Engineering Applications of Artificial Intelligence 133 (2024) 108436

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Survey paper

Deep learning for liver cancer histopathology image analysis: A comprehensive survey

Haoyang Jiang [a,1], Yimin Yin [d,e,1], Jinghua Zhang [b,c,*], Wanxia Deng [b], Chen Li [f]

[a] School of Ocean Engineering, Yantai Institute of Science and Technology, Yantai, Shandong, China
[b] College of Intelligence Science and Technology, National University of Defense Technology, Changsha, Hunan, China
[c] Center for Machine Vision and Signal Analysis, University of Oulu, Oulu, Finland
[d] College of Systems Engineering, National University of Defense Technology, Changsha, Hunan, China
[e] School of Mathematics and Statistics, Hunan First Normal University, Changsha, Hunan, China
[f] College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, Liaoning, China

ARTICLE INFO

Keywords:
Deep learning
Liver cancer
Computational histopathology
Whole slide images
Survey

ABSTRACT

Liver cancer is the predominant cause of cancer-related fatalities globally, wherein Hepatocellular Carcinoma (HCC) and Intrahepatic Cholangiocarcinoma (ICC) emerge as the principal subtypes. Histopathology images, revered as the definitive benchmark for liver cancer diagnosis, yield rich phenotypic information, instrumental in facilitating disease progression prediction and potential survival prognostication. Deep learning has been rapidly developed recently and has become the mainstream technique for liver cancer histopathology image analysis, showing noteworthy accomplishments. This article undertakes a comprehensive examination of over 50 publications within the domain of deep learning-based liver cancer histopathology analysis, systematically discussing many advanced approaches. We commence our exploration by elucidating diverse facets of this field, encompassing problem formulation, general learning paradigms, and main challenges. Subsequently, we present a meticulous summary of publicly accessible datasets and evaluation metrics. To foster a deeper understanding of the research status of this domain, we furnish a taxonomy covering supervised learning and weakly supervised learning approaches within the specific tasks, i.e., classification and localization for histopathology diagnosis as well as deep learning-based survival models for disease prognosis. Finally, we discuss existing open issues and potential future trends within the realm of computational histopathology in liver cancer research.

1. Introduction

In the realm of cancer-related disease examination and diagnosis, Histopathology Image Analysis (HIA), often stained with Haematoxylin and Eosin (H&E), represents the gold standard (Xu et al., 2017). In comparison to cytology images, histopathological images furnish more pathological information and morphological characteristics essential crucial for diagnosis and phenotypic information for prognosis, including indicators like lymphocytic infiltration of cancer (Gurcan et al., 2009). Traditionally, histopathology image diagnosis entails pathologists examining images at varying magnifications under a microscope, with diagnostic outcomes contingent solely upon the subjective judgments of these experts. Notably, variations may arise due to differences in diagnostic methodologies and the varying levels of experience among pathologists. The advent of whole slide scanners presents a promising avenue for standardizing diagnostic procedures. During the nascent stages of histopathology diagnosis, several researchers delved into quantitative image analysis for disease grading, as exemplified by Gleason scores for prostate cancer (Epstein et al., 2016). Furthermore, The Cancer Genome Atlas (TCGA) provides a trove of thousands of digital histopathological images for free utilization, significantly catalyzing the advancement of computerized image analysis and machine learning techniques.

Liver cancer stands as the two leading cause of mortality in cancer-related diseases globally, posing a substantial threat to human well-being (Chhikara and Parang, 2023). Within the realm of primary liver cancer, Hepatocellular Carcinoma (HCC) and Intrahepatic Cholangiocarcinoma (ICC) represent two distinct subtypes. Notably, HCC cases predominate, while ICC cases remain relatively scarce (Liu et al., 2015; Massarweh and El-Serag, 2017). Early-stage intervention holds the promise of optimizing life expectancy and, in some cases, achieving

* Corresponding author at: College of Intelligence Science and Technology, National University of Defense Technology, Changsha, Hunan, China.
E-mail address: zhangjingh@foxmail.com (J. Zhang).
[1] Haoyang Jiang and Yimin Yin contributed equally.

