Applying artificial intelligence to big data in hepatopancreatic and biliary surgery: a scoping review

Citation for published version:

Digital Object Identifier (DOI):
10.20517/ais.2022.39

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Artificial Intelligence Surgery

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Applying artificial intelligence to big data in hepatopancreatic and biliary surgery: a scoping review

Kieran G. McGivern1,8, Thomas M. Drake2,3,8, Stephen R. Knight2,8, James Lucocq6, Miguel O. Bernabeu2,8, Neil Clark2, Cameron Fairfield2, Riinu Pius1, Catherine A. Shaw1, Sohan Seth6, Ewen M. Harrison2,3

1NHS Greater Glasgow and Clyde, Royal Alexandra Hospital, Paisley PA2 9PN, UK.  
2Centre for Medical Informatics, Usher Institute, University of Edinburgh, Edinburgh EH16 4UX, UK.  
3Department of Clinical Surgery, University of Edinburgh, Edinburgh EH16 4SA, UK.  
4NHS Fife, Victoria Hospital, Kirkcaldy KY2 5AH, UK.  
5The Bayes Centre, The University of Edinburgh, Edinburgh EH8 9BT, UK.  
6Institute for Adaptive and Neural Computation, School of Informatics, The University of Edinburgh, Edinburgh EH8 9AB, UK.  
8Authors contributed equally.

Correspondence to: Prof. Ewen M. Harrison, Centre for Medical Informatics, Usher Institute, University of Edinburgh, 51 Little France Crescent, Edinburgh EH16 4SA, UK. E-mail: ewen.harrison@ed.ac.uk


Received: 8 Dec 2022  First Decision: 14 Feb 2023  Revised: 10 Mar 2023  Accepted: 15 Mar 2023  Published: 27 Mar 2023

Abstract

Aim: Artificial Intelligence (AI) and its applications in healthcare are rapidly developing. The healthcare industry generates ever-increasing volumes of data that should be used to improve patient care. This review aims to examine the use of AI and its applications in hepatopancreatic and biliary (HPB) surgery, highlighting studies leveraging large datasets.

Methods: A PRISMA-ScR compliant scoping review using Medline and Google Scholar databases was performed (5th August 2022). Studies focusing on the development and application of AI to HPB surgery were eligible for inclusion. We undertook a conceptual mapping exercise to identify key areas where AI is under active development for use in HPB surgery. We considered studies and concepts in the context of patient pathways - before surgery (including diagnostics), around the time of surgery (supporting interventions) and after surgery (including prognostication).
Results: 98 studies were included. Most studies were performed in China or the USA (n = 45). Liver surgery was the most common area studied (n = 51). Research into AI in HPB surgery has increased rapidly in recent years, with almost two-thirds published since 2019 (61/98). Of these studies, 11 have focused on using “big data” to develop and apply AI models. Nine of these studies came from the USA and nearly all focused on the application of Natural Language Processing. We identified several critical conceptual areas where AI is under active development, including improving preoperative optimization, image guidance and sensor fusion-assisted surgery, surgical planning and simulation, natural language processing of clinical reports for deep phenotyping and prediction, and image-based machine learning.

Conclusion: Applications of AI in HPB surgery primarily focus on image analysis and computer vision to address diagnostic and prognostic uncertainties. Virtual 3D and augmented reality models to support complex HPB interventions are also under active development and likely to be used in surgical planning and education. In addition, natural language processing may be helpful in the annotation and phenotyping of disease, leading to new scientific insights.

Keywords: Artificial Intelligence, big data, surgery, liver, pancreas, biliary

INTRODUCTION

Artificial Intelligence (AI) encompasses a range of computational approaches with the central aim of developing algorithms to process and interpret information. AI methods can be applied to various input data types ranging from tabular datasets and images to multimedia and text. Although termed “intelligence”, these algorithms are in no sense conscious or able to employ “rational thought”, but in most cases, reflect model parameters derived exclusively from input data. Within AI, there are three overlapping fields that arguably have the most potential for HPB surgery: machine learning (ML), computer vision (CV) and natural language processing (NLP). ML uses algorithms to learn, adapt, and draw inferences from patterns in training data. CV allows for supervised or unsupervised image analysis, allowing for features of interest in images to be identified and characterized. For text-based sources of data written as prose or in a “human-readable” format (e.g., radiology or pathology reports), NLP allows computers to interpret human text or spoken language communication[1-5].

The specific areas and applications of AI most likely to deliver a positive impact on patient care currently need to be clarified, as are the barriers limiting the uptake of AI approaches into clinical practice. In 2021 Bari et al. described the applications of AI in hepatopancreatic and biliary (HPB) surgery, proposing the framework of preoperative, intraoperative, and postoperative AI applications. We have adopted this structure for this review[6].

With the increased availability of structured and unstructured healthcare datasets, the opportunity for AI-based approaches widens. Policymakers, healthcare providers, and industry are exploring new AI approaches, seeking to utilize data across a range of applications, including improving outcomes, optimizing the patient experience, and providing cost-effectiveness in delivering care at the health system level[7-9]. In this review, we aim to outline the fundamental AI approaches to pressing questions in HPB surgery, identifying where AI is most likely to have an impact in future patient care.

METHODS

This scoping review was performed in accordance with the PRISMA-ScR guidelines for scoping reviews[10]. The Medline database was searched systematically using the following Medical Subject Headings (MeSH) search terms to ensure the identification of appropriate articles; “Algorithms.mp. or algorithm/” AND
“surgery/ or biliary tract surgery/ or liver surgery/ or pancreas surgery/”. Articles were limited to English language and those published from 2012 onwards to provide contemporary studies that were likely reflective of current approaches in AI. Further supplementary searches were performed using citation lists and the Google Scholar database. The last search was conducted on 5th August 2022.

We defined “HPB surgery”, as the surgical management of benign and malignant diseases of the liver, pancreas, gallbladder, and bile ducts. “Artificial intelligence” refers to the use of various algorithmic methods which could be applied to interpret or process information. We further assessed the identified papers for the element of AI primarily used i.e., machine/deep learning, computer vision or natural language processing[1-5].

Following the literature search, article titles and abstracts were screened by three reviewers (KMcG, SRK, JL) and those meeting the inclusion criteria underwent full-text review. Any disagreements were resolved by consensus within the group. References from included articles were searched to identify any other relevant articles. Conference abstracts were screened to assist in identifying related full-text articles before inclusion. Where more than one article was published from a single data set, the article analyzing the largest cohort of patients was included.

Data were extracted independently using a standardized pro forma. This included the aim of the study, methodology, year of publication, countries represented, the primary organ of focus, AI methods employed, and the number of patients (where applicable). Identified publications were further interrogated to find a shared focus on diagnostics, prognostics, or intervention, allowing further subdivision of the presented research. We then undertook a conceptual mapping exercise to identify areas of crucial importance.

We used a pragmatic approach to further select studies with a sample size equal to or greater than five thousand that satisfied the “velocity, volume and variety” of data points needed to be considered as “big data”. A similar approach used previously, albeit with smaller datasets, acted as a benchmark[7-9,11]. Any disagreements on the selection of these papers were resolved by group consensus.

The Covidence online toolkit was used throughout the data collection and extraction stages of this scoping review[12].

RESULTS
Scoping search results
The search identified 5,221 articles, of which 134 were fully assessed for eligibility. A further 63 articles were identified from article citation lists or by the supplementary search of the Google Scholar database [Figure 1]. Following assessment, 98 studies[13-110] were included in this review, with most studies excluded due to being in conference abstract form only (n = 84).

Characteristics of included studies
Identified studies had a wide geographical distribution, coming from a total of 24 countries, with the majority from China or the USA (45/98). No papers identified originated from the African continent [Figure 2]. Studies on the use of AI in surgical conditions of the liver predominated (n = 51). Research on pancreatic and biliary conditions (n = 23) was included at a comparable frequency to one another. We noted a rapid increase in the number of studies published over the past three years, with almost two-thirds of the identified papers (n = 61) published since 2019 [Figure 3].
Figure 1. PRISMA flow chart detailing study selection and exclusion.

Figure 2. Global distribution of studies on the use of AI in HPB Surgery.

Figure 3. Primary organ of interest over time demonstrating the increase in frequency of publication.
Studies identified were subdivided into groups focusing on diagnostics, prognostics, and interventions. We assessed 23 papers reporting diagnostic uses of AI in HPB surgery. Of these, five focused on the gallbladder, 11 on the liver, and seven on the pancreas. Twenty-nine studies reported prognostication using AI, of which three focused on the gallbladder, 16 on the liver, one on the liver and pancreas, and nine on the pancreas alone. Almost half of the studies identified reported on the interventional use of AI in HPB surgery (n = 46), with 24 studies focusing on the liver, 19 on the gallbladder alone or in conjunction with another organ (n = 4), and three studies looking at the pancreas. A summary of the papers subdivided into the diagnostic, prognostic and intervention cohorts can be found in Tables 1, 2, and 3, respectively.

Regarding sample size, most studies (n = 13) reporting diagnostic applications of AI in HPB surgery utilized data from fewer than 1,000 patients. The smallest number of patients in a focused study of the ultrasound-based classification of liver lesions was 22. Three studies included over 5,000 patients and were included in our big data cohort. Most studies (n = 16) looking at prognostic uses of AI in HPB surgery had fewer than 500 patients. Two studies had fewer than 5,000 patients, but were included in our big data cohort due to the high number of images and image reports included. Eleven studies looking at interventional uses of AI in HPB surgery did not use actual patient data, but used simulations-based approaches. There was little mention or use of “explainable AI” concepts in any of the included studies.

**Conceptual mapping of AI research in HPB surgery**

Following data extraction and study classification, we undertook a conceptual mapping exercise to identify key areas and relationships in AI use. Many of the identified concepts involved outcome prediction (such as the risk of complication, or personalized survival predictions). Others utilized AI to support clinicians in the identification of a condition before, during, or after surgery (such as identifying malignancy, identifying complications early, or even the prevention of these by using AI to alert clinicians to unseen structures intraoperatively). Preoperative planning and surgical simulation were particularly key areas within the intervention grouping. Finally, within the conceptual mapping exercise, we identified several areas where AI may be useful as either a risk stratification tool or as an intervention in future research.

**Diagnostic applications of artificial intelligence**

Diagnostic applications of AI primarily involved interpreting images using computer vision models. AI was used across a range of imaging modalities, including transabdominal ultrasound, endoscopic ultrasound, MRI and CT, to identify lesions or classify lesions into different radiomic subgroups of disease. Although the majority of preoperative, diagnostic AI work focused on imaging, there were studies investigating perioperative risk prediction. However, there were no studies that proposed to use AI as an intervention in preoperative care pathways. Therefore, it should be considered that preoperative AI may also be undertaken with a broader surgical focus, rather than specifically targeted at HPB populations and hence are not discussed in this review.

**Prognostic applications of artificial intelligence**

The majority of prognostic applications for AI were in the prediction of cancer recurrence and survival. This was achieved using a variety of input data, including imaging, genetics, and clinical characteristics. Prediction models were developed for a variety of time points, including the first 30 days following surgery and for longer-term survival.
Table 1. Summary of included studies focusing on diagnostic uses of AI in HPB surgery

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year of publication</th>
<th>Location</th>
<th>Organ</th>
<th>AI method</th>
<th>Aim</th>
<th>Method</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saftoiu et al.</td>
<td>2012</td>
<td>Romania</td>
<td>P</td>
<td>DL/CV</td>
<td>Assessed accuracy of real-time EUS elastography in pancreatic lesions using artificial neural network analysis</td>
<td>Prospective, blinded, multicentric study</td>
<td>EUS images</td>
</tr>
<tr>
<td>Kaizhi et al.</td>
<td>2014</td>
<td>Japan</td>
<td>L</td>
<td>DL/CV</td>
<td>Proposes automatic classification method based on deep learning in contrast-enhanced ultrasonography (CEUS) of focal liver lesions</td>
<td>Case series</td>
<td>CEUS images</td>
</tr>
<tr>
<td>Gatos et al.</td>
<td>2015</td>
<td>Greece</td>
<td>L</td>
<td>ML/CV</td>
<td>Design and implementation of a computer-based image analysis system employing the support vector machine system for the classification of liver lesions</td>
<td>Retrospective study</td>
<td>MRI images</td>
</tr>
<tr>
<td>Roch et al.</td>
<td>2015</td>
<td>USA</td>
<td>P</td>
<td>NLP</td>
<td>Implement an automated Natural Language Processing based pancreatic cyst identification system</td>
<td>Single institution prospective pilot study</td>
<td>Patient records</td>
</tr>
<tr>
<td>Sada et al.</td>
<td>2016</td>
<td>USA</td>
<td>L</td>
<td>NLP</td>
<td>Evaluated whether natural language processing document classification improves HCC identification</td>
<td>Retrospective study</td>
<td>Pathology/radiology reports</td>
</tr>
<tr>
<td>Kondo et al.</td>
<td>2017</td>
<td>Japan</td>
<td>L</td>
<td>ML/CV</td>
<td>Proposes automatic classification method based on machine learning in CEUS of focal liver lesions</td>
<td>Single institution pilot study</td>
<td>CEUS images</td>
</tr>
<tr>
<td>Yang et al.</td>
<td>2017</td>
<td>China</td>
<td>L</td>
<td>NLP</td>
<td>Assess gene expression in HCC using combined data from The Cancer Genome Atlas and NLP identified genes</td>
<td>Description of experiment</td>
<td>Gene library/published literature</td>
</tr>
<tr>
<td>Kuwahara et al.</td>
<td>2019</td>
<td>Japan</td>
<td>P</td>
<td>DL</td>
<td>Investigate whether a deep learning algorithm using EUS images of IPMN could predict the diagnosis of malignancy</td>
<td>Retrospective study</td>
<td>EUS images</td>
</tr>
<tr>
<td>Shen et al.</td>
<td>2019</td>
<td>China</td>
<td>P</td>
<td>ML</td>
<td>Establish and validate a radiomics diagnosis model for the classification of three subtypes of pancreatic lesion</td>
<td>Retrospective study</td>
<td>CT images</td>
</tr>
<tr>
<td>Lei Xu et al.</td>
<td>2019</td>
<td>China/USA</td>
<td>G</td>
<td>ML/CV</td>
<td>Develop and validate a prediction model for preoperative LN status evaluation in ICC patients</td>
<td>Retrospective study</td>
<td>MRI images</td>
</tr>
<tr>
<td>Brown et al.</td>
<td>2019</td>
<td>Canada</td>
<td>L</td>
<td>NLP/ML</td>
<td>Explore natural language processing to predict downstream radiology resource utilization in patients undergoing surveillance for HCC</td>
<td>Retrospective study</td>
<td>Radiology reports</td>
</tr>
<tr>
<td>Watson et al.</td>
<td>2020</td>
<td>USA</td>
<td>P</td>
<td>DL</td>
<td>Use CT-guided deep learning techniques to predict malignancy of PCNs</td>
<td>Retrospective pilot study</td>
<td>CT images</td>
</tr>
<tr>
<td>Liu et al.</td>
<td>2020</td>
<td>China</td>
<td>L</td>
<td>NLP/DL</td>
<td>Designed an NLP pipeline for the direct extraction of clinically relevant features of liver cancer from radiology reports</td>
<td>Retrospective study</td>
<td>Radiology reports</td>
</tr>
<tr>
<td>Mao et al.</td>
<td>2021</td>
<td>China</td>
<td>L</td>
<td>ML</td>
<td>Investigate the performance of an ultrasound-based radiomics approach to differentiate primary liver cancer from metastatic liver cancer</td>
<td>Retrospective study</td>
<td>US images</td>
</tr>
<tr>
<td>Jang et al.</td>
<td>2021</td>
<td>South Korea</td>
<td>G</td>
<td>DL/CV</td>
<td>Evaluate the diagnostic performance of AI in differentiating biliary lesions using EUS images</td>
<td>Retrospective study</td>
<td>EUS images</td>
</tr>
<tr>
<td>Dongyan et al.</td>
<td>2021</td>
<td>China</td>
<td>G</td>
<td>DL/CV</td>
<td>Assessed duodenoscopy assisted</td>
<td>Pilot study</td>
<td>ERCP/</td>
</tr>
</tbody>
</table>
Interventional applications of artificial intelligence

We identified several key concepts around supporting interventions with AI assistance [Table 3 and Figure 4]. Intraoperative vision was a major area, with multiple studies focusing on improving the visualization of unseen structures, which may cause significant patient harm if inadvertently injured (e.g., major blood vessels or the bile duct). This was achieved through virtual or augmented reality, where inputs from other data sources such as CT and MRI are combined (sensor fusion) and overlaid on real-time images (e.g., through laparoscopic/robot-assisted surgery video source) to produce an augmented view of the surgical field.

Preoperative surgical planning and simulation were also identified as key concepts. There were numerous studies that aimed to develop virtual reality models or other digital interventions which permitted surgeons to plan complex operations with the aim of minimizing complications. This was proposed to be achieved through pre-surgery operative simulation/rehearsal (advantages when unusual anatomy identified) or by using AI methods to predict severe complications such as post-hepatectomy liver failure (PHLF).

**Artificial intelligence tasks**

We identified several common AI tasks being applied in HPB surgery. Classification is where data can be assigned to groups based on a defined shared characteristic. Classification algorithms were frequently
<table>
<thead>
<tr>
<th>Authors</th>
<th>Year of publication</th>
<th>Location</th>
<th>Organ</th>
<th>AI method</th>
<th>Aim</th>
<th>Design</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singal et al.</td>
<td>2013</td>
<td>USA</td>
<td>L</td>
<td>ML</td>
<td>Develop and compare predictive models for HCC development among cirrhotic patients using conventional regression analysis and machine-learning algorithms</td>
<td>Prospective study</td>
<td>Patient factors</td>
</tr>
<tr>
<td>Banerjee et al.</td>
<td>2015</td>
<td>USA</td>
<td>L</td>
<td>ML/CV</td>
<td>RVI was assessed for its ability to predict MVI and outcomes in patients with HCC who underwent surgical resection or liver transplant</td>
<td>Prospective evaluation of a retrospective cohort</td>
<td>CT images</td>
</tr>
<tr>
<td>Walczak et al.</td>
<td>2017</td>
<td>USA</td>
<td>P</td>
<td>ML</td>
<td>Assess the accuracy of artificial neural networks in predicting survival in patients with pancreatic cancer using both clinical and patient-centered data</td>
<td>Retrospective study</td>
<td>Patient factors</td>
</tr>
<tr>
<td>Ying Zhou et al.</td>
<td>2017</td>
<td>China</td>
<td>L</td>
<td>ML/CV</td>
<td>Develop a CT-based radiomics signature and assess its ability to preoperatively predict the early recurrence (≤1 year) of hepatocellular carcinoma (HCC)</td>
<td>Retrospective study</td>
<td>CT images</td>
</tr>
<tr>
<td>Zheng et al.</td>
<td>2018</td>
<td>China</td>
<td>L</td>
<td>ML/CV</td>
<td>Developed a CT–based radiomic nomogram to predict recurrence-free survival rates for HCC after resection, ablation, and transplant</td>
<td>Retrospective study</td>
<td>CT images</td>
</tr>
<tr>
<td>Ivanics et al.</td>
<td>2019</td>
<td>Canada</td>
<td>L</td>
<td>ML</td>
<td>Leverage machine learning to develop an accurate post-transplantation HCC recurrence prediction calculator</td>
<td>Retrospective study</td>
<td>Patient factors</td>
</tr>
<tr>
<td>Sala Elarre et al.</td>
<td>2019</td>
<td>Spain</td>
<td>P</td>
<td>ML</td>
<td>Evaluated the 2-year relapse risk for pancreatic cancer patients based on a machine-learning algorithm</td>
<td>Retrospective study</td>
<td>Patient factors</td>
</tr>
<tr>
<td>Marinelli et al.</td>
<td>2019</td>
<td>USA</td>
<td>L</td>
<td>NLP/DL</td>
<td>Determine if weakly supervised learning/active transfer learning can hasten clinical deployment of deep learning models for liver segmentation</td>
<td>Retrospective study</td>
<td>Radiology reports/CT images</td>
</tr>
<tr>
<td>Naseif et al.</td>
<td>2019</td>
<td>USA</td>
<td>P</td>
<td>ML/CV</td>
<td>Develop a delta-radiomic process based on machine learning to predict the treatment response of pancreatic cancer</td>
<td>Retrospective study</td>
<td>CT images</td>
</tr>
<tr>
<td>Shan et al.</td>
<td>2019</td>
<td>China</td>
<td>L</td>
<td>ML/CV</td>
<td>A Prediction model based on peritumoral radiomics signatures from CT - investigate its efficiency in predicting early recurrence of HCC after curative treatment</td>
<td>Retrospective study</td>
<td>CT images</td>
</tr>
<tr>
<td>Chen et al.</td>
<td>2020</td>
<td>China</td>
<td>L</td>
<td>CV/ML</td>
<td>Establish a radiomics-based clinical model for preoperative prediction of PHLF in HCC</td>
<td>Retrospective study</td>
<td>MRI images</td>
</tr>
<tr>
<td>Han et al.</td>
<td>2020</td>
<td>South Korea</td>
<td>P</td>
<td>ML</td>
<td>Risk prediction model for POPF using AI</td>
<td>Retrospective study</td>
<td>Patient factors</td>
</tr>
<tr>
<td>Kambakamba et al.</td>
<td>2020</td>
<td>Switzerland</td>
<td>P</td>
<td>ML</td>
<td>The potential of machine learning-based approaches to describe the pancreatic texture and to predict POPF</td>
<td>Retrospective study</td>
<td>CT images</td>
</tr>
<tr>
<td>Merath et al.</td>
<td>2020</td>
<td>USA</td>
<td>L/P</td>
<td>ML</td>
<td>Assess ML algorithm to predict the patient risk of developing complications following liver, pancreatic or colorectal surgery</td>
<td>Retrospective study</td>
<td>Patient factors</td>
</tr>
<tr>
<td>Saillard et al.</td>
<td>2020</td>
<td>France</td>
<td>L</td>
<td>DL</td>
<td>Evaluate the effectiveness of AI algorithms to predict survival following HCC resection</td>
<td>Development and testing of AI models</td>
<td>Histology images</td>
</tr>
<tr>
<td>Cesaretti et al.</td>
<td>2020</td>
<td>France</td>
<td>L</td>
<td>ML/DL/CV</td>
<td>Automatizing liver-graft</td>
<td>Prospective</td>
<td>Surgery images</td>
</tr>
</tbody>
</table>
Italy segmentation from smartphone images and validating the robustness of this approach

Mai et al.\(^{(52)}\)
2020
China
L
DL
Establish and validate an artificial neural network model to predict severe post-hepatectomy liver failure in patients with hepatocellular carcinoma who underwent hemi-hepatectomy
Retrospective study
Patient factors

Liu et al.\(^{(53)}\)
2020
Taiwan
L
ML
Devise a predictive model to predict postoperative survival within 30 days based on the patient’s preoperative physiological measurement values
Retrospective study
Patient factors

Schoenberg et al.\(^{(54)}\)
2020
Germany
L
ML
Developing and validating a machine-learning algorithm to predict which patients are sufficiently treated by LR
Retrospective study
Patient factors

Szapkowski et al.\(^{(55)}\)
2020
USA
G
NLP
Determine the growth pattern of GPs and their association with GBC
Retrospective study
Radiology reports

Capretti et al.\(^{(56)}\)
2021
Italy
Portugal
P
CV/ML
Develop a reliable and reproducible machine learning-based multimodal risk model capable of predicting CR-POPF by combining radiomic features and morphologic features
Retrospective study
CT images/patient factors

Sun et al.\(^{(57)}\)
2021
China
L
DL
Develop a model to predict HCC recurrence
Retrospective study
Patient factors

Xie et al.\(^{(58)}\)
2021
USA
P
NLP
Develop and apply a natural language processing algorithm for the characterization of patients diagnosed with chronic pancreatitis
Retrospective study
Radiology reports

Hayashi et al.\(^{(59)}\)
2022
Japan
P
ML
Predict recurrence and metastatic sites in pancreatic cancer following curative surgery
Retrospective study
Histology images

Li et al.\(^{(60)}\)
2022
China
P
ML
Develop and validate clinical-radiomics models that preoperatively predict 1 and 2-year recurrence of PDAC
Retrospective study
CT images/patient factors

Noh et al.\(^{(61)}\)
2022
South Korea
L
ML
Machine learning-based survival rate prediction of hepatocellular carcinoma patients
Retrospective study
Patient factors

Morris-Stiff et al.\(^{(62)}\)
2022
USA
G
NLP
Develop a clinical prediction model for asymptomatic gallstones
Retrospective study
Radiology reports

Narayan et al.\(^{(63)}\)
2022
USA
L
ML/CV
Developed an objective, computer vision artificial intelligence (CVAI) platform to score donor liver steatosis and compared its capability for predicting EAD against pathologist steatosis scores
Retrospective study
Histology images

Cotter et al.\(^{(64)}\)
2022
USA
G
ML
Machine-based learning approach to stratify patients with gallbladder cancer into distinct prognostic groups using preoperative variables
Retrospective study
Patient factors

CV: Computer vision; CR-POPF: clinically relevant postoperative pancreatic fistula; EAD: early allograft dysfunction; G: gallbladder; GPs: gallbladder polyps; GBC: gallbladder cancer; L: liver; LR: liver resection; ML: machine learning; MVI: microvascular invasion; P: pancreas; PHLF: post-hepatectomy liver failure; POPF: postoperative pancreatic fistula; PDAC: pancreatic ductal adenocarcinoma; RVI: radiogenomic venous invasion.
<table>
<thead>
<tr>
<th>Author</th>
<th>Year of publication</th>
<th>Location</th>
<th>Organ</th>
<th>AI method</th>
<th>Aim</th>
<th>Design</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spinczyk et al.</td>
<td>2012</td>
<td>Poland</td>
<td>L</td>
<td>ML</td>
<td>Measurement of liver motion during surgery</td>
<td>Single center feasibility study</td>
<td>Surgery videos</td>
</tr>
<tr>
<td>Okamato et al.</td>
<td>2012</td>
<td>Japan</td>
<td>L/G</td>
<td>CV</td>
<td>Evaluate the utility of an image display system for augmented reality in hepatobiliary surgery under laparotomy</td>
<td>Case series</td>
<td>CT images</td>
</tr>
<tr>
<td>Fang et al.</td>
<td>2013</td>
<td>China</td>
<td>L</td>
<td>CV</td>
<td>Assess the use of 3d planning for hepatectomy for hepatolithiasis</td>
<td>Retrospective study</td>
<td>CT images</td>
</tr>
<tr>
<td>Zein et al.</td>
<td>2013</td>
<td>USA</td>
<td>L</td>
<td>CV</td>
<td>Establish anatomical precision and volumetric accuracy in 3D-printed models for donors and recipients undergoing LDLT</td>
<td>Prospective paired case series</td>
<td>CT/MRI images</td>
</tr>
<tr>
<td>Shahin et al.</td>
<td>2014</td>
<td>Germany</td>
<td>L</td>
<td>ML</td>
<td>Develop a navigation approach to quickly compensate for tumor movements due to surgical manipulation</td>
<td>Description of experiment</td>
<td>US images</td>
</tr>
<tr>
<td>Yang et al.</td>
<td>2014</td>
<td>South Korea</td>
<td>L</td>
<td>ML</td>
<td>Develop a user-centered 3D virtual liver surgery planning system algorithm</td>
<td>Pilot study</td>
<td>CT images</td>
</tr>
<tr>
<td>Fang et al.</td>
<td>2014</td>
<td>China</td>
<td>P</td>
<td>CV</td>
<td>Investigate the clinical significance of 3-dimensional reconstruction of peripancreatic vessels for patients with suspected pancreatic cancer</td>
<td>Randomized parallel single-blind study</td>
<td>CT images</td>
</tr>
<tr>
<td>Begin et al.</td>
<td>2014</td>
<td>Canada</td>
<td>L</td>
<td>CV</td>
<td>Evaluate an alternative automatic technique of liver volumetry based on a novel 3D virtual planning software and compare it to the manual technique</td>
<td>Prospective study</td>
<td>CT images</td>
</tr>
<tr>
<td>Bliznakova et al.</td>
<td>2015</td>
<td>Bulgaria</td>
<td>L</td>
<td>CV/ML</td>
<td>Develop and test a software application for evaluation of the residual function of the liver prior to the intervention of the surgeons</td>
<td>Case series</td>
<td>CT images</td>
</tr>
<tr>
<td>Katic et al.</td>
<td>2015</td>
<td>Germany</td>
<td>P/G</td>
<td>DL</td>
<td>Demonstrate the usefulness of deep learning model to identify surgical steps during laparoscopic cholecystectomy and pancreatic resections</td>
<td>Case series</td>
<td>Surgery videos</td>
</tr>
<tr>
<td>Song et al.</td>
<td>2015</td>
<td>UK</td>
<td>L</td>
<td>ML</td>
<td>Describe a freehand laparoscopic ultrasound-based system that registers liver vessels in ultrasound with MR/CT data</td>
<td>Description of experiment and case series</td>
<td>US/CT/MRI images</td>
</tr>
<tr>
<td>Wang et al.</td>
<td>2015</td>
<td>China</td>
<td>L</td>
<td>ML</td>
<td>Demonstrate the potential of homotopy-based SSC for shape-prior modeling in the liver surgical planning system</td>
<td>Description of experiment</td>
<td>CT images</td>
</tr>
<tr>
<td>Fang et al.</td>
<td>2015</td>
<td>China</td>
<td>L</td>
<td>CV</td>
<td>Compare outcomes of surgery on centrally located HCC with and without 3D planning</td>
<td>Retrospective study</td>
<td>CT/MRI images</td>
</tr>
<tr>
<td>Zhang et al.</td>
<td>2015</td>
<td>China</td>
<td>G</td>
<td>CV</td>
<td>Assess the use of 3d planning in surgery on bile duct cancer</td>
<td>Case series</td>
<td>CT images</td>
</tr>
<tr>
<td>Okuda et al.</td>
<td>2015</td>
<td>Japan</td>
<td>G</td>
<td>CV</td>
<td>Evaluate the impact of 3D CT cholangiography on operative planning and outcomes of biliary malignancies</td>
<td>Retrospective study</td>
<td>CT images</td>
</tr>
<tr>
<td>Okamoto et al.</td>
<td>2015</td>
<td>Japan</td>
<td>P</td>
<td>CV</td>
<td>Evaluate the utility of navigation surgery using augmented reality technology for pancreatectomy</td>
<td>Case series</td>
<td>CT images</td>
</tr>
<tr>
<td>Fortmeier et al.</td>
<td>2016</td>
<td>Germany</td>
<td>G/L</td>
<td>CV</td>
<td>Creation of a visuo-haptic simulation framework for the training and planning of the first steps of PTCD</td>
<td>Description of experiment</td>
<td>X-ray/US/CT images</td>
</tr>
<tr>
<td>Authors</td>
<td>Year</td>
<td>Country</td>
<td>IGS Method</td>
<td>Description of Experiment</td>
<td>Images</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>------</td>
<td>---------</td>
<td>------------</td>
<td>---------------------------</td>
<td>--------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusaglia et al.</td>
<td>2016</td>
<td>Switzerland</td>
<td>L CS</td>
<td>Present a novel LRS-based IGS system for laparoscopic liver procedures</td>
<td>Laparoscopic surgery images</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ntourakis et al.</td>
<td>2016</td>
<td>France</td>
<td>L CS</td>
<td>Investigate the potential of AR-based navigation to help locate and resect colorectal liver metastases</td>
<td>Prospectve pilot study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastmeyer et al.</td>
<td>2017</td>
<td>Germany</td>
<td>L ML</td>
<td>Compare axial force errors of simulated needle insertion for liver biopsy</td>
<td>Description of experiment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sauer et al.</td>
<td>2017</td>
<td>Germany</td>
<td>L CS</td>
<td>Evaluate the application of a mixed reality head-mounted display for the visualization of anatomical structures during liver surgery</td>
<td>Case study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cai et al.</td>
<td>2017</td>
<td>China</td>
<td>L CS</td>
<td>Report experience of using a 3d visualization system during hepatic resection</td>
<td>Case series</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miyamoto et al.</td>
<td>2017</td>
<td>Japan</td>
<td>P CV</td>
<td>3d planning - compared the pancreatic duct diameter and location with the intraoperative findings</td>
<td>Retrospective study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu et al.</td>
<td>2018</td>
<td>China</td>
<td>L CS</td>
<td>Assess the use of 3d planning for specific hepatectomy</td>
<td>Retrospective study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mise et al.</td>
<td>2018</td>
<td>Japan</td>
<td>L CV</td>
<td>Assess how virtual hepatectomy conducted using surgical planning software influences the outcomes of liver surgery</td>
<td>Retrospective study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mascagani et al.</td>
<td>2019</td>
<td>Italy/France</td>
<td>G DL</td>
<td>Develop and test a method for consistent critical view of safety evaluation and reporting in videos which could be developed into the deep learning model</td>
<td>Pilot study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teatini et al.</td>
<td>2019</td>
<td>Norway</td>
<td>L ML</td>
<td>Test if intraoperative imaging is necessary for accurate surgical navigation for laparoscopic liver resection</td>
<td>Description of experiment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ho et al.</td>
<td>2020</td>
<td>New Zealand</td>
<td>L CV</td>
<td>Describe the computational pipeline that integrates into silico liver models and algorithms to aid surgical planning for liver resection</td>
<td>Description of experiment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prevost et al.</td>
<td>2020</td>
<td>Switzerland</td>
<td>L CV</td>
<td>Evaluate the technical feasibility and the clinical impact of a new augmented reality system for laparoscopic liver surgery</td>
<td>Pilot study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sandal et al.</td>
<td>2021</td>
<td>Turkey</td>
<td>G ML</td>
<td>Determine the usefulness of fuzzy logic algorithm to evaluate risk in patients undergoing laparoscopic cholecystectomy</td>
<td>Case series</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cervantes-Sanchez et al.</td>
<td>2021</td>
<td>Mexico/Germany</td>
<td>L/G ML/DL</td>
<td>Machine/deep learning methods are combined with HSI-goal is the automatic discrimination using HSI of the bile duct from the gallbladder and liver</td>
<td>Description of experiment and case series</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tokuyasu et al.</td>
<td>2021</td>
<td>Japan</td>
<td>G DL/CV</td>
<td>Develop a system that outlines laparoscopic cholecystectomy landmarks on endoscopic images in real time</td>
<td>Description of experiment and case report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guzman-Garcia et al.</td>
<td>2021</td>
<td>Spain</td>
<td>G NLP/DL</td>
<td>Assess if analysis of surgeons’ speech using natural language processing provide deeper insight into the surgical decision-making processes during laparoscopic cholecystectomy</td>
<td>Description of experiment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imler et al.</td>
<td>2021</td>
<td>USA</td>
<td>G NLP/ML</td>
<td>Demonstrate the feasibility of using NLP to measure adherence</td>
<td>Retrospective study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authors</td>
<td>Year</td>
<td>Country</td>
<td>Methodology</td>
<td>Case Study/Validation</td>
<td>Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>-------</td>
<td>------------------</td>
<td>---------------</td>
<td>-----------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ruzzenente et al.</td>
<td>2022</td>
<td>Italy</td>
<td>L ML</td>
<td>Case series</td>
<td>Evaluate four difficulty scoring systems in liver surgery and determine the most important characteristics using random forest models.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mascagani et al.</td>
<td>2022</td>
<td>France/Italy</td>
<td>G DL/CV</td>
<td>Multicentre retrospective validation</td>
<td>Creation of an assessment tool for CVS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mascagani et al.</td>
<td>2022</td>
<td>France/Italy</td>
<td>G DL/CV</td>
<td>Case series</td>
<td>Develop a deep learning model to automatically segment hepatocystic anatomy and assess the criteria defining the critical view of safety (CVS).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tranter-Entwistle et al.</td>
<td>2022</td>
<td>New Zealand/Australia</td>
<td>G ML/CV</td>
<td>Case series</td>
<td>Use a commercially available ML-driven platform to evaluate a subjective grading of operative difficulty in laparoscopic cholecystectomy.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liu et al.</td>
<td>2022</td>
<td>China</td>
<td>G ML/CV</td>
<td>Pilot study</td>
<td>Develop model and preliminarily verify its potential surgical guidance ability by comparing its performance with surgeons during laparoscopic cholecystectomy.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ugail et al.</td>
<td>2022</td>
<td>UK</td>
<td>L ML/DL/CV</td>
<td>Pilot study</td>
<td>Present the use of deep learning for the non-invasive evaluation of donor liver organs.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mojtahed et al.</td>
<td>2022</td>
<td>USA/Netherlands/Portugal</td>
<td>L DL/CV</td>
<td>Retrospective study</td>
<td>Demonstrate the accuracy and precision of liver segment volume measurements.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Han et al.</td>
<td>2022</td>
<td>China</td>
<td>L DL/CV</td>
<td>Retrospective study</td>
<td>Develop and validate a three-dimensional convolutional neural network model for automatic liver segment segmentation.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ward et al.</td>
<td>2022</td>
<td>USA</td>
<td>G DL/CV</td>
<td>Development and testing of AI models</td>
<td>Trained model to identify PGS.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Madani et al.</td>
<td>2022</td>
<td>Canada/USA/UK</td>
<td>G DL/CV</td>
<td>Development and testing of AI models</td>
<td>Develop and evaluate the performance of models that can identify safe and dangerous zones of dissection during laparoscopic cholecystectomy.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loukas et al.</td>
<td>2022</td>
<td>Greece</td>
<td>G DL/CV</td>
<td>Development and testing of AI models</td>
<td>Framework for vascularity classification of the gallbladder wall from intraoperative images of laparoscopic cholecystectomy.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Golany et al.</td>
<td>2022</td>
<td>Israel</td>
<td>G DL/CV</td>
<td>Development and testing of AI models</td>
<td>Developed algorithm and evaluated its performance in recognizing surgical phases of laparoscopic cholecystectomy.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AR: Augmented reality; CVS: Critical view of safety; G: gallbladder; HSI: hyperspectral images; IGS: image guided surgery; L: liver; LDLT: living donor liver transplant; LC: laparoscopic cholecystectomy; LRS: laser range scanners; PTCD: percutaneous transhepatic biliary drain; PGS: parkland grading scale for cholecystitis; P: pancreas; SSC: sparse shape composition.

derived from imaging to group lesions into disease subgroups\cite{15,18,21}. In another example, decision tree models were used to predict the occurrence of any complication and of specific complications in patients undergoing liver, pancreatic and colorectal surgery\cite{49}. These algorithms were superior to the American Society of Anaesthesiologists (ASA) classification at predicting the chance of any complication. They performed well for specific complications, with c-statistics ranging from 0.76 to 0.98. As described in our conceptual mapping exercise, the augmentation of surgical fields to highlight relevant anatomy is a key area.
of research. This is an example of object detection and is a task well suited to laparoscopic cholecystectomy. Madani et al. describe a deep learning algorithm that intraoperatively recognizes “go,” or “no-go” areas of dissection to minimize the risk of adverse events such as bile duct injury.

Figure 4. Conceptual mapping of areas of AI research in HPB surgery, stratified by treatment timing. This exercise identified several areas of overlap (dashed arrows) across different divisions, in addition to several areas where AI would be useful for future research (purple free text). CT: Computed tomography; IPMN: intraductal papillary mucinous neoplasm; MRI: magnetic resonance imaging; PHLF: post hepatectomy liver failure; POPF: postoperative pancreatic fistula.

Figure 5. Number of studies utilizing large datasets stratified by AI approach.
The intersection of AI and big data in HPB surgery

We identified eleven studies utilizing large datasets in HPB surgery applications [Table 4]. Nine have been published since 2020. Eight of the 11 identified papers utilized NLP to extract data from large numbers of reports, mainly with the aim of identifying patients with a specific condition, either for phenotyping or to identify patient cohorts. The majority originated from the USA (n = 9; 82%), with one study from China and one from South Korea [Figure 5].

An example of the use of an NLP algorithm to identify patient cohorts and devise a means of following-up incidental scan results was by Kooragayala et al.[35]. This study used a keyword search associated with suspicious pancreatic lesions in over 18,000 patients who underwent a CT scan following trauma over a 10-year period. The approach identified pancreatic lesions in the reports of 232 patients, of which 48 were intraductal papillary mucinous neoplasms (IPMNs). In addition, this paper proposed a management flowchart for incidentally found pancreatic lesions. A further example of the use of NLP in high-volume data was demonstrated by Morris-Stiff et al.[62], who used NLP to identify asymptomatic gallstones from a cohort of 49,414 patients. They were then able to identify risk factors for progression to symptomatic gallstone disease in this asymptomatic cohort and showed an approximately 2% risk of symptomatic progression per year.

DISCUSSION

This review has identified a rapid increase in the quantity of AI research conducted within HPB surgery. Much of this is focused on intraoperative applications of AI, such as the use of image analysis and computer vision to address diagnostic and prognostic uncertainties. In addition, the use of 3D reconstruction and augmented reality models coupled with data-driven prediction algorithms has emerged as an important area, particularly in preoperative planning and intraoperative decision-making in liver surgery. Artificial intelligence methods have the most to offer in the distillation of multi-dimensional information to tractable knowledge that can be applied to individual treatment decisions. HPB surgery represents a good target for these technologies, given the frequently complex disease patterns and diverse treatment pathways employed.

Most artificial intelligence approaches rely on large volumes of data for training purposes. Of the commonly described features of big data, the included studies reflect “volume” and “variety” with fewer utilizing real-time rapidly changing data (“velocity”). Data sources included large pre-existing databases, collated images and imaging reports. Two notable databases used were the Cancer Genome Atlas and the American College of Surgeons National Surgical Quality Improvement Program (NSQIP) database, which were widely used across a range of studies. Natural language processing was frequently employed to extract information from imaging reports and other healthcare text sources. In one study, NLP was used to identify concerning pancreatic lesions in historical imaging reports[35]. This demonstrates the depth and flexibility in AI techniques to adapt to changes in patient management over time - the malignant potential of particular pancreatic cysts has only been appreciated in recent years. Moreover, these approaches may be adapted to help non-specialists managing HPB conditions, particularly in low-resource settings with limited access to tertiary HPB services. As computer vision approaches improve, the supplementation of local imaging and pathology reporting with AI-derived diagnostic support may leapfrog the requirement for massive and often unaffordable training of humans to perform these tasks.

There are, however, genuine risks of bias arising with the development of these techniques. We found significant geographical variation in current research, with no studies incorporating data from low- and middle-income countries (LMICs). If the benefits of AI are to be shared equitably across contexts, then investigators must consider how solutions can broadly generalize between populations and avoid
<table>
<thead>
<tr>
<th>Author</th>
<th>Year of publication</th>
<th>Study description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roch et al.</td>
<td>2015</td>
<td>566,233 CT reports from 50,669 patients analysed for keywords associated with Pancreatic Cysts using NLP</td>
</tr>
<tr>
<td>Yang et al.</td>
<td>2017</td>
<td>The Cancer Genome Atlas catalogs genes associated with 33 cancers. Genes associated with HCC were extracted from database and checked for overlap with genes identified in 35 years of published literature using NLP</td>
</tr>
<tr>
<td>Merath et al.</td>
<td>2020</td>
<td>15,657 patients undergoing liver, pancreatic or colorectal surgery (685 liver and 6,012 pancreatic) retrospectively were identified from the American College of Surgeons National Surgical Quality Improvement Program database. Risk-prediction Machine Learning model created from pre-op characteristics</td>
</tr>
<tr>
<td>Szpakowski et al.</td>
<td>2020</td>
<td>365 Gallbladder Cancer and 35,970 Gallbladder Polyp patients were identified from 622,227 patients in a Californian health system. NLP was used to identify Polyps from Ultrasound reports</td>
</tr>
<tr>
<td>Xie et al.</td>
<td>2021</td>
<td>58,085 imaging reports from 6,346 Chronic Pancreatitis patients were used to develop an NLP algorithm that could characterize features of Chronic Pancreatitis</td>
</tr>
<tr>
<td>Yamashita et al.</td>
<td>2021</td>
<td>430,426 imaging reports from 199,783 patients were used to create an NLP algorithm to identify the presence and size of Pancreatic Cysts</td>
</tr>
<tr>
<td>Imler et al.</td>
<td>2021</td>
<td>23,674 ERCP reports were analyzed for quality measures using NLP</td>
</tr>
<tr>
<td>Noh et al.</td>
<td>2022</td>
<td>Machine learning-based prediction models for survival applied to 10,742 HCC patients</td>
</tr>
<tr>
<td>Morris-Stiff et al.</td>
<td>2022</td>
<td>Ultrasound reports identified 49,414 patients with gallstones. NLP algorithm trained to identify asymptomatic patients (22,257)</td>
</tr>
<tr>
<td>Narayan et al.</td>
<td>2022</td>
<td>25,494 images from 90 liver biopsies were used to develop Machine Learning Computer Vision models to score liver steatosis</td>
</tr>
<tr>
<td>Kooragayala et al.</td>
<td>2022</td>
<td>NLP was used to identify pancreatic lesions from 18,769 adult trauma CT reports</td>
</tr>
</tbody>
</table>

CT: Computed tomography; ERCP: endoscopic retrograde cholangiopancreatography; HCC: hepatocellular carcinoma; NLP: natural language processing.

Exacerbating pre-existing healthcare disparities. This is a widely discussed and controversial topic in the broader AI field. Inherent systematic biases in datasets clearly exist, with some of the most obvious reflecting racial, socioeconomic and gender-based prejudices. Addressing these complex issues is crucial across all AI work, including in HPB surgery. The majority of HPB disease occurs in LMICs, so it is essential that these populations are better represented in current HPB research more broadly and AI research specifically.

In addition to geographical disparities, concerns around the transparency of AI algorithms and lack of explainability are likely to hamper uptake and trust in clinical practice. The need for explainability is rooted in evidence-based medicine, which relies on transparency and reproducibility in decision-making. Without explainable AI, patient trust in healthcare will erode. Others have argued that true explainability represents a false hope, and that explainability methods cannot deliver meaningful patient-level interpretability. The focus should be on robust internal and external validation. In this review, we found little reference to concepts of explainability in included studies. It is important that these issues are explored and addressed, particularly when developing algorithms orientated toward patient-facing prognostication. As AI systems transition from research to clinical practice, transparency and reliability are paramount if trust is to be built and maintained.

The ability to understand and reproduce scientific findings is imperative, yet reporting the quality of included studies was variable. A number of useful reporting guidelines now exist, specifically orientated toward AI. In 2019, a rigorous process of literature review, expert consultation, Delphi survey, and consensus meeting resulted in the SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials - Artificial Intelligence) and CONSORT-AI (Consolidated Standards of Reporting Trials - Artificial Intelligence) standards. In addition, two additional tools are currently under
development: Transparent Reporting of a multivariable prediction model of Individual Prognosis Or Diagnosis AI extension (TRIPOD-AI) and the Prediction model Risk Of Bias Assessment Tool (PROBAST-AI). These promise to provide standardization and assessment tools that will greatly increase the quality of clinically-orientated AI study reporting.

Where should AI research in HPB be focussed? Most studies in this review concentrated on image analysis. While this is an important area, there are many other challenges in HPB which could benefit from the application of AI. Research prioritization in AI must be determined by broad stakeholder groups, led primarily by the patient and public representatives, accounting for a range of viewpoints and actively engaging non-technical individuals in the design and delivery of research studies. We found little mention of engagement with stakeholder groups in included studies (e.g., patients, clinicians and the wider HPB community), which is crucial if these complex interventions are to move into clinical practice successfully. Moreover, included studies focused on the development of AI models rather than on the implementation of AI systems. While this is understandable given the current stage of development, future work should focus on how broader AI-driven systems can be implemented safely into clinical pathways and be clear about the function they serve.

Our study has several limitations. First, there is significant heterogeneity in the content and outcomes of the various studies included. While meaningful comparisons are challenging, a useful overview of common issues and themes affecting AI research in HPB is provided. Second, as is the nature of a scoping review, it is possible that studies meeting the inclusion criteria have been omitted, leading to an incomplete presentation of the current literature. For example, papers focusing on NLP and the gallbladder were relatively poorly represented in exploratory literature searches, possibly reflecting poor search descriptors and study labeling. Finally, as AI and associated concepts are undergoing rapid development, study inclusion criteria are in flux. Improving formal definitions in these emerging fields will help study classification and ease of literature identification.

The use of AI and big data in HPB surgery and medicine, more generally, is rapidly expanding. AI promises benefits in the delivery of clinical care and may result in future improvement of healthcare outcomes. This review identifies crucial interlinking conceptual areas of AI as applied to HPB surgery. Future research must address issues of bias, transparency, and explainability and ensure that innovation is representative of HPB patient populations across the world.

DECLARATIONS
Authors’ contributions
Participated in the design of the study, data collection, screening, interpretation and presentation, writing of the manuscript and submitted the manuscript: McGivern KG
Participated in the design of the study, data collection, screening, interpretation and presentation, and critical evaluation of the manuscript: Knight SR
Participated in the writing and critical evaluation of the manuscript: Drake TM
Participated in data screening and presentation: Lucocq J
Participated in the critical evaluation of the manuscript: Bernabeu MO, Clark N, Fairfield C, Pius R, Shaw C, Seth S
Participated in the design of the study and critical evaluation of the manuscript: Harrison EM
All authors approved the final version of the manuscript
Availability of data and materials
Not applicable.

Financial support and sponsorship
None.

Conflicts of interest
All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Copyright
© The Author(s) 2023.

REFERENCES
34. Chang Y, Wu Q, Chi L, Hoo H, Li Q. Adoption of combined detection technology of tumor markers via deep learning algorithm in diagnosis and prognosis of gallbladder lesions. *J Supercomput* 2022;78:3955-75.  DOI
preoperative non-contrast enhanced computed tomography. *HPB* 2020;22:5384. DOI


77. Fang CH, Tao HS, Yang J, et al. Impact of three-dimensional reconstruction technique in the operation planning of centrally located


