



Systematic Review

Artificial Intelligence-Based Opportunities in Liver Pathology—A Systematic Review

Pierre Allaume ^{1,*}, Noémie Rabilloud ², Bruno Turlin ^{1,3} , Edouard Bardou-Jacquet ^{3,4}, Olivier Loréal ³, Julien Calderaro ^{5,6}, Zine-Eddine Khene ^{2,7}, Oscar Acosta ², Renaud De Crevoisier ^{2,8}, Nathalie Rioux-Leclercq ¹, Thierry Pecot ⁹  and Solène-Florence Kammerer-Jacquet ^{1,2}

- ¹ Department of Pathology CHU de Rennes, Rennes 1 University, Pontchaillou Hospital, 2 rue Henri Le Guilloux, CEDEX 09, 35033 Rennes, France
 - ² Impact TEAM, Laboratoire Traitement du Signal et de l'Image (LTSI) INSERM, Rennes 1 University, Pontchaillou Hospital, 35033 Rennes, France
 - ³ Research Unit n°UMR1341 NuMeCan–Nutrition, Métabolismes et Cancer, Rennes 1 University, Pontchaillou Hospital, 2 rue Henri Le Guilloux, CEDEX 09, 35033 Rennes, France
 - ⁴ Department of Liver Diseases CHU de Rennes, Rennes 1 University, Pontchaillou Hospital, 35033 Rennes, France
 - ⁵ Assistance Publique-Hôpitaux de Paris, Department of Pathology Henri Mondor, 94000 Créteil, France
 - ⁶ INSERM U955, Team Pathophysiology and Therapy of Chronic Viral Hepatitis and Related Cancers, 94000 Créteil, France
 - ⁷ Department of Urology, CHU de Rennes, Rennes 1 University, Pontchaillou Hospital, 2 rue Henri Le Guilloux, CEDEX 09, 35033 Rennes, France
 - ⁸ Department of Radiotherapy, Centre Eugène Marquis, 35033 Rennes, France
 - ⁹ Biosit Platform UAR 3480 CNRS US18 INSERM U955, Rennes 1 University, Pontchaillou Hospital, 35033 Rennes, France
- * Correspondence: pierre.allaume@chu-rennes.fr



Citation: Allaume, P.; Rabilloud, N.; Turlin, B.; Bardou-Jacquet, E.; Loréal, O.; Calderaro, J.; Khene, Z.-E.; Acosta, O.; De Crevoisier, R.; Rioux-Leclercq, N.; et al. Artificial Intelligence-Based Opportunities in Liver Pathology—A Systematic Review. *Diagnostics* **2023**, *13*, 1799. <https://doi.org/10.3390/diagnostics13101799>

Academic Editor: Dechang Chen

Received: 5 April 2023

Revised: 4 May 2023

Accepted: 12 May 2023

Published: 19 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Background: Artificial Intelligence (AI)-based Deep Neural Networks (DNNs) can handle a wide range of applications in image analysis, ranging from automated segmentation to diagnostic and prediction. As such, they have revolutionized healthcare, including in the liver pathology field. Objective: The present study aims to provide a systematic review of applications and performances provided by DNN algorithms in liver pathology throughout the Pubmed and Embase databases up to December 2022, for tumoral, metabolic and inflammatory fields. Results: 42 articles were selected and fully reviewed. Each article was evaluated through the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool, highlighting their risks of bias. Conclusions: DNN-based models are well represented in the field of liver pathology, and their applications are diverse. Most studies, however, presented at least one domain with a high risk of bias according to the QUADAS-2 tool. Hence, DNN models in liver pathology present future opportunities and persistent limitations. To our knowledge, this review is the first one solely focused on DNN-based applications in liver pathology, and to evaluate their bias through the lens of the QUADAS2 tool.

Keywords: digital pathology; liver; hepatology; deep learning; artificial intelligence; performance metrics

1. Introduction

1.1. Introduction to AI

Artificial Intelligence (AI) is probably the next revolution in pathology. “Artificial Intelligence” is a broad term encompassing many different approaches for problem-solving. Among them, Machine Learning (ML), and Deep Neural Networks (DNN) in particular, are capable of handling huge amounts of data by increasingly complex mathematical systems that have proven time and time again to at least perform similarly to pathologists for the segmentation of clinically relevant regions and classification of tumors. Moreover, they might contribute to standardizing care by reducing inter-observer lack of reproducibility.

Inspired by neuroanatomy, algorithms known as Deep Learning Algorithms (DLA) rely on architectures known as Deep Neural Networks (DNN) and are able to produce astounding results in imagery recognition and classification. As a mostly image-based discipline, pathology is particularly suitable for the application of such AIs. Therefore, there is a growing field of interest, both academic and industrial, for the development of efficient, fast, robust and precise AI-based solutions for pathologists, among them liver pathology [1].

1.2. Principles of Deep Neural Networks Algorithms

In pathology, most DNN-based tasks can be separated into three categories: segmentation, classification and prediction [2]. Briefly, segmentation algorithms aim to identify anatomical structures on a virtual slide. For example, they may be useful for detecting cell boundaries, tumoral areas or structures such as portal tracts or fibrotic septa [3,4]. Classification algorithms aim to discriminate between, in general, binary modalities, “tumoral vs. non tumoral” being the most common example. Non-binary modalities can also be investigated through the lens of artificial intelligence, some examples in liver pathology being fibrosis grading and hepatitis severity assessment. Prediction algorithms function similarly to classification algorithms, taking input data and proposing an output answer. The difference between classification and prediction algorithms lies in how the “ground truth” is established; in a classification model, “ground truth” can be established directly from the input data (in pathology: a Whole Slide Image), whereas in a prediction model other techniques are to be performed (such as molecular testing). Moreover, “prediction” models include a temporal aspect, because their predicted outcome will not be measurable immediately (either because additional analysis is required, or because clinical follow-up is needed).

Notwithstanding their assigned goal, all DNN-based algorithms share some similarities [5]. DNN systems can be either supervised (Supervised Learning SL) or unsupervised (UL). SL requires the pre-definition of a “ground truth” through the annotations of the developmental dataset. In UL, the model is intended to find a pattern by itself through the numerous unannotated examples it is given, so that it is generalizable to the data ensemble. Development of a DNN-based algorithm is divided into training, testing and validation phases. On the training dataset, the algorithm is given inputs (annotated example of interest if the model is supervised) so as to determine the optimal combinations of features giving the most correct outputs. The validation step aims to fine-tune the hyper-parameters of the algorithm (for example, the number of artificial neuron layers) and to prevent overfitting on the training dataset. Finally, the testing dataset evaluates the model’s performance metrics. In order to obtain high precision, a huge amount of data is often required. Moreover, in supervised learning, the dataset must be labeled to establish a “ground truth” for training the model. The SL can be either “supervised”, with many refined annotations made by an expert (for example, manual definition of precise Region Of Interest for a segmentation model) or “semi-supervised” (for example, a “tumor” or “non-tumor” label for a classification model). Hence, establishing a sufficiently large and labeled dataset is time-consuming and may introduce bias, such as lack of reproducibility by experts [6]. Some data augmentation principles (mathematical transformations of images such as artificial blurring, rotations and inversions, changes of color...) have been proposed and are commonly used to try and mitigate such issues [5].

1.3. Evaluation of Algorithms by Performance Metrics

Throughout the literature, many performance metrics have been used to evaluate DNN algorithms. What they refer to and how they compare to one another is sometimes challenging. The most commonly used performance metrics and their major characteristics are summarized in Table 1. A 2×2 contingency table may be used to either present or explicit the metrics; such an example is given in Figure 1. It is to be noted that no performance metric is widely recognized as superior to the others, nor are they comparable to one another, and as such they are often used in conjunction. In ideal conditions, per-

formance assessment is to be performed on an external validation set, so as to generalize the results and not risk a bias of overfitting the algorithm to the developmental dataset and causing an inability to generalize performance with unseen data (staining differences, artefacts, etc.) [7].

