s20236955

15. Lam KHB, Leon AJ, Hui W, Lee SC, Batruch I, Faust K, et al. Topographic mapping of the glioblastoma proteome reveals a triple-axis model of intra-tumoral heterogeneity. *Nat Commun* (2022) 13(1):116. doi: 10.1038/s41467-021-27667-w

16. Li X, Strasser B, Neuberger U, Vollmuth P, Bendszus M, Wick W, et al. Deep learning super-resolution magnetic resonance spectroscopic imaging of brain metabolism and mutant isocitrate dehydrogenase glioma. *Neuro-Oncol Adv* (2022) 4 (1):vdac071. doi: 10.1093/noajnl/vdac071

17. Latini F, Fahlström M, Berntsson SG, Larsson E-M, Smits A, Ryttlefors M. A novel radiological classification system for cerebral gliomas: the brain-grid. *PloS One* (2019) 14:e0211243. doi: 10.1371/journal.pone.0211243

18. Cho H-h, Lee S-h, Kim J, Park H. Classification of the glioma grading using radiomics analysis. *PeerJ* (2018) 6:e5982. doi: 10.7717/peerj.5982

19. Melin BS, Barnholtz-Sloan JS, Wrensch MR, Johansen C, Il'yasova D, Kinnersley B, et al. Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nat Genet* (2017) 49(5):789–94. doi: 10.1038/ng.3823

20. Karami Fath M, Karimfar N, Fazlollahpour Naghibi A, Shafa S, Ghasemi Shiran M, Ataei M, et al. Revisiting characteristics of oncogenic extrachromosomal DNA as mobile enhancers on neuroblastoma and glioma cancers. *Cancer Cell Int* (2022) 22 (1):200. doi: 10.1186/s12935-022-02617-8

21. Garofano L, Migliozzi S, Oh YT, D'Angelo F, Najac RD, Ko A, et al. Pathway-based classification of glioblastoma uncovers a mitochondrial subtype with therapeutic vulnerabilities. *Nat Cancer* (2021) 2(2):141–56. doi: 10.1038/s43018-020-00159-4

22. Yanovich-Arad G, Ofek P, Yeini E, et al. Proteogenomics of glioblastoma associates molecular patterns with survival. *Cell Rep* (2021) 34:108787. doi: 10.1016/j.celrep.2021.108787

23. Ramazzotti D, Lal A, Wang B, Batzoglou S, Sidow A. Multi-omic tumor data reveal diversity of molecular mechanisms that correlate with survival. *Nat Commun* (2018) 9(1):4453. doi: 10.1038/s41467-018-06921-8

24. Rathore S, Akbari H, Rozycki M, Abdullah KG, Nasrallah MP, Binder ZA, et al. Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1. *Sci Rep* (2018) 8(1):5087. doi: 10.1038/s41598-018-22739-2

25. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci USA* (2018) 115(13):E2970–9. doi: 10.1073/pnas.1717139115

26. Chen RJ, Lu MY, Wang J, Williamson DFK, Rodig SJ, Lindeman NI, et al. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Trans Med Imag* (2022) 41(4):757–70. doi: 10.1109/TMI.2020.3021387

27. Boehm KM, Khosravi P, Vanguri R, Gao J, Shah SP. Harnessing multimodal data integration to advance precision oncology. *Nat Rev Cancer* (2022) 22:114–26. doi: 10.1038/s41568-021-00408-3

28. Cheerla A, Gevaert O. Deep learning with multimodal representation for pancancer prognosis prediction. *Bioinformatics* (2019) 35:i446–54. doi: 10.1093/bioinformatics/btz342

29. Liang C, Li Y, Luo J. A novel method to detect functional microRNA regulatory modules by bicliques merging. *IEEE/ACM Trans Comput Biol Bioinform* (2016) 13:549–56. doi: 10.1109/TCBB.2015.2462370

30. Mao XG, Xue XY, Wang L, Lin W, Zhang X. Deep learning identified glioblastoma subtypes based on internal genomic expression ranks. *BMC Cancer* (2022) 22:86. doi: 10.1186/s12885-022-09191-2

31. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med* (2021) 13:152. doi: 10.1186/s13073-021-00968-x

32. Nagy M, Radakovich N, Nazha A. Machine learning in oncology: what should clinicians know. *JCO Clin Cancer Inform* (2020) 4:799–810. doi: 10.1200/CCI.20.00049

Frontiers in Oncology

Advances knowledge of carcinogenesis and tumor progression for better treatment and management

The third most-cited oncology journal, which highlights research in carcinogenesis and tumor progression, bridging the gap between basic research and applications to imrpove diagnosis, therapeutics and management strategies.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact