Applications of Artificial Intelligence in liver cancer: a scoping review – part I, screening and diagnosis.

Andrea Chierici, Fabien Lareyre, Antonio Iannelli, Elise Poggi, Sebastien Goffart, Lisa Guzzi, Hervé Delingette, Juliette Raffort

Submitted to: Journal of Medical Internet Research
on: February 25, 2024

Table of Contents

Applications of Artificial Intelligence in liver cancer: a scoping review – part I, screening and diagnosis.

Andrea Chierici[1, 2]; Fabien Lareyre[3, 4, 5]; Antonio Iannelli[1, 6]; Elise Poggi[3]; Sebastien Goffart[7]; Lisa Guzzi[7, 8]; Hervé Delingette[7, 8]; Juliette Raffort[4, 5, 8, 9]

[1]Digestive Surgery and Liver Transplant Unit - University Hospital of Nice Nice FR
[2]Digestive Surgery Unit - Antibes Juan-les-Pins Hospital Antibes FR
[3]Department of Vascular Surgery, Hospital of Antibes Juan-les-Pins, Antibes, France Antibes FR
[4]Université Côte d'Azur, Inserm U1065, C3M, Nice, France Nice FR
[5]Fédération Hospitalo-Universitaire FHU Plan&Go, Nice, France Nice FR
[6]Université Côte d'Azur, Inserm U1065, Team 8 “Hepatic complications of obesity and alcohol”, Nice, France Nice FR
[7]Université Côte d'Azur, Inria, Epione Team, Sophia Antipolis, France Sophia Antipolis FR
[8]3IA Institute, Université Côte d'Azur, Nice, France Nice FR
[9]Clinical Chemistry Laboratory, University Hospital of Nice, Nice, France Nice FR

Corresponding Author:
Andrea Chierici
Digestive Surgery and Liver Transplant Unit - University Hospital of Nice
151, route Saint Antoine de Ginèstiere
Nice
FR

Abstract

Artificial intelligence (AI) has emerged as a powerful tool in various fields, including medicine, offering the potential to revolutionize the way diseases are diagnosed and treated. This scoping review explores the applications of AI in primary liver cancer, focusing on screening and diagnosis. Liver cancer, particularly hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC), presents significant challenges due to late-stage diagnoses and limited treatment options. AI-driven models have been developed to enhance screening efforts, utilizing machine learning (ML) algorithms trained on clinical, biochemical, and radiological data to identify high-risk patients. These models demonstrate promising results in early HCC detection, especially in populations with chronic hepatitis B virus (HBV) infection or metabolic dysfunction-associated fatty liver disease (MAFLD). Additionally, AI applications in liver imaging, utilizing deep learning (DL) algorithms such as convolutional neural networks (CNN), have shown remarkable accuracy in segmenting and classifying liver lesions on CT and MR images. However, challenges remain in model validation, standardization, and reproducibility, with many studies lacking external validation and consistency in reporting performance metrics. Furthermore, the transition from model development to real-world implementation poses significant hurdles, highlighting the need for a more rigorous and transparent approach in AI model development. Despite these challenges, AI-driven models hold immense potential to improve early detection and diagnosis of primary liver cancer, ultimately leading to better patient outcomes. Further research and collaboration are warranted to address the current limitations and facilitate the integration of AI into clinical practice for liver cancer management.

(JMIR Preprints 25/02/2024:57617)
DOI: https://doi.org/10.2196/preprints.57617

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Applications of Artificial Intelligence in liver cancer: a scoping review – part I, screening and diagnosis.

Andrea Chierici, MD[1,2], Fabien Lareyre, MD, PhD[3,4,5], Antonio Iannelli, PhD, MD[1,6,7], Elise Poggi, MD[3], Sebastien Goffart[8], Lisa Guzzi[8,9], Hervé Delingette, PhD[8,9], Juliette Raffort MD, PhD[4,5,9,10].

[1] Department of Digestive Surgery, University Hospital of Nice, Nice, France

[2] Department of Digestive Surgery, Hospital of Antibes Juan-les-Pins, Antibes, France

[3] Department of Vascular Surgery, Hospital of Antibes Juan-les-Pins, Antibes, France

[4] Université Côte d'Azur, Inserm U1065, C3M, Nice, France

[5] Fédération Hospitalo-Universitaire FHU Plan&Go, Nice, France

[6] Université Côte d'Azur, Inserm U1065, Team 8 “Hepatic complications of obesity and alcohol”, Nice, France

[7] ADIPOCIBLE Study Group

[8] Université Côte d'Azur, Inria, Epione Team, Sophia Antipolis, France

[9] 3IA Institute, Université Côte d'Azur, Nice, France

[10] Clinical Chemistry Laboratory, University Hospital of Nice, Nice, France

Corresponding author:

Andrea Chierici, MD

Department of Digestive Surgery, University Hospital of Nice, Nice – France.

Hopital l’Archet 2 – 151, route Saint’Antoine de Ginèstiere, 06200 Nice – France.

chierici.ap@chu-nice.fr

Number of words: 6354.

Number of Tables: 2.

Number of references: 55.

Abstract

Artificial intelligence (AI) has emerged as a powerful tool in various fields, including medicine, offering the potential to revolutionize the way diseases are diagnosed and treated. This scoping review explores the applications of AI in primary liver cancer, focusing on screening and diagnosis. Liver cancer, particularly hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC), presents significant challenges due to late-stage diagnoses and limited treatment options. AI-driven models have been developed to enhance screening efforts, utilizing machine learning (ML) algorithms trained on clinical, biochemical, and radiological data to identify high-risk patients. These models demonstrate promising results in early HCC detection, especially in populations with chronic hepatitis B virus (HBV) infection or metabolic dysfunction-associated fatty liver disease (MAFLD). Additionally, AI applications in liver imaging, utilizing deep learning (DL) algorithms such as convolutional neural networks (CNN), have shown remarkable accuracy in segmenting and classifying liver lesions on CT and MR images. However, challenges remain in model validation, standardization, and reproducibility, with many studies lacking external validation and consistency in reporting performance metrics. Furthermore, the transition from model development to real-world implementation poses significant hurdles, highlighting the need for a more rigorous and transparent approach in AI model development. Despite these challenges, AI-driven models hold immense potential to improve early detection and diagnosis of primary liver cancer, ultimately leading to better patient outcomes. Further research and collaboration are warranted to address the current limitations and facilitate the integration of AI into clinical practice for liver cancer management.

Introduction

Artificial Intelligence (AI) is a discipline that aims at developing computer systems able to perform tasks normally requiring human intelligence through iterative learning. Its application in a multiple of settings is increasing constantly and exponentially and medicine is deeply affected by this phenomenon. AI is a large term comprising many different algorithms that can be deeply different in terms of structure, function, and application. The most common form of AI exploited in medicine is machine learning (ML), a subset of AI that allows making predictions once the model is fed and trained with data. Artificial neural networks (ANN) are a subset of ML, whose structure mimics how neurons interact in the brain signal. ANN are composed by layers of nodes that receive an input and elaborate an output which is sent to the next layer if the activation threshold value is reached. To furtherly improve their abilities, ANN have been implemented with multiple layers with different specific skills that make a specific neural network more suitable of doing a particular task. When one ANN is composed of at least 3 layers it is defined as a deep learning (DL) neural network.

Cancer is one of the leading cause of mortality worldwide, with 19.3 millions new cancer diagnosis and 10 millions cancer-related deaths in 2020 (1). Among them, liver cancer is a primary worldwide health issue. It is the seventh most common cancer diagnosed but it is the second cause of cancer-related death after lung cancer (1). Primary liver cancer is mainly represented by two different histotypes: hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC), developing respectively from hepatocytes and cholangiocytes and responsible for about 70% and 15% of all primary liver cancers, respectively (2). Although many different interventions, procedures, and systemic therapies exists and new ones are constantly developed to improve liver cancer treatment results,

both HCC and ICC are still burdened by elevated mortality (3). This can be partly attributed to several factors. Liver cancer is often underdiagnosed or diagnosed at a late stage due to its indolent development and heterogeneity of clinical symptoms (4). Furthermore, once liver cancer is diagnosed, it is hard to define a precise prognosis prediction as it is affected by many variables, from cancer's and patient's characteristics to the choice of available and performable treatments. In case of HCC, expected survival is affected by many variables including the number of nodules, the size of the tumors, and liver function, and it is reported to vary from > 5 years in case of single ≤ 2cm lesions to < 3 months in case of any tumor burden associated with end stage liver function (5). For ICC, survival can also vary from > 5 years in case of limited, intra-hepatic, resectable cancer to a median of 16 months in case of metastatic disease to the peritoneum; however, only 20-30% of patients have resectable ICC at the time of diagnosis (6). In this context, there is an increasing interest to develop AI-applications to improve screening and diagnostic tools to identify liver cancer at an early stage, to guide the decision making and propose personalized care.

This scoping review aims is to summarize the most recent findings on AI applied to screening, diagnosis, treatment planning, treatment efficacy assessment, prognosis prediction, and follow-up for primary liver cancer. In this first part, we focus on how AI can contribute to improve screening and diagnosis performances in primary liver cancer.

Screening

Primary liver cancer and especially HCC develops in well-known predisposing conditions. Independently from the cause, liver cirrhosis with fibrosis and chronic inflammation, is the principal favoring substrate for HCC. However, it has also been shown that chronic inflammation secondary to HBV infection and liver steatosis in the setting of metabolic disfunction with insulin resistance associated fatty liver disease can favor HCC before the

stage of liver cirrhosis is reached (7). Concerning ICC, chronic hepatitis, cirrhosis, biliary inflammatory disease, and hepatobiliary flukes are the most important risk factors although, frequently, none of these conditions is identified when ICC is diagnosed (8).

Screening programs have been developed worldwide for different types of cancers in order to identify suspect lesions at an early stage in patients with predisposing factors. This is especially the case of patients with chronic liver disease for whom biochemical and radiological tests are routinely and regularly performed to early detect liver cancer.

Recently, AI models have been used to exploit clinical, biochemical, and radiologic data to identify liver cancer in asymptomatic patients. These models are essentially based on different ML algorithms that are trained on pre-existent datasets of patients with or without liver cancer whose characteristics are exploited to make predictive algorithms. Table 1 summarizes the main studies that have reported the use of AI-driven models to perform primary liver cancer screening.

With the actual obesity epidemics in the Western countries and the associated increase in the incidence of metabolic dysfunction-associated fatty liver disease (MAFLD) which are well-known favoring substrates for liver cancer (9,10), Okanoue et al. (11) developed a DL based tool to precociously identify HCC in patients with MAFLD at the stage of steatohepatitis during follow-up. This model was based on patients' age, sex, height, weight, AST, ALT, GGT, triglyceride, cholesterol levels, platelet count, diabetes status, and IgM-free apoptosis inhibitor of macrophage level. The sensitivity and specificity for the presented algorithm were 95.2% and 100% respectively.

Another setting where AI algorithms can be useful for early HCC detection is patients with chronic HBV infection. HBV and HCV infections still represent the most frequent cause of HCC worldwide and in Eastern countries in particular, even though the vaccination is changing HBV epidemiology in young adults (12). Many clinical-biochemical models have

been developed to predict HCC in both patients with treated and untreated HBV infection (13), but only a few are based on AI techniques. Lee et al. (14) used a retrospective dataset of treatment-naïve HBV patients who started Entecavir or Tenofovir and were followed for 5 years for HCC surveillance to build such a model. Different algorithms (Random Forest, XGBoost, and logistic regression) were exploited and the model was validated on an external cohort of patients. The included variables were age, sex, cirrhosis, HBV DNA levels, ALT levels, total and serum bilirubin levels, AFP levels, HBeAg positivity, platelet count, and INR, and the area under the receiver operator curve was 0.9, indicating an excellent predictive value. However, the main limitation of this model is that it was developed and validated on Korean patients followed in academic centers, thus possibly affecting the generalization of the model's performance on other populations. To overcome this limitation, Kim et al. (15) developed a HCC predictive model using the Gradient-Boosting Machine algorithm on Korean treatment-naïve HBV patients who started Entecavir or Tenofovir that was then externally validated firstly on Korean and then on Caucasian patients from 11 Western institutions. Including 10 parameters (presence of cirrhosis, age, platelet count, antiviral agent used, sex, ALT levels, baseline serum levels of HBV DNA, serum levels of albumin and bilirubin, and HBeAg status) in the model, the Authors obtained a model with a C-index (i.e., a goodness of fit measure to evaluate risk models in survival analysis) of 0.79 for the Korean population and 0.81 for the Caucasian population, which was better than the predictive performance of some non-AI developed models validated and used worldwide (16–18).

Concerning ICC prediction, Hu et al. realized a model for patients with primary sclerosing cholangitis, which is a well-known risk factor for bile ducts cancer (19). Cox's regression, Random Forest, and Gradient-Boosting algorithms were used to produce three separate predictive models. The variables exploited in the models were the presence and

the duration of inflammatory bowel disease (commonly but not perpetually associated with primary sclerosing cholangitis), the duration of primary sclerosing cholangitis, total bilirubin and CA19.9 levels, and age. The proposed models showed a C-index of 0.69, 0.68, and 0.71 in predicting ICC at baseline, 2 years, and 5 years respectively. Nonetheless, the absence of an external validation of the proposed models represents an important limitation to the actual applicability of the models in clinical practice.

Author	Country	Data	Patients (n)	Model	Setting	Outcome			
						Se (%)	Sp (%)	AUC	c-index
Okanoue, 2023[9]	Japan	clinical, biochemical	230	DL	HCC, NASH	95.2	100		
Lee, 2023[12]	Korea	clinical, biochemical	2897	ML	HCC, HBV			0.872	
Kim, 2022[13]	International	clinical, biochemical	6051	ML	HCC, HBV				0.79/0.81
Hu, 2023[17]	US	clinical, biochemical	1459	ML	ICC, PSC				0.67

Table 1: characteristics of the studied describing AI-driven models to perform primary liver cancer screening. DL: deep learning; ML: machine learning; HCC: hepatocellular carcinoma; ICC: intrahepatic cholangiocarcinoma; NASH: non-alcoholic steatohepatitis; HBV: hepatitis B virus; PSC: primary sclerosing cholangitis; Se: sensitivity; Sp: specificity; AUC: area under the curve.

Liver cancer diagnosis

The diagnosis of liver primary cancer is based on the combination of different clinical, biochemical, radiological, and pathological features that combined allow to correctly detect and classify one or more liver mass(es). Many different instruments based on AI have been recently developed in order to exploit a different combination of this information to identify liver cancer with elevated accuracy. Table 2 reports the characteristics of the models used to perform primary liver cancer diagnosis based either on medical images or p*athological and omics data.*

Imaging models

Liver imaging has a key role in the detection of hepatic masses and their characterization in order to establish the most adapted workup and treatment. Depending on the nature of the liver mass, specific findings on a multiphasic CT-scan or MR can be enough to diagnose liver cancer. In case of HCC, the Liver Imaging Reporting and Data System (LI-RADS) Management Working Group has defined an algorithm based on fixed characteristics permitting to discriminate benign and malignant lesions in patients at risk for HCC (20) without need for invasive procedures to confirm diagnosis as reported in international liver cancer management guidelines (21,22). However, it has been shown that the sensitivity and specificity of both multiphasic CT-scan and MR is not optimal, and can be affected by many patients-, lesion-, technical-, and operator-related factors. For example, a meta-analysis by Chen et al. (23) showed that MR and CT-scan sensitivity is respectively 77% and 63% and that their specificity is 93% and 94% in patients at risk for HCC, but that sensitivity decreases to 69% and 60% when considering ≤ 20mm diameter liver lesions only.

Furthermore, when it comes to consider ICC, which is the second most frequent intrahepatic liver cancer after HCC (24), a pathological evidence is still considered mandatory although CT-scan and MR have elevated accuracy in identifying this specific

histotype of liver cancer (25).

There are examples of several applications of AI in liver imaging, exploiting the three principal imaging techniques routinely used for the diagnosis of liver cancer: ultrasonography (US), CT-scan, and MR. We can resume it at five main tasks, alone or combined, that contribute to improve diagnostic accuracy: 1) enhancement of liver abnormalities through imaging noise reduction; 2) Segmentation (organ and bounderies location and definition) of liver lesions, anatomy, and vascular structures; 3) Classification of liver lesions; 4) Measurement of liver and tumor volume; 5) 3-dimensional (3D) rendering of the liver and its anatomical and pathological structures. Nowadays, these assignments are mainly fulfilled through the application of DL, a subset of ML based on ANN consisting of 3 or more layers, groups of nodes that represent the basic unit of a network receiving inputs that are elaborated in order to produce an output that will be used by the following layer. The most common subtypes of neural networks adopted in this setting are convolutional neural network (CNN), fully convolutional network (FCN), and recurrent neural network (RNN). CNN has been specifically developed to analyze image and audio inputs with a dramatic reduction of computational work and a reduced need for image enhancement through convolution compared to other non-convolutional algorithms. When a network is exclusively based on convolution it is called FCN. Finally, a RNN is a subtype of neural network, in contrast with feedforward neural network, whose layers are connected bi-directionally, so that one-layer's output can be used as an input for a previous layer in the network, whose output will be elaborated by the following layer in a circular system.

One recent example of the power of DL-mediated diagnosis is the one by Midya et al. (26) based on the utilization of CNN on a dataset of 814 preoperative portal venous phase CT-scans of patients undergoing surgery for liver tumor. The accuracy of the model was 95.83% for ICC, 93.94% for HCC, and 85.71% for benign lesions compared to the result of

histological analysis on surgical specimens. Moreover, when compared to experienced radiologists diagnosis, the CNN algorithm seems to perform better (overall accuracy: 96.27% vs. 92.55%) without being significantly demonstrated. However, the described algorithm functions in a semiautomated manner, needing for the intervention of a radiologist to perform the initial manual pre-processing. Instead, Wang et al. (27) described a fully-automated CNN-based algorithm focused on CT-scan HCC detection using a large institutional and different public datasets made of normal and pathologic (HCC, ICC, metastases, benign lesions) images. Although this method permits to avoid human intervention, HCC diagnosis overall accuracy resulted to be 81.3%, with a predictive power that decreased in case of single lesions, lesion's diameter < 5 cm, low-grade liver fibrosis. This did not differ from the accuracy (80-84%) of segmentation provided by the radiologists and used as gold standard for comparison. Similar results are reported by Gao et al. (28) whose proposed model is based on a CNN and RNN architecture. Tumor segmentation was performed with 86.2% accuracy and distinction between ICC and HCC was correctly outlined in 82.9% of cases, better than what was reported for radiologists (70.8%). Far better results in terms of diagnosis concordance between CT-scan images and postoperative pathology were obtained by Zhang et al. (29) who reported a 96.55% accuracy for their reverse-CNN algorithm. However, the author described the accuracy of the model to distinguish between different benign and malignant liver lesions but no distinction between ICC and HCC was reported. High accuracy for segmentation and classification of liver lesions were also reported (30,31). In the study by Othman et al. (30), the authors developed two models integrating several pre-trained DL models for tumor segmentation in order to take advantage of the characteristics of the different methods. The automated CNN model issued was tested on the public image datasets LiTS17 (32) and IRCADb-01 (33), with a 99.5% accuracy in diagnosing benignant and HCC lesions. In the second study (31), the authors produced an automated CNN model focused on

analyzing the changes in Hounsfield Unit on liver lesions in the different contrast-enhanced phasis of a CT-scan. Tumor segmentation process resulted to have 100% accuracy while tumor classification accuracy was 95.1%. Despite presenting impressive accuracy results, these methods reliability is affected by the absence of an external validation and by the lack of information regarding the characteristics of the training and testing dataset used.

Although most of the literature concerning AI algorithms to segment and classify liver lesions is based on CT-scan images, some authors also focused on US images. A meta-analysis by Campello et al. (34) explored the accuracy of AI applied to US to classify benignant and malignant liver lesions. Overall, sensitivity and specificity of US to identify malignant liver lesions were 81.7% and 84.8% respectively, while they increased to 87.1% and 87% with contrast-enhanced US. Interestingly, when the meta-analysis focused exclusively on ML and DL applied to contrast-enhanced US, pooled sensitivity and specificity increased to 92.4% and 88.2% respectively. However, the only recent randomized controlled study analyzing the benefit of AI associated to US in the surveillance for HCC in patients affected by liver cirrhosis, chronic HBV infection, or ≥ stage 3 fibrosis, has given less encouraging results (35). A CNN model based on a supervised learning method and trained on a dataset of 8510 US images of liver lesions was used by expert and non-expert radiologist in order to implement liver lesions diagnosis. While AI significantly increased the lesion detection rate for non-expert operators (36.9% vs. 21.4%), no significant difference was found for expert radiologists (66.7% vs. 66.3%) overall and for HCC specifically (70.5% vs. 70.5%).

The association of AI and MR, despite the fact that MR is known to perform better than CT-scan at least for HCC detection (36), has been less explored. Stollmayer et al. (37) elaborated a CNN model to perform differential diagnosis among focal nodular hyperplasia, HCC, and liver metastases from gadoxetate disodium-enhanced arterial, portal, and hepatobiliary phase 2D and 3D MR images. Combining the different sequences

in multi-channel images, sensitivity and specificity performance for detecting and correctly classifying HCC were 80% and 100%, respectively, using the 2D images, which was better than what was obtained with the correspondent 3D images (75% and 95%, respectively). Thus, with this research the Author highlighted how exploiting multiple phases of the MR permits to obtain elevated accuracy results with 2D images whose use is less computational demanding compared to 3D ones. Another example of the good use of the different phases of liver MR images is the one reported by Takenaga et al. (38). The proposed FCN algorithm detects and classifies liver lesions from 5 different MR phases included in the dataset. Among others benign and malignant lesions, the model permitted to correctly predict the presence of a HCC in 27/40 cases, while in the other 13/40 it mistakenly classified the lesion as a colorectal liver metastasis. No details concerning possible images' or lesions' characteristics that possibly affected the model's accuracy were provided. Liu et al. (39) also developed a DL algorithm for the diagnosis of intra-parenchymal liver masses using 2D MR images. In this case, no multiphasic approach was used; instead, a preprocessing method based on a radiologist manual definition of the region of interest was adopted in order to permit the algorithm to classify between HCC and ICC. The reported overall accuracy of this method was 92.26%.

As diagnosis is not only based on imaging techniques, research based on AI is also developing in the direction of using multiple sources of data to identify with the best possible precision cancer, although very few has been published concerning liver cancer till now. Khan et al. (40) recently made a step in this direction producing a multi-modal DL network implementing portal-phase CT-scan images and pathological data in order to train a model capable of recognizing HCC and colorectal liver metastases. For HCC, model's precision was 96.47% and 95.23% whether the patient had single or multiple lesions, respectively.

Despite the fact that current literature on liver cancer and AI is almost exclusively centered

on DL methods, other AI methods have been used to improve the diagnostic accuracy of imaging tests. This is mainly based on the concept of radiomics, a field of quantitative imaging that uses imaging features to describe tumor phenotypes. An example of the application of AI to liver cancer radiomics is represented by the research by Mahmoudi et al. (41) and Lysdahlgaard (42). In the first case, cluster analysis and different ML models are applied to a dataset of liver CT-scans of patients with either HCC or ICC. Before analysis, an expert radiologist performs manual segmentation of the tumor to focus the analysis for the research of radiomics features only on the tumor region of interest. The best results were obtained with a logistic regression model that presented 84.6% precision in tumor classification. With an equivalent process but selecting radiomics features with principal component analysis, Lysdahlgaard (42) reported 100% accuracy in liver cancer classification between primary liver cancer and healthy liver when using ML models as support vector machine and logistic regression on 131 portal phase CT-scans. Although these examples show an elevated reliability of these methods for the classification problem, compared to DL, they often demand a more important manual intervention to preprocess the input images. Moreover, they cannot be used for preemptive tumor identification and segmentation that remain a completely human-dependent task. Finally, it is worth to say that radiomics has some important intrinsic limitations that can affect models' performances (43): 1) radiomic feature quantification is sensitive to a number of technical factors such as modalities, acquisition modes, reconstruction parameters, smoothing, and segmentation thresholds; 2) many radiomic features are unstable; 3) images voxel intensity has to be discretized to improve computation, thus modifying the quantification of radiomic features; 4) the same radiomic feature can be differently computed in different studies; 5) some radiomics features depend on the size fo the segmentation which may affect the quantification and generate heterogeneity between different size lesions.

Ultimately, a totally different application of AI in liver imaging for cancer diagnosis is described by Bae et al (44) and Cao et al. (45). Through DL models, the Authors obtained a significantly reduced liver image noise in low radiation dose images compared to standard dose ones without affecting images' quality and clinical usability. These results show the important benefits of AI on the possibility to reduce patients' radiation exposure which could be extremely advantageous for those patients requiring repeated exams for cancer screening.

Pathological and omics models

The advances in biomedical science of the last decades have highlighted how cancer occurrence is the combination not only of genes modifications, but also of posttranslational and epigenetics protein modifications and the consequent metabolic changes involved. The relative biomedical sciences genomic, proteomic, and metabolomic are reunited in the wider term omics. Several studies have highlighted the interest of AI to analyze omics data and identify patterns that can be useful for cancer diagnosis, prognosis, and treatment (46).

Pathological images can be digitalized and used as an input to train AI models to build tools capable of identifying liver cancer. Dievernich et al. (47), for example, used digitalized sections stained with hemaotxylin and eosin or with immunofluorescence to train a feed forward neural network and a CNN to identify HCC area on sections. The characteristics used as input for the CNN model were the marker expression pattern from immunofluorescence, the cell densities, the area of the cell nuclei, and the lengths of the minor and major axes of the nuclei. This model showed 75% accuracy with low sensitivity (62%) and elevated specificity (87%). Similarly, Diao et al. (48) used hematoxylin and eosin stained whole-slide images with non-tumoral/tumoral annotations to feed a CNN capable to automatically identify HCC area. Dice's coefficient , which is a spatial overlap index and reproducibility validation metric, was 0.861 for this model, which was greater

than the classification manually made by a junior and an intermediate pathologist, and slightly inferior to the one performed by a senior pathologist. However, one tile's classification took an average of 7 seconds for the model compared to 10 minutes for the pathologists, thus permitting an important gain in time. The same objective was also pursued by Aatresh et al. (49) whose proposed CNN called LiverNet is intended not only to identify HCC on hematoxylin and eosin stained slides, but also to classify it in subtypes following the Edmondson and Steiner classification which predict recurrence and survival rate (50). Trained and tested on two public liver pathologic datasets, the proposed model showed 91% and 97% accuracy depending on the dataset used for test. However, as declared by the authors, the datasets exploited are restricted in size and variance which can affect the performance of the model in a clinical situation with increased tumor-characteristics variance.

While the previous research focused on cells features alone or combined with omics characteristics, Zhang et al. (51) exclusively recurred to RNA-sequencing, DNA methylation, and copy number variation data to build a patient similarity network in order to identify omics patterns associated to liver cancer. Then, after pre-processing, these data are used as input of a graph convolutional network, a variety of CNN that can be fed with data represented as graphs, that classic CNN cannot handle. The graph convolutional network finally classifies patients as having or not liver cancer. The proposed model performance was compared on the same dataset to other similar networks produced to identify other cancers based on omics data and to some ML algorithms resulting to be superior with an accuracy of 98.57%. This result with multi-omics data was also superior compared to the same model applied to single omics features. Another highly performing model based, in this case, only on genomics features has been realized by Cheng et al. (52) using ML. Seven hundred and forty-two genetic pathways were used to establish a diagnostic model for HCC. The area under the curve of the model was 0.987 and 0.992 in

the internal and external validation, respectively. The algorithm also allowed to identify 12 genes differentially expressed between HCC and non-cancer tissue related to prognosis whose analysis results are reported in the related section.

Regarding ICC, less literature is available on this subject. One of the most challenging issues in ICC diagnosis is to discern it from colorectal metastases. For this reason, Albrecht et al. (53) developed a CNN model fed with hematoxylin and eosin whole-slide images of patients with either ICC or colorectal metastases. The model was initially trained and tested on a clinical dataset of 456 patients and then externally validated on 159 patients from another institution. The authors reported impressive model accuracy values (96.5% and 98,9% at the internal and external validation respectively) enriched by the adjunct value of the external validation and the comparison of the model to 6 expert pathologists whose accuracy was surpassed.

Author	Country	Data	Patients (n)	Model	Setting	Outcome					
						Se (%)	Sp (%)	Accuracy (%)	AUC	DSC	F1
Imaging models											
Midya, 2023[24]	USA	CT-scan, portal	814	CNN, semiautomatic	Liver nodules	96	98	96.3			
Wang, 2021[25]	China	CT-scan, triphasic	7512	CNN, automatic	Liver nodules (HCC)	89.4	74	81.3	0.883		
Gao, 2021[26]	China	CT-scan, multiphasic	723	CNN, RNN	HCC, ICC, CRLM			82.9			
Zhang, 2022[27]	China	CT-scan, delayed	58	CNN, automatic	Liver nodules (PLC)			96.5			
Othman, 2022[28]	Saudi Arabia	CT-scan, multiphasic	150	CNN, automatic	Liver nodules (HCC)			99.5		0.561	
Phan, 2022[29]	Vietnam	CT-scan, multiphasic	2000	CNN, automatic	Liver nodules (HCC)			95.1			
Campello, 2023[32]	Brazil	CEUS	-	ML, DL	Liver nodules (HCC)	87.1	87				
Tiyarattanachai, 2023[33]	Thailand	US	504	CNN	HCC			70.5			
Stollmayer, 2021[35]	Hungary	MR, multiphasic	69	CNN	FNH, HCC, Met	100	89				
Takenaga, 2021[36]	Japan	MR, multiphasic	184	FCN	Liver nodules (HCC)			67.5			
Liu, 2023[37]	China	MR, T2WI	112	SFFNet	HCC, ICC			92.3	0.968		
Khan, 2023[38]	Canada	CT-scan, portal/pathology	68	CNN, automatic	HCC, Met	96.8	97.8				0.966
Mahmoudi, 2023[39]	Germany	CT-scan radiomics/clinical	94	ML	HCC, ICC				0.819		0.786
Lysdahlgaard, 2022[40]	Denmark	CT-scan radiomics	131	ML	PLC	100	100	100			
Pathological and omics models											
Dievernich, 2023[45]	Germany	Pathology, clinical	12	CNN, automatic	HCC	62	87	75			
Diao, 2021[46]	China	Pathology	100	CNN, semiautomatic	HCC					0.861	
Aatresh, 2021[47]	India	Pathology	-	CNN, automatic	HCC			91-97			
Zhang, 2022[49]	China	Omics	1613	CNN, automatic	HCC			98.6	0.986		0.986
Cheng, 2022[50]	China	Omics	-	ML	HCC	84.9		86.8	0.992		0.918
Albrecht, 2023[51]	Germany	Omics	615	CNN, automatic	ICC, Met			98.9			

Table 2: Characteristics of the studied describing AI-driven models to perform primary liver cancer diagnosis. CT: computed tomography; MR: magnetic resonance; CEUS: contrast-enhanced ultrasonography; DL: deep learning; ML: machine learning; CNN: convolutional neural network; RNN: recurrent neural network; FCN: fully convolutional network; SFFNet: spatial field fusion network; HCC: hepatocellular carcinoma; ICC: intrahepatic cholangiocarcinoma; PLC: primary liver cancer; CRLM: colorectal liver metastases; Met: liver metastases; FNH: focal nodular hyperplasia; Se: sensitivity; Sp: specificity; AUC: area under the curve; DSC: Dice's coefficient.

Conclusion

AI-driven models, from ML to DL, represent one major subject of research development in many medical fields at the moment. The applications of AI in the setting of primary liver cancer are multiple and authors are constantly developing tools that can assist or even substitute clinicians in their daily practice. The results reported concerning models' performance are impressive and it is hard to imagine that future clinical practice could not rely on this technology.

However, this review also highlights many limitations that need to be overcome. First, there is still a gap between model development and real-life implementation. Many publications on AI and liver cancer simply focus on the performance of AI models in executing tasks (i.e.: liver and tumor segmentation) but applicability and clinical benefits need to be evaluated. For example, this issue becomes manifest in screening models that are developed on a selected cohort of patients with excellent performance, but that can lack of accuracy and reproducibility when applied to different populations.

Conversely, in other situations such as the development of models based on imaging data to discern from benign and malignant liver lesions, the inclusion of an excessively heterogeneous population affects model applicability. Diagnostic imaging of patients with different grade and causes of chronic liver disease are very different and training a model on these populations can reduce performances when focusing only on a subgroup of patients. Bias related to patients selection and inclusion are quite common in AI-models for screening and diagnosis (54).

Secondly, most of the models' performance characteristics are reported without a complete validation. One of the commonest weaknesses identified is the absence of an external validation, which permits to overcome bias associated to patients' selection and practice standardization. Furthermore, to assess one's model superiority, authors tend to compare their proposed model's performance to other published algorithms that were

initially developed for other settings (i.e.: different type of cancer, patients with other characteristics) and that frequently did not undergo correct validation in turn.

Thirdly, with this review we pointed out the absence of consensus on how models' performance should be reported. Different research with similar aims tend to use plenty of different performance measures to describe their model, thus making it impossible to simply compare their capacities (55). Moreover, many models present wide discrepancies between sensitivity and specificity, when reported. The different role of these two measures is of primary importance and their relevance changes depending on the aim of the proposed tool (i.e.: elevated sensitivity is auspicable for screening tests while high specificity distinguish accured diagnostic tests) and this is rarely discussed when models' performance is presented.

Finally, it is rare to find research outlining a full and detailed model development process from patients' selection and algorithm choice and application to final validation and practical implementation. When the report focuses on model development frequently few or no detail is given on how it can be used in daily practice; on the other hand, when it points on implementation, little information is given on how the model was produced. This makes this research non-reproducible and slow further developments.

AI-driven models will surely become the basis of the management of liver cancer but a more rigorous model development practice is needed to achieve this objective.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209–49.
2. Massarweh NN, El-Serag HB. Epidemiology of Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma. Cancer Control. 2017;24(3):1–11.

3. Rumgay H, Arnold M, Ferlay J, Lesi O, Cabasag CJ, Vignat J, et al. Global burden of primary liver cancer in 2020 and predictions to 2040. J Hepatol [Internet]. 2022;77(6):1598–606. Available from: https://doi.org/10.1016/j.jhep.2022.08.021
4. Grazzini G, Cozzi D, Flammia F, Grassi R, Agostini A, Belfiore MP, et al. Hepatic tumors: Pitfalls in diagnostic imaging. Acta Biomed. 2020;91(8-S):9–17.
5. Reig M, Forner A, Rimola J, Ferrer-Fàbrega J, Burrel M, Garcia-Criado Á, et al. BCLC strategy for prognosis prediction and treatment recommendation: The 2022 update. J Hepatol [Internet]. 2022;76(3):681–93. Available from: https://doi.org/10.1016/j.jhep.2021.11.018
6. Bressler L, Bath N, Manne A, Miller E, Cloyd JM. Management of locally advanced intrahepatic cholangiocarcinoma: a narrative review. Chinese Clin Oncol. 2023;12(2).
7. Lampimukhi M, Qassim T, Venu R, Pakhala N, Mylavarapu S, Perera T, et al. A Review of Incidence and Related Risk Factors in the Development of Hepatocellular Carcinoma. 2023;15(11).
8. El-Diwany R, Pawlik TM, Ejaz A. Intrahepatic Cholangiocarcinoma. Surg Oncol Clin N Am [Internet]. 2019;28(4):587–99. Available from: https://doi.org/10.1016/j.soc.2019.06.002
9. Sheka AC, Adeyi O, Thompson J, Hameed B, Crawford PA, Ikramuddin S. Nonalcoholic Steatohepatitis: A Review. JAMA - J Am Med Assoc. 2020;323(12):1175–83.
10. Gofton C, Upendran Y, Zheng MH, George J. MAFLD: How is it different from NAFLD? Clin Mol Hepatol. 2023;29:S17–31.
11. Okanoue T, Yamaguchi K, Shima T, Mitsumoto Y, Katayama T, Okuda K, et al. Artificial intelligence/neural network system that accurately diagnoses hepatocellular carcinoma in nonalcoholic steatohepatitis. Hepatol Res. 2023;(July).

12. McGlynn KA, Petrick JL, El-Serag HB. Epidemiology of the hepatocellular carcinoma. Hepatology. 2021;73(Suppl 1):4–13.
13. Yu JH, Cho SG, Jin Y, Lee J. The best predictive model for hepatocelullar carcinoma in patients with chronic hepatitis B infection. Clin Mol Hepatol. 2022;28:351–61.
14. Lee HW, Kim H, Park T, Park SY, Chon YE, Seo YS, et al. A machine learning model for predicting hepatocellular carcinoma risk in patients with chronic hepatitis B. Liver Int. 2023;43(8):1813–21.
15. Kim HY, Lampertico P, Nam JY, Lee HC, Kim SU, Sinn DH, et al. An artificial intelligence model to predict hepatocellular carcinoma risk in Korean and Caucasian patients with chronic hepatitis B. J Hepatol [Internet]. 2022;76(2):311–8. Available from: https://doi.org/10.1016/j.jhep.2021.09.025
16. Papatheodoridis G, Dalekos G, Sypsa V, Yurdaydin C, Buti M, Goulis J, et al. PAGE-B predicts the risk of developing hepatocellular carcinoma in Caucasians with chronic hepatitis B on 5-year antiviral therapy. J Hepatol [Internet]. 2016;64(4):800–6. Available from: http://dx.doi.org/10.1016/j.jhep.2015.11.035
17. Yang HI, Yuen MF, Chan HLY, Han KH, Chen PJ, Kim DY, et al. Risk estimation for hepatocellular carcinoma in chronic hepatitis B (REACH-B): Development and validation of a predictive score. Lancet Oncol [Internet]. 2011;12(6):568–74. Available from: http://dx.doi.org/10.1016/S1470-2045(11)70077-8
18. Lee HW, Kim SU, Park JY, Kim DY, Ahn SH, Han KH, et al. External validation of the modified PAGE-B score in Asian chronic hepatitis B patients receiving antiviral therapy. Liver Int. 2019;39(9):1624–30.
19. Hu C, Iyer RK, Juran BD, McCauley BM, Atkinson EJ, Eaton JE, et al. Predicting cholangiocarcinoma in primary sclerosing cholangitis: using artificial intelligence, clinical and laboratory data. BMC Gastroenterol [Internet]. 2023;23(1):1–13. Available from: https://doi.org/10.1186/s12876-023-02759-7

20. Mitchell DG, Bruix J, Sherman M, Sirlin CB. LI-RADS (Liver Imaging Reporting and Data System): Summary, discussion, and consensus of the LI-RADS Management Working Group and future directions. Hepatology. 2015;61(3):1056–65.

21. Heimbach JK, Kulik LM, Finn RS, Sirlin CB, Abecassis MM, Roberts LR, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. Hepatology. 2018;67(1):358–80.

22. Galle PR, Forner A, Llovet JM, Mazzaferro V, Piscaglia F, Raoul JL, et al. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. J Hepatol [Internet]. 2018;69(1):182–236. Available from: https://doi.org/10.1016/j.jhep.2018.03.019

23. Chen X, Li M, Guo R, Liu W, Li J, Zong X, et al. The diagnostic performance of contrast-enhanced CT versus extracellular contrast agent-enhanced MRI in detecting hepatocellular carcinoma: direct comparison and a meta-analysis. Abdom Radiol [Internet]. 2022;47(6):2057–70. Available from: https://doi.org/10.1007/s00261-022-03484-7

24. Moris D, Palta M, Kim C, Allen PJ, Morse MA, Lidsky ME. Advances in the treatment of intrahepatic cholangiocarcinoma: An overview of the current and future therapeutic landscape for clinicians. CA Cancer J Clin. 2023;73(2):198–222.

25. Bridgewater J, Galle PR, Khan SA, Llovet JM, Park JW, Patel T, et al. Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. J Hepatol [Internet]. 2014;60(6):1268–89. Available from: http://dx.doi.org/10.1016/j.jhep.2014.01.021

26. Midya A, Chakraborty J, Srouji R, Narayan RR, Boerner T, Zheng J, et al. Computerized Diagnosis of Liver Tumors From CT Scans Using a Deep Neural Network Approach. IEEE J Biomed Heal Informatics. 2023;27(5):2456–64.

27. Wang M, Fu F, Zheng B, Bai Y, Wu Q, Wu J, et al. Development of an AI system for accurately diagnose hepatocellular carcinoma from computed tomography imaging data. Br J Cancer [Internet]. 2021;125(8):1111–21. Available from: http://dx.doi.org/10.1038/s41416-021-01511-w

28. Gao R, Zhao S, Aishanjiang K, Cai H, Wei T, Zhang Y, et al. Deep learning for differential diagnosis of malignant hepatic tumors based on multi-phase contrast-enhanced CT and clinical data. J Hematol Oncol [Internet]. 2021;14(1):1–7. Available from: https://doi.org/10.1186/s13045-021-01167-2

29. Zhang H, Luo K, Deng R, Li S, Duan S. Deep Learning-Based CT Imaging for the Diagnosis of Liver Tumor. Comput Intell Neurosci. 2022;2022.

30. Othman E, Mahmoud M, Dhahri H, Abdulkader H, Mahmood A, Ibrahim M. Automatic Detection of Liver Cancer Using Hybrid Pre-Trained Models. Sensors. 2022;22(14).

31. Phan AC, Cao HP, Le TNH, Trieu TN, Phan TC. Improving liver lesions classification on CT/MRI images based on Hounsfield Units attenuation and deep learning. Gene Expr Patterns [Internet]. 2023;47(November 2022):119289. Available from: https://doi.org/10.1016/j.gep.2022.119289

32. Bilic P, Christ P, Li HB, Vorontsov E, Ben-Cohen A, Kaissis G, et al. The Liver Tumor Segmentation Benchmark (LiTS). Med Image Anal. 2023;84.

33. Soler L, Hostettler A, Agnus V, Charnoz A, Fasquel J, Moreau J, et al. 3D Image Reconstruction for Comparison of Algorithm Database: a Patient-Specific Anatomical and Medical Image Database. Tech Rep. 2010;

34. Campello CA, Castanha EB, Vilardo M, Staziaki P V., Francisco MZ, Mohajer B, et al. Machine learning for malignant versus benign focal liver lesions on US and CEUS: a meta-analysis. Abdom Radiol [Internet]. 2023;48(10):3114–26. Available from: https://doi.org/10.1007/s00261-023-03984-0

35. Tiyarattanachai T, Apiparakoon T, Chaichuen O, Sukcharoen S, Yimsawad S, Jangsirikul S, et al. Artificial intelligence assists operators in real-time detection of focal liver lesions during ultrasound: A randomized controlled study. Eur J Radiol [Internet]. 2023;165(May):110932. Available from: https://doi.org/10.1016/j.ejrad.2023.110932

36. Guo J, Seo Y, Ren S, Hong S, Lee D, Kim S, et al. Diagnostic performance of contrast-enhanced multidetector computed tomography and gadoxetic acid disodium-enhanced magnetic resonance imaging in detecting hepatocellular carcinoma: direct comparison and a meta-analysis. Abdom Radiol. 2016;41(10):1960–72.

37. Stollmayer R, Budai BK, Tóth A, Kalina I, Hartmann E, Szoldán P, et al. Diagnosis of focal liver lesions with deep learning-based multichannel analysis of hepatocyte-specific contrast-enhanced magnetic resonance imaging. World J Gastroenterol. 2021;27(35):5978–88.

38. Takenaga T, Hanaoka S, Nomura Y, Nakao T, Shibata H, Miki S, et al. Multichannel three-dimensional fully convolutional residual network-based focal liver lesion detection and classification in Gd-EOB-DTPA-enhanced MRI. Int J Comput Assist Radiol Surg [Internet]. 2021;16(9):1527–36. Available from: https://doi.org/10.1007/s11548-021-02416-y

39. Liu Y, Wang B, Mo X, Tang K, He J, Hao J. A Deep Learning Workflow for Mass-Forming Intrahepatic Cholangiocarcinoma and Hepatocellular Carcinoma Classification Based on MRI. 2023;529–44.

40. Khan RA, Fu M, Brent B, Luo Y, Wu FX. A multi-modal deep neural network for multi-class liver cancer diagnosis. Neural Networks [Internet]. 2023;165:553–61. Available from: https://doi.org/10.1016/j.neunet.2023.06.013

41. Mahmoudi S, Bernatz S, Ackermann J, Koch V, Dos Santos DP, Grünewald LD, et

al. Computed Tomography Radiomics to Differentiate Intrahepatic Cholangiocarcinoma and Hepatocellular Carcinoma. Clin Oncol. 2023;35(5):e312–8.

42. Lysdahlgaard S. Comparing Radiomics features of tumour and healthy liver tissue in a limited CT dataset: A machine learning study. Radiography [Internet]. 2022;28(3):718–24. Available from: https://doi.org/10.1016/j.radi.2022.03.015
43. Yip SSF, Aerts HJWL. Applications and limitations of radiomics. Phys Med Biol. 2016;61(13):R150–66.
44. Bae JS, Lee JM, Kim SW, Park S, Han S, Yoon JH, et al. Low-contrast-dose liver CT using low monoenergetic images with deep learning–based denoising for assessing hepatocellular carcinoma: a randomized controlled noninferiority trial. Eur Radiol. 2023;33(6):4344–54.
45. Cao L, Liu X, Li J, Qu T, Chen L, Cheng Y, et al. A study of using a deep learning image reconstruction to improve the image quality of extremely low-dose contrast-enhanced abdominal CT for patients with hepatic lesions. Br J Radiol. 2021;94(1118):1–6.
46. Bradlow L, Carruba G. Cancer OMICS. OMICS. 2011;15(3):91–2.
47. Dievernich A, Stegmaier J, Achenbach P, Warkentin S, Braunschweig T, Neumann UP, et al. A Deep-Learning-Computed Cancer Score for the Identification of Human Hepatocellular Carcinoma Area Based on a Six-Colour Multiplex Immunofluorescence Panel. Cells. 2023;12(7):1–17.
48. Diao S, Tian Y, Hu W, Hou J, Lambo R, Zhang Z, et al. Weakly Supervised Framework for Cancer Region Detection of Hepatocellular Carcinoma in Whole-Slide Pathologic Images Based on Multiscale Attention Convolutional Neural Network. Am J Pathol [Internet]. 2022;192(3):553–63. Available from: https://doi.org/10.1016/j.ajpath.2021.11.009
49. Aatresh AA, Alabhya K, Lal S, Kini J, Saxena PP. LiverNet: efficient and robust deep

learning model for automatic diagnosis of sub-types of liver hepatocellular carcinoma cancer from H&E stained liver histopathology images. Int J Comput Assist Radiol Surg [Internet]. 2021;16(9):1549–63. Available from: https://doi.org/10.1007/s11548-021-02410-4

50. Zhou L, Rui JA, Zhou WX, Wang S Bin, Chen SG, Qu Q. Edmondson-Steiner grade: A crucial predictor of recurrence and survival in hepatocellular carcinoma without microvascular invasio. Pathol Res Pract [Internet]. 2017;213(7):824–30. Available from: http://dx.doi.org/10.1016/j.prp.2017.03.002
51. Zhang G, Peng Z, Yan C, Wang J, Luo J, Luo H. A novel liver cancer diagnosis method based on patient similarity network and DenseGCN. Sci Rep [Internet]. 2022;12(1):1–10. Available from: https://doi.org/10.1038/s41598-022-10441-3
52. Cheng B, Zhou P, Chen Y. Machine-learning algorithms based on personalized pathways for a novel predictive model for the diagnosis of hepatocellular carcinoma. BMC Bioinformatics [Internet]. 2022;23(1):1–19. Available from: https://doi.org/10.1186/s12859-022-04805-9
53. Albrecht T, Rossberg A, Albrecht JD, Nicolay JP, Straub BK, Gerber TS, et al. Deep Learning–Enabled Diagnosis of Liver Adenocarcinoma. Gastroenterology [Internet]. 2023;(August):1–14. Available from: https://doi.org/10.1053/j.gastro.2023.07.026
54. Kelly BS, Judge C, Bollard SM, Clifford SM, Healy GM, Aziz A, et al. Radiology artificial intelligence: a systematic review and evaluation of methods (RAISE). Eur Radiol. 2022;32(11):8054.
55. Reinke A, Tizabi MD, Baumgartner M, Eisenmann M, Heckmann-Nötzel D, Kavur AE, et al. Understanding metric-related pitfalls in image analysis validation. 2023;1(1). Available from: http://arxiv.org/abs/2302.01790