Table 1. Most commonly used performance metrics used in the literature, with key features summarized.

| Performance Metric | Rational | Interpretation | Key Concepts | Main Applications |
|-----------------------------|--|---|--|---|
| Jaccard Index | Also known as the ratio of Intersection over Union (IoU) | $0 \leq x \leq 1$ the closer to 1 the best | Very simple to use, the Jaccard index is a way of conceptualizing accuracy for object detection. It quantifies the similarity of the algorithm vision with those of the annotated ground truth | object/area segmentation (may also be employed for binary classification) |
| Accuracy | Number of correct predictions (true positives and true negatives) divided by the total number of predictions | $0 \leq x \leq 1$ the closer to 1 the best | Very simple to use, accuracy quantifies the percentage of correct predictions by the algorithm. However, it is not adapted to imbalanced problems (where positive and negative proportions are greatly different) nor to problems where the “cost” of false positives/negatives must be taken into account (such as screening situations). Therefore, it may not be suited for evaluating algorithms in certain medical situations | classification/prediction |
| F1-score | Combines Precision (ratio of true positives/total positives predicted) and Recall (ratio of true positives/total positives in ground truth) | $0 \leq x \leq 1$ the closer to 1 the best enhancing Precision OR Recall leads to a better score | Useful for evaluating performance in situations where Accuracy would be misleading (see above). It is adapted to unbalanced problems. It is, however, difficult to interpret when low: is it because of low Precision (too much false positives) or low Recall (not enough true positives)? | classification/prediction |
| AUROC (or AUC) | Combines Recall (ratio of true positives/total positives in ground truth) and Fallout (ratio of false positives/total negatives in ground truth) | $0 \leq x \leq 1$ the closer to 1 the best enhancing Precision OR diminishing Fallout leads to a better score | Useful for evaluating the diagnostic ability of a (binary) classifier, because it takes both true positives and true negatives into account. Therefore (and contrary to F1-score), diminishing the number of false negatives is taken into account (which is of importance in screening situations) | classification/prediction |
| c-index (concordance index) | Generalization of AUROC for assessing the correct ranking of events | $0 \leq x \leq 1$ the closer to 1 the best | Adapted to datasets with censored data (survival studies, prediction of adverse events, etc.) | |

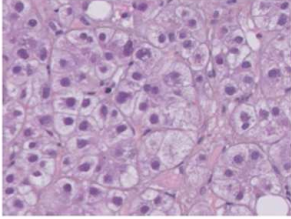
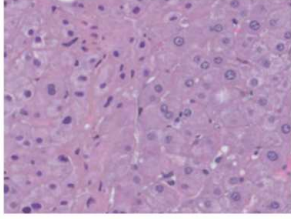
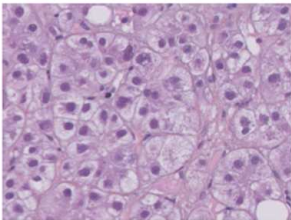
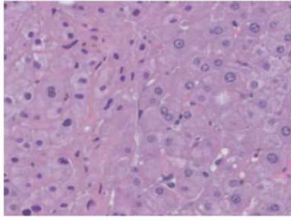
| | | Ground Truth | | |
|------------------|----------|---|---|---|
| | | Positive | Negative | |
| Algorithm Result | Positive | True Positive (TP) (ex : hepatocellular carcinoma is correctly detected)  | False Positive (FP) (ex : hepatocellular carcinoma is detected instead of non tumoral liver)  | Precision (Positive Predictive Value) = $TP/(TP+FP)$ |
| | Negative | False Negative (FN) (ex : non tumoral liver is detected instead of hepatocellular carcinoma)  | True Negative (TN) (ex : non tumoral liver is correctly detected)  | False Omission Rate = $TN/(FN+TN)$ |
| | | Recall (Sensitivity) = $TP/(TP+FN)$ | Fallout (Specificity) = $TN/(FP+TN)$ | |

Figure 1. Example of a 2×2 contingency table used for calculating performance metrics such as precision, recall and fallout. Histological images are courtesy of the CHU Pontchaillou, Rennes, France.

1.4. Aim of the Present Review

As Artificial Intelligence in healthcare is an ever-growing subject, many publications have aimed to review its impact in numerous medical fields, be it radiology, pathology or clinical specialties [8]. The present study aims to provide a synthetic and comprehensive review of applications provided by DNN algorithms in liver pathology throughout the literature, in the tumoral as well as metabolic and inflammatory fields, and to discuss the potential bias and opportunities opened by Artificial Intelligence for liver pathologists.

2. Materials and Methods

This systematic review follows the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) statement (Supplementary Material Table S1) [9].

2.1. Inclusion and Exclusion Criteria

All studies published until 12 January 2022 using deep learning in hepatology on histopathologic slides are included. Exclusion criteria were as follows: (1) not using Whole Slide Images (WSI) of human tissue slides; (2) not published in English; (3) review articles, editorials or other unrelated topics.

2.2. Online Registration Information

The search protocol was not registered online prior to data extraction.

2.3. Data Sources and Literature Search Strategy

The first author (A.P.) searched the Pubmed and Embase databases to identify studies up to 01 December 2022. The search terms used were as follows: (Liver OR Hepatic) AND

(artificial intelligence OR deep learning OR machine learning) AND (Whole slide image OR digital pathology OR pathomics OR pathomic) AND (Human). Manual selection of relevant articles through crosschecking references was additionally performed.

2.4. Studies Selection and Data Extraction

The first author screened the articles according to inclusion/exclusion criteria, removed duplicates and added relevant articles through crosschecking references. For each article, information regarding authors, year of publication, country, algorithm type (segmentation vs. classification vs. prediction), number of WSI used for development and validation, and the most pertinent performance outcomes were extracted. One author (A.P.) extracted data from each study and a second independent author (K-J.S-F.) validated the extracted data. All studies were separated and summarized in two tables: “tumoral” for all studies related to cancer segmentation, classification or prediction (Table 2), and “non-tumoral” for all other liver pathologies (Table 3). The features extracted included author and year of publication, first author’s country, category of algorithm, size of developmental and external validation sets, and main performance metrics as appropriate.

2.5. Assessment of the Risk of Bias and Applicability

Each article was evaluated keeping the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) guidelines in mind, so that any potential bias and limitations could be discussed [10]. Notably, the following criteria were taken into account to stratify as “high risk”: patient selection: developmental set <60 WSI, unicentric study; index test: no external validation set, no use of image augmentation principles; reference standard: definition of “ground truth” solely based on one expert, no use of standardized scoring system when appropriate, lightly-supervised or unsupervised model; flow and timing: for prospective studies only, patient exclusions from final analysis, multiple reference standard provided.

3. Results

3.1. Search Results

The result of the search yielded 167 articles (127 non-duplicated). After excluding articles based on title and abstract screening, 35 articles remained and were reviewed by full text screening. Seven articles were added manually by manual reference checking, amounting to 42 selected articles for systematic review. Twenty-two articles were labeled “tumoral” and 20 articles were labeled “non-tumoral” based on full text review (Figure 2).

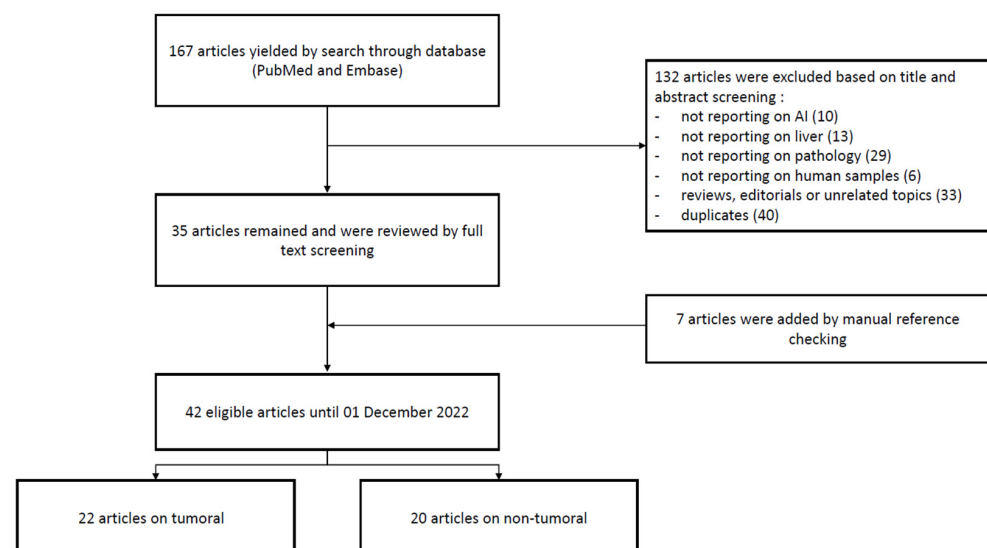


Figure 2. Flowchart of literature search.

3.2. Tumoral

Twenty-two articles were reviewed through full text screening (Table 2) [11–32]. No studies used prospectively collected data. The algorithms primary goal was either segmentation (6/22) [11–16], classification (6/22) [17–22] or prediction (10/22) [23–32]. The majority of studies (18/22) were focused on hepatocellular carcinoma. The others were respectively reporting on liver metastasis (3/22) [12,19,28] or cholangiocarcinoma (1/22) [32]. Six studies used The Cancer Genome Atlas (TCGA) database for either their developmental or external validation set (including the one from Liao et al., whose full-text access was unavailable through our institution [25]). More than half of the studies (13/22) tested their algorithms on at least one external validation set. The DNN models used in each study are reported in Supplementary Material Table S2.

Three published segmentation studies used the same developmental and external testing set, namely the Pathology Artificial Intelligence Platform (PAIP) dataset from the PAIP 2019 Challenge [11,13,14]. PAIP challenge's goal is to evaluate new and existing algorithms for automated liver cancer detection in WSI, either with or without peritumoral/intratumoral stroma-reaction, using a predefined fully annotated dataset from Seoul National University Hospital, South Korea [33]. As of the writing date of the present article, more than 1500 participating teams were numbered, and the PAIP 2019 Challenge is still ongoing. Among the published studies, Roy et al. demonstrated the highest performance, using a classification model to achieve an F1-score of 0.94 [13].

Cheng et al. proposed a DNN-classification model to help discriminate hepatic nodular lesions (HNL), comprising hepatocellular adenoma, hepatocellular carcinoma, dysplastic nodule and focal nodular hyperplasia [22]. Their model was trained and validated on HE biopsy specimens and showed a good discrimination power between (pre)neoplastic lesions and benign hepatocellular proliferation or background with an AUC of 0.94.

DNN-prediction models were shown to be able to predict genetic somatic mutations or signatures directly from HE WSI. Chen et al. demonstrated an algorithm able to predict CTNNB1, FMN2, TP53 and ZFX4 somatic status in HCC slides ailing from TCGA, with AUCs ranging from 0.71 to 0.89 [23]. Zeng et al. developed a DNN-model predicting activation of 6 immune gene signatures in advanced HCC directly from histology [30]. Those immune gene signatures were previously shown to be associated with better response/survival rates after nivolumab therapy (an anti PD-1 antibody).

Prediction of post-surgical recurrence-free interval or survival in HCC was also explored through DNN-based systems. Saillard et al. proposed a DNN-based system computing a risk score of independent prognostic value after hepatocellular surgery based on the resected tumor's WSI (c-index 0.70) [25]. That score was shown to be significant on an external validation dataset, even after stratification of patients on the disease's stage, vascular invasion and presence of satellite nodules, and to outperform a composite score based on clinical, biological and pathological features. In the same vein, Yamashita et al. developed a DNN approach of post-surgical recurrence risk stratification for HCC with WSI from the TCGA and validated their model on the Stanford Department of Pathology slide archive, achieving an AUC of 0.96 and a c-index of 0.67 on the external validation set with a 10-year follow-up.

Among the individual studies, the highest amount of WSI used for the developmental dataset was 2917 WSI in a predictive study from Chen et al. [29]. This study aimed to predict the existence of microvascular invasion in patients with HCC by a DNN analysis of the WSI on the primary tumor. Their model was deemed able to successfully and independently predict microvascular invasion in the external validation set with an AUC of 0.87. Furthermore, Chen et al. strived to implement their model into a clinical setting and showed that the presence of microvascular invasion as predicted by their DNN model is correlated with a poorer outcome.

Table 2. Key features of reported studies regarding the use of Deep Neural Networks (DNN) on Whole Image histology Slides (WSI) in the “tumoral” field.

| Authors (Year) [Reference] | Country | Algorithm Goal | Development Dataset | External Validation Dataset | Performance Metrics |
|-----------------------------------|-------------|------------------------------|---------------------|-----------------------------|-------------------------------------|
| Feng Y. et al. (2021) [11] | France | segmentation (HCC) | 60 WSI | 40 WSI | Jaccard score 0.90 F1 score 0.47 |
| Cancian P. et al. (2021) [12] | Italy | segmentation (LM) | 303 WSI | no | Jaccard score 0.89 |
| Roy M. et al. (2021) [13] | USA | segmentation (HCC) | 60 WSI | 40 WSI | F1 score 0.94 |
| Wang X. et al. (2021) [14] | China | segmentation (HCC) | 60 WSI | 40 WSI | Jaccard score 0.797 |
| Feng S. et al. (2021) [15] | China | segmentation (HCC) | 592 WSI | 157 WSI (TCGA) | Accuracy 0.88 |
| Yang TL. et al. (2022) [16] | Taiwan | segmentation (HCC) | 46 WSI | no | Jaccard score 0.89 |
| Li S. et al. (2017) [17] | China | segmentation (HCC) | 127 WSI | not provided | Accuracy 0.97 F1 score 0.94 |
| Kiani A. et al. (2020) [18] | USA | classification (HCC and CCK) | 70 WSI | 80 WSI | Accuracy 0.84 |
| Schau GF. et al. (2020) [19] | USA | classification (LM) | 257 WSI | no | F1 score 0.77 |
| Ercan C. et al. (2022) [20] | Switzerland | classification (HCC) | 98 WSI | no | Accuracy 0.84 F1 score 0.91 |
| Diao S. et al. (2022) [21] | China | classification (HCC) | 100 WSI (TCGA) | no | AUC 0.92 |
| Cheng N. et al. (2022) [22] | China | classification (HCC) | 649 WSI | 234 WSI | AUC 0.94 |
| Chen M. et al. (2020) [23] | China | prediction (HCC) | 387 WSI | 101 WSI | AUC 0.71 |
| Liao H. et al. (2020) [24] | China | prediction (HCC) | not provided | not provided | AUC 0.89 |
| Saillard C. et al. (2020) [25] | France | prediction (HCC) | 390 WSI | 342 WSI (TCGA) | c-index 0.70 |
| Yamashita et al. (2021) [26] | USA | prediction (HCC) | 299 WSI (TCGA) | 198 WSI | c-index 0.67 |
| Saito A. et al. (2021) [27] | Japan | prediction (HCC) | 158 WSI | no | Accuracy 0.90 |
| Xiao C. et al. (2022) [28] | China | prediction (LM) | 611 WSI | no | AUC 0.85 |
| Chen Q. et al. (2022) [29] | China | prediction (HCC) | 2917 WSI | 504 WSI | AUC 0.87 |
| Zeng Q. et al. (2022) [30] | France | prediction (HCC) | 349 WSI (TCGA) | 139 WSI | AUC 0.92 |
| Qu WF. et al. (2022) [31] | China | prediction (HCC) | 576 WSI | 147 WSI (TCGA) | c-index 0.71 |
| Xie J. et al. (2022) [32] | China | prediction (CCK) | 766 WSI | no | AUC 0.68 |

HCC: Hepatocellular Carcinoma; CCK: Cholangiocarcinoma; LM: Liver Metastasis.

3.3. Non-Tumoral

Twenty articles were reviewed through full text screening (Table 3) [34–53]. As for tumoral studies, none used prospectively collected data. The algorithms primary goal was either segmentation (4/20) [34–37], classification (15/20) [38–52] or prediction (1/20) [53]. The most prevalent thematic included inflammation detection and quantification, especially in Non-Alcoholic SteatoHepatitis (NASH) (6/20) and fibrosis detection and grading (4/20). Two studies reported on the evaluation of steatosis to evaluate the quality of the graft in liver transplantation [36,48]. All studies relied on HE staining, with or without additional immunohistochemical stains, except for one, which relied only on CK7 immunohistochemistry [50]. No study validated their algorithm on an external validation dataset, be it designed for either segmentation, classification or prediction. The DNN models used in each study are reported in Supplementary Material Table S2.

The number of WSI incorporated into the developmental set was disparate among studies. In particular, Puri developed a machine learning model on 1277 digital slides to detect drug-induced liver injuries on the cellular level, based on both rat and human hepatocytes, achieving 0.99 Accuracy [43]. On the other end of the spectrum, studies such as Gawrieh et al. and Perez Sans et al. showed the feasibility of developing a steatosis classifier on a training dataset as low as 18 and 20 WSI, respectively [47,48].

NASH scoring is a combination of multiple cardinal features, namely steatosis, lobular inflammation, hepatocyte ballooning and fibrosis. Attempts to develop automated DNN-based classification algorithms for NASH scoring occurred as early as 2015 by Vanderbeck et al., whose model achieved an AUC of 0.95 for lobular inflammation and 0.98 for hepatocyte ballooning detection [39]. Other teams used machine learning programs to help discriminate between NAFLD and NASH [45,46], chronic hepatitis and NASH [49] or alcoholic steatohepatitis and NASH [51]. Heinemann et al. proposed an algorithm not only capable of identifying the cardinal features of NASH but also able to score NASH as a discrete pathologist-like score (as per the Kleiner score of NASH activity score, NAS) [52]. Inter-observer agreement between the DNN algorithm and pathologists was best for steatosis and fibrosis class assessment (kappa 0.92 and 0.81, respectively), but dropped for ballooning and inflammation class assessment (kappa 0.42 and 0.40, respectively).