https://doi.org/10.1016/j.engappai.2024.108436
Received 10 December 2023; Received in revised form 7 February 2024; Accepted 8 April 2024
Available online 18 April 2024

complete remission (Goceri et al., 2016). However, most liver cancer patients are diagnosed with middle or advanced stage, significantly decreasing the overall survival (Singal et al., 2014). Various diagnostic modalities are employed in the assessment of liver cancer, including tumor biomarkers like Alpha-fetoprotein (AFP) (Yamamoto et al., 2010) and Carbohydrate Antigen 19-9 (CA 19-9) (Malaguarnera et al., 2013) for the early diagnosis and prognosis of HCC and ICC, respectively. Additionally, medical imaging, such as Ultrasonography (US), Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Positron Emission Tomography (PET) emerges as another effective diagnostic method. However, tumor biomarker analysis entails a time-consuming and costly procedure, while medical imaging, although valuable, offers only preliminary diagnoses within the radiology department. As a comparison, histopathological images are commonly used as the final diagnosis, and still a gold standard.

In recent years, data-driven Deep Learning (DL), particularly Convolutional Neural Networks (CNNs), has found widespread application across various computer vision domains, demonstrating remarkable performance in areas such as natural image classification (Krizhevsky et al., 2017), object detection (Liu et al., 2020), and image segmentation (Minaee et al., 2021). Among the realm of medical image analysis within radiology, CADs based on DL techniques have achieved notable milestones (Chen et al., 2022b) and gained preliminary exploration in histopathology (Srinidhi et al., 2021). For liver cancer histopathology, DL-based approaches have primarily been applied in diagnosis and prognosis, mainly involving disease classification (Sun et al., 2019; Liao et al., 2020) and lesion localization (Wang et al., 2021a; Feng et al., 2021) within diagnostic models, and risk stratification in prognostic models (Saillard et al., 2020; Muhammad et al., 2021). Most of these studies have adopted supervised learning techniques, but training a high-performance model necessitates access to substantial training data and high-quality annotated images. The scarcity of liver cancer histopathological images constitutes a fundamental challenge for DL models, bringing a hurdle for the broader deployment of supervised learning models. To overcome this challenge, transfer learning and weakly supervised learning approaches have gained prominence in recent years, relieving the limitations arising from the lack of training data and the acquisition of high-quality annotated images to a certain degree. Despite the achievements of DL-based methods on specific datasets, the application of DL-based methods in the context of liver cancer HIA remains infancy, and there is a need to further explore the setting in the real clinical application.

Within this comprehensive survey, we focus on the application of DL-based approaches in the domain of liver cancer HIA, encompassing both HCC and ICC. It is noteworthy that approximately 85% of the studies within this survey have directed their attention towards the application of DL techniques in the context of HCC, reflecting the scarcity of ICC patients for study. In contrast to several recent surveys involving liver cancer HIA that have offered limited insights and lacked systematic summaries, our approach aims for a more comprehensive examination. For instance, Srinidhi et al. (2021) comprehensively summarized the application of DL techniques in multiple cancer HIA, encompassing breast, colon, lung, and others. However, this survey ignored the specific discussion of liver cancer HIA, only involving two papers. Calderaro and Kather (2021) and Calderaro et al. (2022) successively reported DL techniques applied in liver cancer HIA, but provided only brief introductions, without engaging in an in-depth discussion of these DL models. Therefore, our endeavor seeks to bridge this gap by offering a comprehensive survey of DL-based approaches specifically tailored for liver cancer HIA. In doing so, we aim to not only shed light on the current research status in this field but also provide guidance and inspiration for innovative explorations in liver cancer HIA, with potential applicability extending to other cancer types. The contributions of this survey are:

- A systematic overview of advanced DL approaches was provided in the context of liver cancer HIA;
- Problem definition, learning paradigms, mainstream publicly available datasets, and evaluation metrics within this domain were introduced;
- We presented a well-structured taxonomy for various DL techniques, covering two main applications: cancer diagnosis and prognosis;
- Finally, we deeply explored existing open issues, alongside feasible countermeasures, while discussing future directions.

We comprehensively surveyed more than 50 papers encompassing a wide spectrum of DL applications within the domain of liver cancer HIA, virtually encompassing the entirety of existing DL-based methodologies. These papers mainly span the years 2020 to 2023, reflecting the rapid proliferation of research in this field. These pertinent journal papers and conference papers have systematically been searched by *Google Scholar* and databases such as *IEEE Xplore, Elsevier, SPIE, Springer, Nature, Wiley online library*. Keywords used to research related papers are: *("liver cancer" or "hepatocellular carcinoma" or "intrahepatic cholangiocarcinoma") and ("histopathology" or "Whole-slide Images (WSIs)" or "digital pathology") and ("deep learning" or "convolutional neural networks" or "deep neural networks")*. To present an accurate representation of the research landscape in this domain, our review prioritizes papers published in top-tier journals and conferences, as well as the most influential, enlightening, and cutting-edge studies. This survey encompasses a broad spectrum of methodologies to discuss specific applications, including diagnosis and prognosis. In order to ensure the survey remains highly relevant for an extended period, our primary focus is on cutting-edge works covering up to and including December 2023.

The remainder of this survey is structured as follows. In Section 2, we provide an overarching introduction that outlines the problem definition, learning paradigms, and main challenges. Besides, the taxonomy is introduced to generalize the overall methodology in the context of DL liver cancer HIA. Section 3 presents mainstream datasets and several frequently-used evaluation metrics. Moving on to Section 4, we commence by introducing several pre-processing methods (Section 4.1). Subsequently, we embark on a detailed discussion of DL models applied in liver cancer histopathology diagnosis, further categorizing them into supervised learning and weakly supervised learning, segregated by specific tasks, i.e., classification (Section 4.2) and localization (Section 4.3). Section 5 discuss some survival models implemented by DL approaches for disease prognosis. In Section 6, we discuss several existing open issues and future directions in liver cancer HIA. To end with a conclusion in Section 7.

2. Overview

This section aims to provide a preliminary introduction to relational organizational structure in the context of DL-based liver cancer HIA. This first involves the problem definition in Section 2.1. Main learning paradigms applied in liver cancer HIA introduced in Section 2.2. Section 2.3 discussed several core challenges. Finally, Section 2.4 presents a structured taxonomy covering learning methodologies corresponding to specific tasks.

2.1. The problem

The primary objective in liver cancer HIA is to develop DL models capable of acquiring histological representation features and underlying characteristics. This endeavor can be divided into two distinct sub-problems: classification and localization. When dealing with the classification problem, the initial step involves partitioning WSIs into numerous patches, which are then fed into the DL model for forming the patch-level outcomes. These entail the model assigning each patch to its designated category (e.g., cancerous or normal) within the local region. The image-level (or slide-level) outcomes involve amalgamating

the patch-level results by the corresponding aggregation approaches. Conversely, localization revolves around devising algorithms, often leveraging heatmaps or segmentation techniques, that can accurately pinpoint target tissue regions, such as cancerous areas. Building upon the classification problem, the network is adapted to produce probability distribution heatmaps for highlighting the abnormal regions. In the realm of liver cancer HIA, segmentation can be likened to a pixel-wise classification problem. Initially, the patches are input into the segmentation model, and subsequently, they are assembled based on their spatial distribution to generate the segmentation map of WSIs. A key challenge in this context lies in the extraction of discriminative histological structural features using DL models.

2.2. Learning paradigms

In the context of liver cancer HIA, this study aims to elucidate the key learning paradigms of DL models, primarily focusing on supervised and weakly supervised learning. While supervised learning dominates current research, a notable segment explores weakly supervised models. The paper comprehensively details these paradigms, additionally discussing the aggregation approaches used to form the slide-level prediction and the general architectures in the domain of liver cancer HIA.

Supervised learning. In supervised learning, a model is trained with a set of N training samples $\{x_n, y_n\}_{n=1}^{N}$ of input patches x cropped from each WSI and the labels y. In binary classification tasks, this y is commonly a class label with a scalar {0, 1} whereas pixel-wise masks in localization tasks. The goal of supervised learning is to train a model $f_\theta : x \rightarrow y$ to predict the data based on loss function $L(\hat{y}, y)$, where the y denotes the real label and the $\hat{y}$ denotes the output with the predicted probability distribution of the model.

Weakly supervised learning. As for liver cancer histopathology scenarios, the scarcity of high-quality annotated data poses a serious hurdle to the development of DL models. Weakly supervised learning largely mitigates this issue and needs only coarse-grained (image-level) labels. Specifically, given the training dataset $\{X_i, Y_i\}_{i=1}^{N}$, where X_i represents WSIs and Y_i represents the corresponding labels (i.e., weak labels), and the DL model $f_\theta : x \rightarrow y$ automatically infers each fine-grained (patch/pixel-level) labels x_i.

Aggregation. Within the classification problem, DL models often predict each patch as a target category and form the patch-level outcomes. Among the literature we surveyed, there are two main approaches used to generate the image-level outcome, including the average of all outcomes of the predicted patches from WSIs (1) and the selection of the most predictive patches of the WSI (2). These two approaches have the capability to classify histopathology images into the target category using a predefined threshold.

$$g(\hat{y}) = \frac{1}{N}\sum_{i=1}^{N}\hat{y}_i \tag{1}$$

$$g(\hat{y}) = \frac{1}{N}\sum max_{i=1}^{N}\hat{y}_i \tag{2}$$

General architectures. Two main architectural frameworks are widely used in liver cancer HIA. The first framework comprises a feature extractor coupled with a classifier, typically employing a softmax classifier. Conversely, the second framework is the encoder–decoder structure. When addressing classification problems, the first framework is commonly used. Within this framework, feature vectors are extracted by the backbone, and the classifier is responsible for the classification of target classes. The cross-entropy loss function is generally employed to weigh the difference between the prediction value and the actual labels. The minimization of the loss function in DL-based models is accomplished iteratively through backpropagation algorithms (Rumelhart et al., 1986). In pursuit of fostering the acquisition of both local and global contextual information, several studies (Yang et al., 2022; Diao et al., 2022; Feng et al., 2021) adopted dual parallel backbones to learn features from both low-scale images and high-scale images. Binary classification and multiclass classification are the primary applications within the classification problem domain. Distinguishing images into cancerous or normal constitutes a common application for binary classification. In contrast, multiclass classification typically involves further categorizing abnormal images into poorly differentiated, moderately differentiated, and well-differentiated classes. The second framework, the encoder–decoder structure, is usually utilized in segmentation models for addressing localization problems. In this framework, the encoder focuses on feature extraction, while the decoder up-samples the images to map the original size, thus forming segmentation maps. These two architectural frameworks serve as the cornerstone of most studies, employed in both classification and localization tasks. A clear comparison of these two architectures is presented in Fig. 1.

2.3. Main challenges

Liver cancer HIA poses a formidable challenge for training DL models. The oversized nature of histopathological images renders it impractical to directly input them into DL models while preserving intricate details of texture and nucleus features. Additionally, histopathology images often exhibit color inconsistency, even within patches of the same image, thereby significantly complicating DL model identification. The scarcity of training data and the limited availability of high-quality annotated images present a substantial challenge in the realm of liver cancer HIA, notably impacting supervised learning. This section elucidates the specific details of these challenges.

Oversized histopathological image. A histopathological image contains over 10 000 × 10 000 pixels, even up to 100 000 × 100 000 pixels (Das et al., 2020; Courtiol et al., 2018). Given the constraints imposed by limited computational resources and memory capacities, it becomes impractical to directly input the entirety of WSIs into a DL model. One straightforward approach is to down-sample the WSI to the appropriate size (e.g., 256 × 256 pixels), resulting in the loss of valuable contextual information. The other alternative, widely adopted by a majority of studies, is to crop WSIs into a multitude of "patches" (also known as "tiles") with a fixed size. Within the domain of natural image analysis, each image essentially represents such individual Region Of Interest (ROI) that DL models can easily extract the corresponding features, thereby striving to attain the ideal performance. However, unlike natural images, histopathological images necessitate a nuanced consideration of both ROIs, such as microvascular invasion regions, and the surrounding texture information. Therefore, these patches may offer limited feature information, notably in the context of accurately localizing lesions, thus underscoring the intricacies of HIA.

Color inconsistency. In the process of histopathological imaging, manual operations (e.g., tissue section fixed, embedding, cutting, and staining) are the critical steps. Nevertheless, influence factors, such as the diversity of the staining procedure, may lead to the emergence of a challenging issue—namely, intra-class color discrepancy (Zanjani et al., 2018). An illustrative depiction of color-inconsistent images is presented in Fig. 2. While this issue can be deemed acceptable by pathologists during the diagnostic process, its impact on the performance of DL models is substantial. Consequently, the endeavor to construct robust and universally applicable DL models for liver cancer HIA presents a formidable challenge for researchers.

Insufficient training data. Boosting the generalization capability of DL models necessitates the availability of ample training data coupled with high-quality labels, a circumstance that regrettably remains unsatisfactory in the domain of liver cancer histopathology. This challenge is multifaceted, encompassing several critical factors. For example, the undertaking of labeling the WSIs, patches, or even individual pixels demands the expertise of highly trained pathologists, entailing labor-intensive and costly efforts. Additionally, training a high-performance DL model mandates extensive and diverse histopathological images, often involving multi-institutions. However,

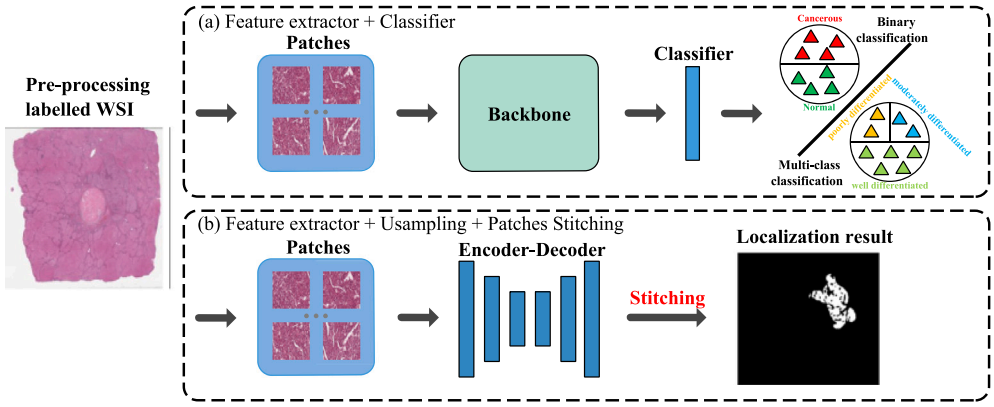

Fig. 1. Two main architectures applied in liver cancer histopathology. (a) is "Feature extractor" + "Classifier", which is commonly used in classification or partial localization tasks. (b) is "Encoder + Decoder", whose structure is usually used as the segmentation model to locate lesion regions of image patches that are stitched as the segmentation map of WSIs.

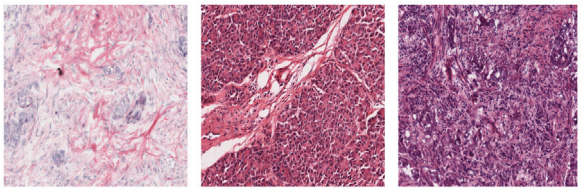

Fig. 2. Examples of color inconsistent images from TCGA-LIHC (Erickson et al., 2016).

some studies cannot make their datasets for public utilization due to the need to adhere to stringent institutional privacy protection requirements safeguarding patient data.

Summary. Among the above-mentioned challenges, we can see that these challenges largely hamper the development of DL models in liver cancer HIA. Although various countermeasures are used in liver cancer histopathology, they partly relieve a specific challenge and do not fully overcome these challenges. Subsequently, the existing countermeasures are summarized in Section 4, including pre-processing methods (Section 4.1) and learning-efficient paradigms, and future feasible countermeasures are discussed in Section 6.

2.4. Taxonomy

To attain a comprehensive understanding of the research status pertaining to DL-based methodologies in the domain of liver cancer HIA, we present a taxonomy elucidating existing approaches. Our taxonomy meticulously organizes DL-based methods, focusing on two main clinical applications: cancer diagnosis and prognosis. As delineated in Fig. 3, we methodically categorize these existing approaches into two principal categories, namely classification and localization, aligning with their specific tasks within diagnostic models. Furthermore, we subdivide them into supervised learning approaches and weakly supervised learning methods. Among deep survival learning for prognosis, we provide an in-depth examination of survival models implemented by DL-based methods, specifically tailored for the prognosis of liver cancer patients.

3. Datasets and evaluation metrics

This section introduced several mainstream publicly available datasets and evaluation metrics. The details of these datasets were presented, encompassing characteristics, sources, corresponding tasks, and other pertinent information. Subsequently, the evaluation metrics relevant to specific applications, such as diagnostic and prognostic models, were introduced.

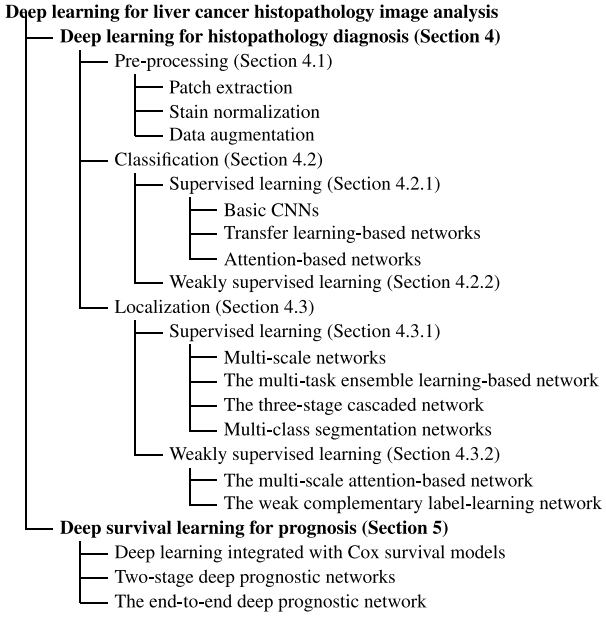

Fig. 3. The taxonomy of representative DL-based methods applied in liver cancer HIA.

3.1. Datasets

To continuously enhance the performance of DL models, the imperative prerequisite is access to a substantial dataset that encompasses a significant number of meticulously annotated images. The collection of digitalized histopathology images typically involves medical institutions, which frequently include medical centers affiliated with educational institutions or specialized hospitals. The annotation of these images, whether at the image-level, patch-level, or pixel-level, is meticulously carried out by expert pathologists. In the context of classification tasks, both patch-level and image-level annotations are frequently utilized. Typically, professional pathologists delineate abnormal regions, and subsequently, patches are extracted from the corresponding tissue regions, such as cancerous or normal regions, to provide patch-level annotations. Image-level annotations can be regarded as weak labels since only the category of the histopathology images is provided, without specifying the type of the particular tissue regions. For localization tasks, accurate ground-truth binary pixel masks, referred to as pixel-level annotations, are typically necessary. Similarly, patches are cropped from the corresponding tissue regions

Table 1
Summary of publicly available datasets in liver cancer histopathology.

Dataset/Year	Cancer subtype	Stain	Dimension	Image format	Task	Annotation types	Description
TCGA-LIHC	HCC	H&E	Varying	SVS	Classification	Image-level	(1) Presenting tissue WSIs annotated with the corresponding category (cancerous or normal) from 377 cases and diagnostic WSIs from 365 cases, but lacking detailed pixel-level annotations for segmentation tasks; (2) Providing clinical, biological, survival, and pathological data, benefiting for researches on survival analysis and others.
PAIP 2019	HCC	H&E	Varying	SVS	Segmentation	Pixel-level	(1) Encompassing 100 WSIs (50 training slides, 40 testing slides, and 10 validating slides); (2) Furnishing ground-truth binary pixel masks and delineating tissue regions into the two categories: the comprehensive tumor area and the viable tumor area by distinct color lines.
KMC 2021	HCC	H&E	1920 × 1440	–	Classification and segmentation	–	Providing 257 WSIs with four sub-types (70 WSIs belonging to sub-type 0, 80 WSIs belonging to sub-type 1, 83 WSIs belonging to sub-type 2, and 24 WSIs belonging to sub-type 3) and 80 WSIs with annotated nuclei.

to facilitate training a DL model. In this section, we present a comprehensive summary of three liver cancer histopathology datasets that are accessible to the public. Furthermore, an exhaustive account of the specific details of these datasets is elaborated upon subsequently, as presented in Table 1 for enhanced clarity and reference.

The **TCGA-LIHC** (Erickson et al., 2016) stands as a pivotal milestone as the initial publicly accessible dataset for HCC histopathological images. This dataset is the project of TCGA from the National Cancer Institute Genomic Data Commons. Comprising a comprehensive compilation of tissue slide images sourced from 377 patients and diagnostic slide images sourced from 365 patients, TCGA-LIHC offers image-level annotations elucidating the categorization of WSIs. However, it is noteworthy that the dataset does not furnish detailed segmentation annotations specifically delineating liver cancer tissue regions. Beyond WSIs, this dataset also provides clinical, biological, survival, and pathological data. This multifaceted dataset thereby significantly contributes to the advancement of prognostic models and related research endeavors.

The **PAIP 2019** (Kim et al., 2021), a constituent of the MICCAI 2019 Grand Challenge for Pathology, comprising 100 WSIs. This dataset elaborately divides a training set featuring 50 slides, a testing set comprising 40 slides, and a validation set including 10 slides. Each WSI is stained using the H&E and subsequently subjected to scanning through the Aperio AT2 scanner at a magnification factor of × 20. A distinctive hallmark of the PAIP dataset is its provision of pixel-level annotations implemented by expert pathologists affiliated with the Seoul National University Hospital. These annotations discerningly designate tissue regions within WSIs. Moreover, the dataset delineates tissue regions into two distinct categories: the comprehensive tumor area and the viable tumor area, which are thoughtfully demarcated using distinct color lines.

The **KMC** dataset (Aatresh et al., 2021; Lal et al., 2021) encompasses 257 HCC original H&E stained WSIs. Each WSI within this dataset boasts dimensions of 1920 × 1440 pixels, and they are sourced from distinct patients. The dataset categorizes these WSIs into four distinct sub-types, with each sub-type exhibiting varying quantities. Specifically, sub-type 0 comprises 70 WSIs, sub-type 1 encompasses 80 WSIs, sub-type 2 includes 83 WSIs, and sub-type 3 contains 24 WSIs. Furthermore, there are 80 liver cancer H&E stained histopathological images with annotated nuclei within each WSI. These images are acquired at the same microscopic zoom level (40x) using an Olympus scanner. These images were collected and annotated by a clinical pathologist from Kasturba Medical College (KMC), Mangalore, Manipal Academy of Higher Education (MAHE), Manipal, Karnataka, India.

Table 2
Summary of commonly-used evaluation metrics in liver cancer histopathology.

Task	Metric	Definition
Classification	Accuracy (Acc)	$\frac{TP+TN}{FP+FN+TP+TN}$
	Precision (Pr)	$\frac{TP}{FP+TP}$
	Recall (Re) = Sensitivity (Sens)	$\frac{TP}{FN+TP}$
	Specificity (Spec)	$\frac{TN}{FP+TN}$
	F1-score (F1)	$\frac{2\times Precision\times Recall}{Precision+Recall} = \frac{2TP}{FP+FN+2TP}$
	ROC curve	True Positive Rate (TPR), False Positive Rate (FPR)
	AUC	Area under the ROC curve
Localization	Jaccard index (Jaccard or IOU)	$\frac{A\cap B}{A\cup B} = \frac{TP}{TP+FP+FN}$
	Dice score (Dice)	$2\frac{\lvert A\cap B\rvert}{\lvert A\rvert+\lvert B\rvert} = \frac{2TP}{FP+FN+2TP}$