Constantinescu et al. were the sole proposer of a predictive deep-learning based model to predict early (within 30 days) post-surgical complications after liver surgery [53]. To that end, their model assessed neoangiogenesis and inflammation through both HE and immunohistochemical stains and was able to predict the arising of complications with an accuracy of 0.97 and an AUC of 0.97 on the internal testing dataset.

Table 3. Key features of reported studies regarding the use of Deep Neural Networks (DNN) on Whole Image histology Slides (WSI) in the “non tumoral” field.

| Authors (Year) [Reference] | Country | Algorithm Goal | Development Dataset | External Validation Dataset | Performance Metrics |
|-------------------------------|----------------|---------------------------------------|---------------------|-----------------------------|-------------------------------------|
| Guo X. et al. (2018) [34] | USA | segmentation (steatosis) | 451 WSI | no | Jaccard score 0.77 F1 score 0.66 |
| Jirik M. et al. (2020) [35] | Czech Republic | segmentation (intra vs. extralobular) | 33 WSI | no | Accuracy 0.91 |
| Roy M. et al. (2020) [36] | USA | segmentation (steatosis) | 36 WSI | no | F1 score 0.94 |
| Yu H. et al. (2022) [37] | USA | segmentation (portal tracts) | 53 WSI | no | Jaccard score 0.80 F1 score 0.89 |

Table 3. Cont.

| Authors (Year) [Reference] | Country | Algorithm Goal | Development Dataset | External Validation Dataset | Performance Metrics |
|--------------------------------------|-------------|---|---------------------|-----------------------------|---|
| Vanderbeck S. et al. (2014) [38] | USA | classification (steatosis, bile ducts, vascular structures) | 47 WSI | no | AUC 0.83 |
| Vanderbeck S. et al. (2015) [39] | USA | classification (NASH) | 59 WSI | no | AUC 0.98 |
| Wang TH et al. (2015) [40] | Taiwan | classification (fibrosis) | 175 WSI | no | AUC 0.82 |
| Munsterman I. et al. (2019) [41] | Netherlands | classification (NASH) | 79 WSI | no | AUC 0.97 |
| Klimov S. et al. (2019) [42] | USA | classification (fibrosis) | 115 WSI | no | AUC 0.70 |
| Puri M. (2020) [43] | USA | classification (DILI) | 1277 WSI | no | Accuracy 0.99 |
| Forlano et al. (2020) [44] | UK | classification (NASH) | 246 WSI | no | AUC 0.80 |
| Teramoto T. et al. (2020) [45] | Japan | classification (NASH) | 79 WSI | no | AUC 0.85 |
| Salvi M. et al. (2020) [46] | Italy | classification (steatosis) | 385 WSI | no | Accuracy 0.97 |
| Gawrieh S. et al. (2020) [47] | USA | classification (NASH) | 18 WSI | no | AUC 0.79 |
| Perez-Sans F. et al. (2021) [48] | Spain | classification (steatosis) | 20 WSI | no | AUC 0.98 |
| Marti-Aguado D. et al. (2021) [49] | Spain | Classification (chronic hepatitis) | 156 WSI | no | AUC 0.75 (NASH) AUC 0.99 (Chronic Hepatitis model) |
| Sjöblom N. et al. (2021) [50] | Finland | classification (chronic cholestasis) | 210 WSI | no | Accuracy 0.93 |
| Ramkissoon R. et al. (2022) [51] | USA | classification (NASH) | 97 WSI | no | AUC 0.96 |
| Heinemann F. et al. (2022) [52] | USA | classification (NASH) | 467 WSI | no | F1 score 0.37 to 0.85 |
| Constantinescu C. et al. (2022) [53] | Romania | prediction (liver surgery complications) | 500 WSI | no | AUC 0.97 |

NASH: Non-Alcoholic Steatohepatitis; NAFLD: Non-Alcoholic Fatty Liver Disease; DILI: Drug Induced Liver Injuries.

3.4. Assessment of the Risk of Bias and Applicability through the QUADAS-2 Tool

The overall assessment of the reviewed studies is presented in Figure 3. When the overall risk of bias was measured with the QUADAS-2 tool, a majority of studies (30/42) presented at least one high-risk factor. Algorithms developed on unicentric datasets inferior to 60 WSI were deemed at high risk for the patient selection domain. All studies without an external validation set were judged at high risk for the “index test domain”. For the “reference standard” domain, high risk was applied if “ground truth” was determined by only one pathologist and on criteria deemed either too subjective or that were notoriously low-reproducible. Finally, as no study fitted the criteria for evaluation by the “flow and

timing” domain, it was categorized as “not applicable”. Per study evaluation with the QUADAS-2 tool is provided in Table S2.

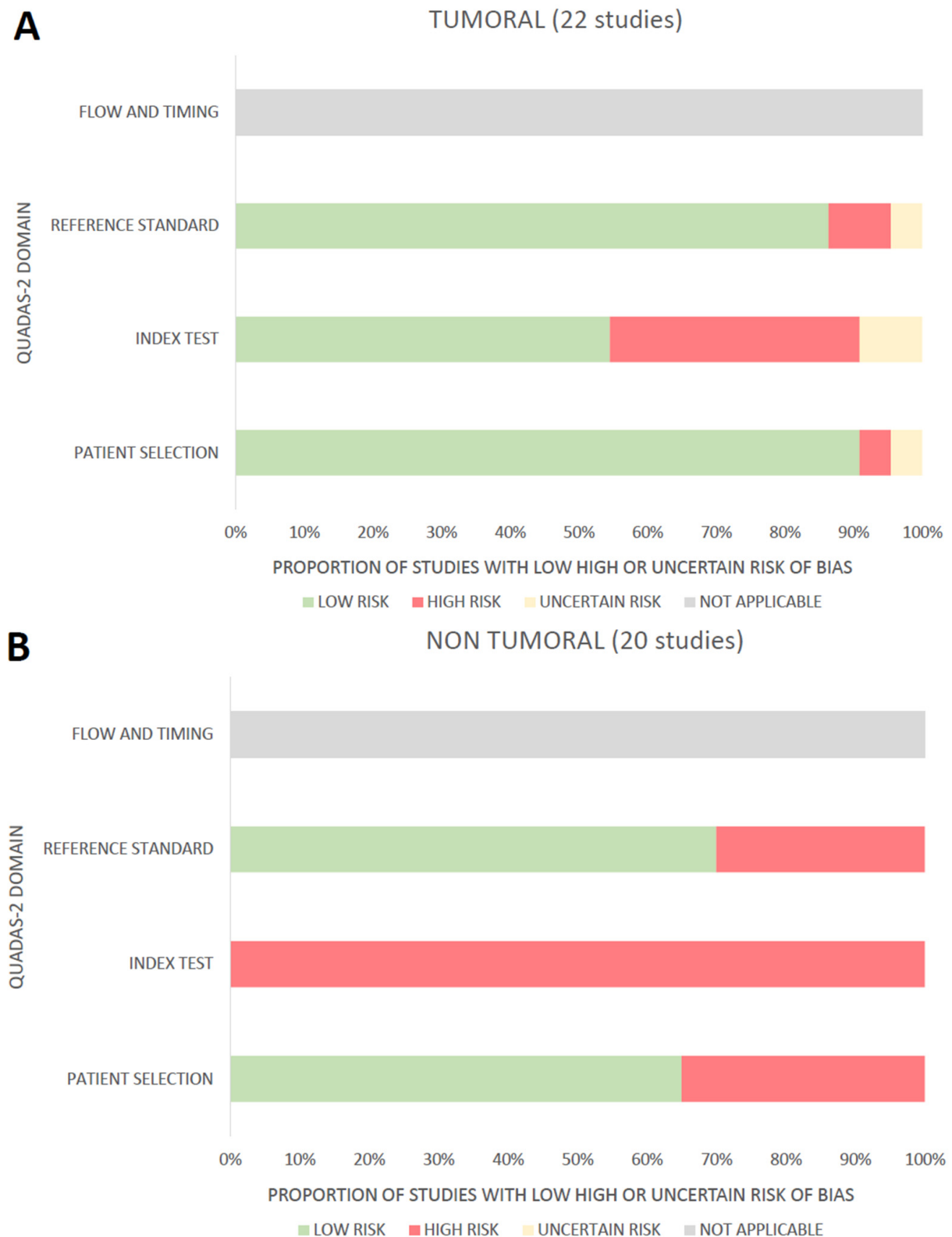


Figure 3. Global overview of reviewed studies, including evaluation of risk of bias through the QUADAS-2 tool for the “tumoral” (A) and “non tumoral” (B) fields.

4. Discussion

As shown, the applications and opportunities offered by AI are manifold, even for a field as specialized as liver pathology, in terms of both the tumoral (hepatocellular carcinoma, cholangiocarcinoma and secondary tumors in liver) and non-tumoral (NAFLD and NASH, chronic hepatitis, fibrosis staging, iatrogenic-induced lesions) aspects. DNN-based models are such promising tools that they have led to numerous publications throughout the last decade, with an increasing rate over the past few years. We have strived to provide a comprehensive, systematic and updated review of AI in liver pathology. Pathologists may sometimes feel estranged regarding this topic, because AI is foremost a mathematical and computational discipline, with a wide array of performance metrics, architectures and other specificities. Hence, navigating the literature might sometimes prove challenging. Previous reviews have discussed the opportunities of AI in liver diseases as a whole, or as part of clinical, radiological and sometimes pathological applications in gastroenterology and hepatology [2–8]. To our knowledge, this review is the first to solely focus on DNN-based applications in liver pathology, and to evaluate their bias through the lens of the QUADAS2 tool.

4.1. AI in Tumoral Liver Pathology: What to Remember

The practical value of segmentation algorithms in the tumoral liver pathology field may seem poor, because most of the time pathologists have no trouble identifying HCC areas in the liver. In a similar fashion, classification algorithms such as Kiani et al.'s [18], which aimed to distinguish HCC and CCK, may seem of limited value, because there are already robust immunohistochemical tools available. However, the major contribution of such algorithms might be a gaining of time in everyday practice, and one might argue that this alone is pertinent enough in the context of a shortage in liver pathologists, as has been raised by several authors in other domains of pathology [54–57]. The development of classification algorithms targeting lesions of difficult diagnosis such as benign hepatocellular proliferations (hepatocellular adenoma and focal nodular hyperplasia) might prove the most beneficial, especially on biopsy specimens, such as in Chens et al.'s study [22].

The development of DNN-based algorithms tackling the tumoral field in liver pathology seem to be more advanced and robust throughout the literature than those in the non-tumoral field. Indeed, our global evaluation through the QUADAS-2 tool showed that studies in the tumoral field were less prone to risk of bias in all domains. The reasons for this include the availability of public online databases such as The Cancer Genome Atlas (TCGA), which compiles molecular data and whole slide images of numerous cancers, which are an invaluable resource for the development and external validation of many DNN-based algorithms for pathology, including hepatocellular carcinoma [15,26,32]. Likewise, multicentric studies with a development cohort in one hospital and an external testing dataset in another may be easier to design for DNN-based studies in the tumoral rather than the metabolic field [22,24,30].

The most promising uses for DNN-based algorithms in tumoral hepatology seem to lie in prediction algorithms. Predicting recurrence [23] or survival [26] from histological WSI is especially appealing, because its clinical applications in patient's care are obvious and, as of today, out of reach. The prediction of gene signature without resorting to molecular testing [31] could also prove a gain of both time and money. As of yet, there is no targeted therapy in HCC. The prediction of genomic alterations in CCK might, on the other hand, yield therapeutic value [58,59]. One must keep in mind that the predictive information of survival or recurrence is complex and most certainly multimodal. As such, DNN-based algorithms could provide yet another tool for patient management, in association with existing pathological (TNM classification), clinical (Performance Status) or radiology scoring systems [60,61]. Hence, multimodal approaches such as DNN-based systems compiling radiological and pathological imaging data (reflected through the concepts of “radiomics” and “pathomics”, respectively) have been proposed for non-small lung cancer [62], cervical cancer [63], breast cancer [64] and glioblastoma [65] but have yet to

be developed in liver pathology. Radiological imaging data are particularly suited to the development and use of deep learning artificial intelligence and share many of its concepts and limitations with the field of pathology [66].

4.2. AI in Non-Tumoral Liver Pathology: What to Remember

On the other hand, classification DNN-based systems are of paramount interest in the metabolic liver pathology field. For example, steatohepatitis diagnosis and grading has been shown to present poor inter-observer reproducibility [67]. In particular, hepatocyte ballooning recognition seems to be poorly reproducible, even between expert pathologists, hindering DNN development on the subject, as acknowledged by authors such as Brunt et al. [68]. DNN-based models may prove useful in the future to standardize such difficult diagnosis, yet a “reference standard bias” may represent a pitfall in their development if overlooked. To avoid this, “ground truth” (i.e., the annotations) should be defined by standard scoring systems such as the Non-Alcoholic Fatty Liver Disease Activity Score for NAFLD and by a consensus of more than one pathologist, and this could be built once again on a multimodal level including clinical, biological and radiological data.

While external validation is feasible enough for studies on tumoral material (13/22 in this review), it seems an acute limitation for those on inflammatory or metabolic diseases. In fact, among the papers we reviewed, none (0/20) of them tested their model on an external validation set. There are in fact, at the time of this article’s writing, no online database similar to TCGA for metabolic or inflammatory liver diseases. Setting up such a public annotated database might prove beneficial to put DNN-based algorithms to trial.

4.3. Persistent Limitations to the Implementation of AI in Daily Practice

AI integration into healthcare increases at an ever-growing speed, and DNN-based models seem to be promising tools in the image analysis field as a whole, in radiology and, since the arising of whole slide imaging, in pathology. At present, image-analysis DNN-tools do not outmatch trained physicians, but they might prove useful for alleviating complex and time-consuming tasks [69]. As great as their promises may be, DNN systems are hampered by some limitations. Most of the studies reported in this review were stratified as having at least one domain of high risk of bias according to the QUADAS-2 tool. Some steps should be taken when designing a DNN-based study, among them (but not limited to) building a multicentric, sufficiently large development database with a “ground truth” defined by consensual, if possible multimodal, criteria and systematically validated on an external dataset. However, economic and practical constraints should be taken into consideration, for most DNN models require a curated and sufficiently large developmental dataset, which can prove time-consuming and costly [70]. The explainability of predictive algorithms is also a common shortcoming that some authors have begun to tackle by looking back at the most predictive histological tiles of their models [25,29,30]. To further that end, the role of pathologists will be of paramount importance. Ethical questions such as minorities and low-income country integration should also always be kept in mind [71], because AI’s usefulness relies on its success in a clinical setting, and its scope is always increasing [72,73]. As such, predictive risk-stratification models in particular may be of the utmost interest.

4.4. Present Review’s Limitations

We focused on two databases, namely Pubmed and Embase, and may have missed papers published in specialized, non-healthcare-related journals. It is also to be noted that the QUADAS-2 tool, though encouraged by the PRISMA guidelines, is not perfectly suited for the qualitative assessment of AI-related studies. Indeed, the QUADAS-2 guidelines do not provide precise criteria to assess the robustness of AI-related studies. Our criteria stratifying “low risk of bias” and “high risk of bias” are therefore subjective. QUADAS-IA is an extension of QUADAS-2 and is specifically designed to tackle this issue, and to date is still to be published [74]. When available, it should prove an even more precise tool

to evaluate diagnostic research in the AI field. Cabitza and Campagner also proposed a checklist of 30 items for the qualitative assessment of medical AI studies [75]. While not specifically designed for pathology, it may be useful when designing an AI-based study in healthcare; it is available online and was additionally performed in the present review for each study (Supplementary Material Table S3) [76].

5. Conclusions

For the moment, contrary to other fields of pathology, there are no commercially available AI-based solutions in liver pathology. Studies are still impaired by a lack of external validation cohorts. Therefore, multicentric studies and expert collaborations are needed to further develop DNN-based algorithms.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/diagnostics13101799/s1>, Table S1: Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) statements as reported in the present study. Table S2: Specific models used in each study. If multiple models were tested, the model achieving the best results is marked by an asterisk (*). Table S3: Per-study review according to the IJMEDI checklist. Table S4: Individual review evaluation of risks of bias through the QUADAS-2 tool.

Author Contributions: Conceptualization, P.A. and S.-F.K.-J.; methodology, P.A., N.R. and T.P.; validation, S.-F.K.-J., E.B.-J., O.L. and N.R.-L.; writing—original draft preparation, P.A.; writing—review and editing, P.A., N.R., B.T., E.B.-J., J.C., Z.-E.K., O.A., R.D.C., N.R.-L., T.P. and S.-F.K.-J.; supervision, S.-F.K.-J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available in the Supplementary Materials.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Su, T.H.; Wu, C.H.; Kao, J.H. Artificial intelligence in precision medicine in hepatology. *J. Gastroenterol. Hepatol.* **2021**, *36*, 569–580. [CrossRef]
2. Balsano, C.; Alisi, A.; Brunetto, M.R.; Invernizzi, P.; Burra, P.; Piscaglia, F. The application of artificial intelligence in hepatology: A systematic review. *Dig. Liver Dis.* **2022**, *54*, 299–308. [CrossRef]
3. Calderaro, J.; Kather, J.N. Artificial intelligence-based pathology for gastrointestinal and hepatobiliary cancers. *Gut* **2021**, *70*, 1183–1193. [CrossRef]
4. Lee, H.W.; Sung, J.J.Y.; Ahn, S.H. Artificial intelligence in liver disease. *J. Gastroenterol. Hepatol.* **2021**, *36*, 539–542. [CrossRef] [PubMed]
5. Le Berre, C.; Sandborn, W.J.; Aridhi, S.; Devignes, M.-D.; Fournier, L.; Smaïl-Tabbone, M.; Danese, S.; Peyrin-Biroulet, L. Application of Artificial Intelligence to Gastroenterology and Hepatology. *Gastroenterology* **2020**, *158*, 76–94.e2. [CrossRef] [PubMed]
6. Christou, C.D.; Tsoulfas, G. Challenges and opportunities in the application of artificial intelligence in gastroenterology and hepatology. *World J. Gastroenterol.* **2021**, *27*, 6191–6223. [CrossRef] [PubMed]
7. Cao, J.-S.; Lu, Z.-Y.; Chen, M.-Y.; Zhang, B.; Juengpanich, S.; Hu, J.-H.; Li, S.-J.; Topatana, W.; Zhou, X.-Y.; Feng, X.; et al. Artificial intelligence in gastroenterology and hepatology: Status and challenges. *World J. Gastroenterol.* **2021**, *27*, 1664–1690. [CrossRef]
8. Ahn, J.C.; Connell, A.; Simonetto, D.A.; Hughes, C.; Shah, V.H. Application of Artificial Intelligence for the Diagnosis and Treatment of Liver Diseases. *Hepatology* **2021**, *73*, 2546–2563. [CrossRef]
9. Liberati, A.; Altman, D.G.; Tetzlaff, J.; Mulrow, C.; Gøtzsche, P.C.; Ioannidis, J.P.A.; Clarke, M.; Devereaux, P.J.; Kleijnen, J.; Moher, D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ* **2009**, *339*, b2700. [CrossRef]
10. Bristol U of. QUADAS-2. University of Bristol. Available online: <https://www.bristol.ac.uk/population-health-sciences/projects/quadas/quadas-2/> (accessed on 29 December 2022).

11. Feng, Y.; Hafiane, A.; Laurent, H. A deep learning based multiscale approach to segment the areas of interest in whole slide images. *Comput. Med. Imaging Graph.* **2021**, *90*, 101923. [\[CrossRef\]](#)
12. Cancian, P.; Cortese, N.; Donadon, M.; Di Maio, M.; Soldani, C.; Marchesi, F.; Savevski, V.; Santambrogio, M.D.; Cerina, L.; Laino, M.E.; et al. Development of a Deep-Learning Pipeline to Recognize and Characterize Macrophages in Colo-Rectal Liver Metastasis. *Cancers* **2021**, *13*, 3313. [\[CrossRef\]](#)
13. Roy, M.; Kong, J.; Kashyap, S.; Pastore, V.P.; Wang, F.; Wong, K.C.L.; Mukherjee, V. Convolutional autoencoder based model HistoCAE for segmentation of viable tumor regions in liver whole-slide images. *Sci. Rep.* **2021**, *11*, 139. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Wang, X.; Fang, Y.; Yang, S.; Zhu, D.; Wang, M.; Zhang, J.; Tong, K.; Han, X. A hybrid network for automatic hepatocellular carcinoma segmentation in H&E-stained whole slide images. *Med. Image Anal.* **2021**, *68*, 101914. [\[PubMed\]](#)
15. Feng, S.; Yu, X.; Liang, W.; Li, X.; Zhong, W.; Hu, W.; Zhang, H.; Feng, Z.; Song, M.; Zhang, J.; et al. Development of a Deep Learning Model to Assist With Diagnosis of Hepatocellular Carcinoma. *Front. Oncol.* **2021**, *11*, 762733. [\[CrossRef\]](#)
16. Yang, T.-L.; Tsai, H.-W.; Huang, W.-C.; Lin, J.-C.; Liao, J.-B.; Chow, N.-H.; Chung, P.-C. Pathologic liver tumor detection using feature aligned multi-scale convolutional network. *Artif. Intell. Med.* **2022**, *125*, 102244. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Li, S.; Jiang, H.; Pang, W. Joint multiple fully connected convolutional neural network with extreme learning machine for hepatocellular carcinoma nuclei grading. *Comput. Biol. Med.* **2017**, *84*, 156–167. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Kiani, A.; Uyumazturk, B.; Rajpurkar, P.; Wang, A.; Gao, R.; Jones, E.; Yu, Y.; Langlotz, C.P.; Ball, R.L.; Montine, T.J.; et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. *NPJ Digit. Med.* **2020**, *3*, 23. [\[CrossRef\]](#)
19. Schau, G.F.; Burlingame, E.A.; Thibault, G.; Anekpuranang, T.; Wang, Y.; Gray, J.W.; Corless, C.; Chang, Y.H. Predicting primary site of secondary liver cancer with a neural estimator of metastatic origin. *J. Med. Imaging* **2020**, *7*, 012706. [\[CrossRef\]](#)
20. Ercan, C.; Coto-Llerena, M.; Piscuoglio, S.; Terracciano, L.M. Establishing quantitative image analysis methods for tumor microenvironment evaluation. *J. Hepatol.* **2022**, *77*, S660–S661. [\[CrossRef\]](#)
21. Diao, S.; Tian, Y.; Hu, W.; Hou, J.; Lambo, R.; Zhang, Z.; Xie, Y.; Nie, X.; Zhang, F.; Racoceanu, D.; 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.* **2022**, *192*, 553–563. [\[CrossRef\]](#)
22. Cheng, N.; Ren, Y.; Zhou, J.; Zhang, Y.; Wang, D.; Zhang, X.; Chen, B.; Liu, F.; Lv, J.; Cao, Q.; et al. Deep Learning-Based Classification of Hepatocellular Nodular Lesions on Whole-Slide Histopathologic Images. *Gastroenterology* **2022**, *162*, 1948–1961.e7. [\[CrossRef\]](#)
23. Chen, M.; Zhang, B.; Topatana, W.; Cao, J.; Zhu, H.; Juengpanich, S.; Mao, Q.; Yu, H.; Cai, X. Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning. *NPJ Precis. Oncol.* **2020**, *4*, 14. [\[PubMed\]](#)
24. Liao, H.; Xiong, T.; Peng, J.; Xu, L.; Liao, M.; Zhang, Z.; Wu, Z.; Yuan, K.; Zeng, Y. Classification and Prognosis Prediction from Histopathological Images of Hepatocellular Carcinoma by a Fully Automated Pipeline Based on Machine Learning. *Ann. Surg. Oncol.* **2020**, *27*, 2359–2369. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Saillard, C.; Schmauch, B.; Laifa, O.; Moarii, M.; Toldo, S.; Zaslavskiy, M.; Pronier, E.; Laurent, A.; Amaddeo, G.; Regnault, H.; et al. Predicting Survival After Hepatocellular Carcinoma Resection Using Deep Learning on Histological Slides. *Hepatology* **2020**, *72*, 2000–2013. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Yamashita, R.; Long, J.; Saleem, A.; Rubin, D.L.; Shen, J. Deep learning predicts postsurgical recurrence of hepatocellular carcinoma from digital histopathologic images. *Sci. Rep.* **2021**, *11*, 2047. [\[CrossRef\]](#)
27. Saito, A.; Toyoda, H.; Kobayashi, M.; Koiwa, Y.; Fujii, H.; Fujita, K.; Maeda, A.; Kaneoka, Y.; Hazama, S.; Nagano, H.; et al. Prediction of early recurrence of hepatocellular carcinoma after resection using digital pathology images assessed by machine learning. *Mod. Pathol.* **2021**, *34*, 417–425. [\[CrossRef\]](#)
28. Xiao, C.; Zhou, M.; Yang, X.; Wang, H.; Tang, Z.; Zhou, Z.; Tian, Z.; Liu, Q.; Li, X.; Jiang, W.; et al. Accurate Prediction of Metachronous Liver Metastasis in Stage I–III Colorectal Cancer Patients Using Deep Learning with Digital Pathological Images. *Front. Oncol.* **2022**, *12*, 67. [\[CrossRef\]](#)
29. Chen, Q.; Xiao, H.; Gu, Y.; Weng, Z.; Wei, L.; Li, B.; Liao, B.; Li, J.; Lin, J.; Hei, M.; et al. Deep learning for evaluation of microvascular invasion in hepatocellular carcinoma from tumor areas of histology images. *Hepatol. Int.* **2022**, *16*, 590–602. [\[CrossRef\]](#)
30. Zeng, Q.; Klein, C.; Caruso, S.; Maille, P.; Laleh, N.G.; Sommacale, D.; Laurent, A.; Amaddeo, G.; Gentien, D.; Rapinat, A.; et al. Artificial intelligence predicts immune and inflammatory gene signatures directly from hepatocellular carcinoma histology. *J. Hepatol.* **2022**, *77*, 116–127. [\[CrossRef\]](#)
31. Qu, W.-F.; Tian, M.-X.; Qiu, J.-T.; Guo, Y.-C.; Tao, C.-Y.; Liu, W.-R.; Tang, Z.; Qian, K.; Wang, Z.-X.; Li, X.-Y.; et al. Exploring pathological signatures for predicting the recurrence of early-stage hepatocellular carcinoma based on deep learning. *Front. Oncol.* **2022**, *12*, 968202. [\[CrossRef\]](#)
32. Xie, J.; Pu, X.; He, J.; Qiu, Y.; Lu, C.; Gao, W.; Wang, X.; Lu, H.; Shi, J.; Xu, Y.; et al. Survival prediction on intrahepatic cholangiocarcinoma with histomorphological analysis on the whole slide images. *Comput. Biol. Med.* **2022**, *146*, 105520. [\[CrossRef\]](#) [\[PubMed\]](#)
33. PAIP 2019—Grand Challenge. Available online: <https://paip2019.grand-challenge.org/> (accessed on 28 December 2022).
34. Guo, X.; Wang, F.; Teodoro, G.; Farris, A.B.; Kong, J. Liver steatosis segmentation with deep learning methods. *Proc. IEEE Int. Symp. Biomed Imaging* **2019**, *2019*, 24–27. [\[PubMed\]](#)

35. Jirik, M.; Gruber, I.; Moulisova, V.; Schindler, C.; Cervenková, L.; Palek, R.; Rosendorf, J.; Arlt, J.; Bolek, L.; Dejmek, J.; et al. Semantic Segmentation of Intralobular and Extralobular Tissue from Liver Scaffold H&E Images. *Sensors* **2020**, *20*, 7063.
36. Roy, M.; Wang, F.; Vo, H.; Teng, D.; Teodoro, G.; Farris, A.B.; Castillo-Leon, E.; Vos, M.B.; Kong, J. Deep-learning-based accurate hepatic steatosis quantification for histological assessment of liver biopsies. *Lab. Investig.* **2020**, *100*, 1367–1383. [[CrossRef](#)] [[PubMed](#)]
37. Yu, H.; Sharifai, N.; Jiang, K.; Wang, F.; Teodoro, G.; Farris, A.B.; Kong, J. Artificial intelligence based liver portal tract region identification and quantification with transplant biopsy whole-slide images. *Comput. Biol. Med.* **2022**, *150*, 106089. [[CrossRef](#)] [[PubMed](#)]
38. Vanderbeck, S.; Bockhorst, J.; Komorowski, R.; Kleiner, D.E.; Gawrieh, S. Automatic classification of white regions in liver biopsies by supervised machine learning. *Hum. Pathol.* **2014**, *45*, 785–792. [[CrossRef](#)]
39. Vanderbeck, S.; Bockhorst, J.; Kleiner, D.; Komorowski, R.; Chalasani, N.; Gawrieh, S. Automatic quantification of lobular inflammation and hepatocyte ballooning in nonalcoholic fatty liver disease liver biopsies. *Hum. Pathol.* **2015**, *46*, 767–775. [[CrossRef](#)]
40. Wang, T.H.; Chen, T.C.; Teng, X.; Liang, K.H.; Yeh, C.T. Automated biphasic morphological assessment of hepatitis B-related liver fibrosis using second harmonic generation microscopy. *Sci. Rep.* **2015**, *5*, 12962. [[CrossRef](#)]
41. Munsterman, I.D.; van Erp, M.; Weijers, G.; Bronkhorst, C.; de Korte, C.L.; Drenth, J.P.H.; van der Laak, J.A.W.M.; Tjwa, E.T.T.L. A Novel Automatic Digital Algorithm that Accurately Quantifies Steatosis in NAFLD on Histopathological Whole-Slide Images. *Cytometry B Clin. Cytom.* **2019**, *96*, 521–528. [[CrossRef](#)]
42. Klimov, S.; Farris, A.; Chen, H.; Chen, Y.; Jiang, Y. THU-083-Predicting advanced liver fibrosis using deep learning based biopsy image analysis. *J. Hepatol.* **2019**, *70*, e196. [[CrossRef](#)]
43. Puri, M. Automated Machine Learning Diagnostic Support System as a Computational Biomarker for Detecting Drug-Induced Liver Injury Patterns in Whole Slide Liver Pathology Images. *Assay Drug Dev. Technol.* **2020**, *18*, 1–10. [[CrossRef](#)] [[PubMed](#)]
44. Forlano, R.; Mullish, B.H.; Giannakeas, N.; Maurice, J.B.; Angkathunyakul, N.; Lloyd, J.; Tzallas, A.T.; Tsipouras, M.; Yee, M.; Thursz, M.R.; et al. High-Throughput, Machine Learning-Based Quantification of Steatosis, Inflammation, Ballooning, and Fibrosis in Biopsies from Patients with Nonalcoholic Fatty Liver Disease. *Clin. Gastroenterol. Hepatol.* **2020**, *18*, 2081–2090.e9. [[CrossRef](#)] [[PubMed](#)]
45. Teramoto, T.; Shinohara, T.; Takiyama, A. Computer-aided classification of hepatocellular ballooning in liver biopsies from patients with NASH using persistent homology. *Comput. Methods Programs Biomed.* **2020**, *195*, 105614. [[CrossRef](#)] [[PubMed](#)]
46. Salvi, M.; Molinaro, L.; Metovic, J.; Patrono, D.; Romagnoli, R.; Papotti, M.; Molinari, F. Fully automated quantitative assessment of hepatic steatosis in liver transplants. *Comput. Biol. Med.* **2020**, *123*, 103836. [[CrossRef](#)] [[PubMed](#)]
47. Gawrieh, S.; Sethunath, D.; Cummings, O.W.; Kleiner, D.E.; Vuppalanchi, R.; Chalasani, N.; Tuceryan, M. Automated quantification architectural pattern detection of hepatic fibrosis in, N.A.F.L.D. *Ann. Diagn. Pathol.* **2020**, *47*, 151518. [[CrossRef](#)]
48. Pérez-Sanz, F.; Riquelme-Pérez, M.; Martínez-Barba, E.; de la Peña-Moral, J.; Salazar Nicolás, A.; Carpes-Ruiz, M.; Esteban-Gil, A.; Legaz-García, M.D.C.; Parreño-González, M.A.; Ramírez, P.; et al. Efficiency of Machine Learning Algorithms for the Determination of Macrovesicular Steatosis in Frozen Sections Stained with Sudan to Evaluate the Quality of the Graft in Liver Transplantation. *Sensors* **2021**, *21*, 1993. [[CrossRef](#)]
49. Marti-Aguado, D.; Fernández-Patón, M.; Alfaro-Cervello, C.; Mestre-Alagarda, C.; Bauza, M.; Gallen-Peris, A.; Merino, V.; Benlloch, S.; Pérez-Rojas, J.; Ferrández, A.; et al. Digital Pathology Enables Automated and Quantitative Assessment of Inflammatory Activity in Patients with Chronic Liver Disease. *Biomolecules* **2021**, *11*, 1808. [[CrossRef](#)]
50. Sjöblom, N.; Boyd, S.; Manninen, A.; Knuuttila, A.; Blom, S.; Färkkilä, M.; Arola, J. Chronic cholestasis detection by a novel tool: Automated analysis of cytokeratin 7-stained liver specimens. *Diagn. Pathol.* **2021**, *16*, 41. [[CrossRef](#)]
51. Ramkissoon, R.; Ahn, J.; Noh, Y.-K.; Yeon, Y.; Meroueh, C.; Thanneer, P.; Das, A.; Hipp, J.; Allen, A.; Simonetto, D.; et al. Application of machine learning algorithms to classify steatohepatitis on liver biopsy. *J. Hepatol.* **2022**, *77*, S138–S139. [[CrossRef](#)]
52. Heinemann, F.; Birk, G.; Stierstorfer, B. Deep learning enables pathologist-like scoring of NASH models. *Sci. Rep.* **2019**, *9*, 18454. [[CrossRef](#)]
53. Constantinescu, C.; Udristoiu, A.; Gruionu, L.G.; Pirici, D.; Gruionu, G.; Iacob, A.V.; Udristoiu, S.; Surlin, V.; Ramboiu, S.; Saftoiu, A. Tu1315: Deep learning assessment of inflammation and angiogenesis in liver steatosis as an accurate predictor of liver surgery complications. *Gastroenterology* **2022**, *162*, S1272. [[CrossRef](#)]
54. Song, Z.; Zou, S.; Zhou, W.; Huang, Y.; Shao, L.; Yuan, J.; Gou, X.; Jin, W.; Wang, Z.; Chen, X.; et al. Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning. *Nat. Commun.* **2020**, *11*, 4294. [[CrossRef](#)] [[PubMed](#)]
55. Sakamoto, T.; Furukawa, T.; Lami, K.; Pham, H.H.N.; Uegami, W.; Kuroda, K.; Kawai, M.; Sakanashi, H.; Cooper, L.A.D.; Bychkov, A.; et al. A narrative review of digital pathology and artificial intelligence: Focusing on lung cancer. *Transl. Lung Cancer Res.* **2020**, *9*, 2255–2276. [[CrossRef](#)] [[PubMed](#)]
56. De Logu, F.; Ugolini, F.; Maio, V.; Simi, S.; Cossu, A.; Massi, D.; Italian Association for Cancer Research (AIRC) Study Group; Nassini, R.; Laurino, M. Recognition of Cutaneous Melanoma on Digitized Histopathological Slides via Artificial Intelligence Algorithm. *Front. Oncol.* **2020**, *10*, 1559. [[CrossRef](#)] [[PubMed](#)]

57. Ström, P.; Kartasalo, K.; Olsson, H.; Solorzano, L.; Delahunt, B.; Berney, D.M.; Bostwick, D.G.; Evans, A.J.; Grignon, D.J.; Humphrey, P.A.; et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: A population-based, diagnostic study. *Lancet Oncol.* **2020**, *21*, 222–232. [CrossRef]
58. Massironi, S.; Pilla, L.; Elvevi, A.; Longarini, R.; Rossi, R.E.; Bidoli, P.; Invernizzi, P. New and Emerging Systemic Therapeutic Options for Advanced Cholangiocarcinoma. *Cells* **2020**, *9*, 688. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7140695/> (accessed on 16 February 2023). [CrossRef] [PubMed]
59. Kabbara, K.W.; Cannon, T.; Winer, A.; Wadlow, R.C. Molecular Pathogenesis of Cholangiocarcinoma: Implications for Disease Classification and Therapy. *Oncology* **2022**, *36*, 492–498.
60. Park, H.J.; Park, B.; Lee, S.S. Radiomics and Deep Learning: Hepatic Applications. *Korean J. Radiol.* **2020**, *21*, 387–401. [CrossRef]
61. Wei, J.; Jiang, H.; Gu, D.; Niu, M.; Fu, F.; Han, Y.; Song, B.; Tian, J. Radiomics in liver diseases: Current progress and future opportunities. *Liver Int.* **2020**, *40*, 2050–2063. [CrossRef]
62. Alvarez-Jimenez, C.; Sandino, A.A.; Prasanna, P.; Gupta, A.; Viswanath, S.E.; Romero, E. Identifying Cross-Scale Associations between Radiomic and Pathomic Signatures of Non-Small Cell Lung Cancer Subtypes: Preliminary Results. *Cancers* **2020**, *12*, 3663. [CrossRef]
63. Zhang, Z.; Li, X.; Sun, H. Development of machine learning models integrating PET/CT radiomic and immunohistochemical pathomic features for treatment strategy choice of cervical cancer with negative pelvic lymph node by mediating COX-2 expression. *Front. Physiol.* **2022**, *13*, 994304. [CrossRef]
64. Tran, W.T.; Jerzak, K.; Lu, F.-I.; Klein, J.; Tabbarah, S.; Lagree, A.; Wu, T.; Rosado-Mendez, I.; Law, E.; Saednia, K. Personalized Breast Cancer Treatments Using Artificial Intelligence in Radiomics and Pathomics. *J. Med. Imaging Radiat. Sci.* **2019**, *50* (Suppl. 2), S32–S41. [CrossRef] [PubMed]
65. Brancato, V.; Cavaliere, C.; Garbino, N.; Isgro, F.; Salvatore, M.; Aiello, M. The relationship between radiomics and pathomics in Glioblastoma patients: Preliminary results from a cross-scale association study. *Front. Oncol.* **2022**, *12*, 1005805. [CrossRef] [PubMed]
66. Castiglioni, I.; Rundo, L.; Codari, M.; Di Leo, G.; Salvatore, C.; Interlenghi, M.; Gallivanone, F.; Cozzi, A.; D’Amico, N.C.; Sardanelli, F. AI applications to medical images: From machine learning to deep learning. *Phys. Med.* **2021**, *83*, 9–24. [CrossRef]
67. Davison, B.A.; Harrison, S.A.; Cotter, G.; Alkhoury, N.; Sanyal, A.; Edwards, C.; Colca, J.R.; Iwashita, J.; Koch, G.G.; Dittrich, H.C. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. *J. Hepatol.* **2020**, *73*, 1322–1332. [CrossRef] [PubMed]
68. Brunt, E.M.; Clouston, A.D.; Goodman, Z.; Guy, C.; Kleiner, D.E.; Lackner, C.; Tiniakos, D.G.; Wee, A.; Yeh, M.; Leow, W.Q.; et al. Complexity of ballooned hepatocyte feature recognition: Defining a training atlas for artificial intelligence-based imaging in NAFLD. *J. Hepatol.* **2022**, *76*, 1030–1041. [CrossRef] [PubMed]
69. Car, J.; Sheikh, A.; Wicks, P.; Williams, M.S. Beyond the hype of big data and artificial intelligence: Building foundations for knowledge and wisdom. *BMC Med.* **2019**, *17*, 143. [CrossRef]
70. Willemink, M.J.; Koszek, W.A.; Hardell, C.; Wu, J.; Fleischmann, D.; Harvey, H.; Folio, L.R.; Summers, R.M.; Rubin, D.L.; Lungren, M.P. Preparing Medical Imaging Data for Machine Learning. *Radiology* **2020**, *295*, 4–15. [CrossRef] [PubMed]
71. Leslie, D.; Mazumder, A.; Peppin, A.; Wolters, M.K.; Hagerty, A. Does “AI” stand for augmenting inequality in the era of covid-19 healthcare? *BMJ* **2021**, *372*, n304. [CrossRef]
72. Liu, X.; Faes, L.; Kale, A.U.; Wagner, S.K.; Fu, D.J.; Bruynseels, A.; Mahendiran, T.; Moraes, G.; Shamdas, M.; Kern, C.; et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. *Lancet Digit. Health* **2019**, *1*, e271–e297. [CrossRef]
73. Wilson, A.; Saeed, H.; Pringle, C.; Eleftheriou, I.; Bromiley, P.A.; Brass, A. Artificial intelligence projects in healthcare: 10 practical tips for success in a clinical environment. *BMJ Health Care Inform.* **2021**, *28*, e100323. [CrossRef] [PubMed]
74. Sounderajah, V.; Ashrafian, H.; Rose, S.; Shah, N.H.; Ghassemi, M.; Golub, R.; Kahn, C.E.; Esteva, A.; Karthikesalingam, A.; Mateen, B.; et al. A quality assessment tool for artificial intelligence-centered diagnostic test accuracy studies: QUADAS-AI. *Nat. Med.* **2021**, *27*, 1663–1665. [CrossRef] [PubMed]
75. Cabitza, F.; Campagner, A. The need to separate the wheat from the chaff in medical informatics: Introducing a comprehensive checklist for the (self)-assessment of medical AI studies. *Int. J. Med. Inform.* **2021**, *153*, 104510. [CrossRef] [PubMed]
76. Cabitza, F.; Campagner, A. The IJMEDI Checklist for Assessment of Medical AI. 28 May 2021. Available online: <https://zenodo.org/record/4835800> (accessed on 12 January 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.