3.2. Evaluation metrics

Various evaluation metrics serve as quantitative measures to assess the performance of DL models for the specific task. Taking the binary classification task as an example, the terminology includes True Positive (TP), which signifies accurately predicted cancerous conditions, True Negative (TN), denoting accurately predicted normal conditions, False Positive (FP), representing inaccurately predicted cancerous conditions, and False Negative (FN), indicating inaccurately predicted normal conditions. Notably, Accuracy (Acc) stands as the primary metric for quantifying the proportion of correct predictions across all samples in classification models. For localization models, the widely adopted metric is the Jaccard index, also known as the Intersection Over Union (IOU), which quantifies the degree of overlap between predicted images and the ground-truth labels. In contrast, for prognostic models, the Concordance Index (C-Index) serves as a prevalent evaluation metric. The C-Index assesses the alignment between the predicted ranks and the actual observed survival outcomes (Steck et al., 2007; Blanche et al., 2019), where a value of 0.5 suggests complete randomness and a value of 1 signifies a perfect prediction. For a comprehensive overview of evaluation metrics commonly employed in the assessment of DL models within the domain of liver cancer histopathology, please refer to Table 2.

4. Deep learning for histopathology diagnosis

The objective of this section is to furnish a comprehensive and methodical overview of DL-based methodologies employed in the diagnosis of liver cancer histopathology. As highlighted in Section 2.3, numerous challenges significantly impede liver cancer HIA. To address these

challenges, several studies have initiated efforts to adopt diverse approaches, including pre-processing (Section 4.1) and learning-efficient methods. Concerning learning-efficient methods, these methodologies can be categorized into two distinct groups: classification (Section 4.2) and localization (Section 4.3), aligning with the specific nature of the addressed problem. Subsequent sections delve into the substantive contributions and innovations inherent in these DL-based approaches. Moreover, we underscore their performance in the context of specific applications on either public datasets or private datasets. Towards the culmination of this section, we briefly review these methodologies and discuss their inherent limitations. More details are presented as follows.

4.1. Pre-processing

Due to the predominant challenges primarily associated with training data, numerous studies have inclined towards employing pre-processing methods to augment the quality of these data, thereby bolstering the generalization capabilities of DL models. When considering the adoption of pre-processing approaches, key methods include patch extraction, stain normalization, and data augmentation.

Patch extraction for addressing oversized histopathological images. A gigapixel histopathological image is unlikely directly fed into a DL-based model; for this reason, the patches with fixed size (e.g., 256 × 256 pixels) cropped from WSIs may be more appropriate. The simplest approach is to use the sliding window with the fixed size and step to crop WSIs into non-overlapping (Wang et al., 2021a) or overlapping patches (Chen et al., 2021). Alternatively, an effective approach involves the utilization of the OpenSlide library (Goode et al., 2013), a tool that facilitates the processing of WSIs for patch extraction. Furthermore, partial studies have embraced Otsu's method (Otsu, 1979) to segregate tissue regions and filter out extraneous blank backgrounds before patch cropping, thereby significantly curbing computational cost.

Stain normalization for addressing color inconsistency. Addressing the pervasive challenge of color heterogeneity intrinsic to histopathological images due to variations in staining workflows across multiple institutions (BenTaieb and Hamarneh, 2017), an effective approach is to use Histogram Equalization Algorithm (HEA), a strategy consistently applied across several studies. Notably, Feng et al. (2021) introduced an innovative approach that recognizes the impact of white backgrounds on tissue region normalization with respect to color characteristics. To circumvent this interaction, they proposed a partial color normalization technique grounded in linear transformations, with a specialized focus on tissue regions by a specific mask.

Data augmentation for addressing insufficient training data. In the domain of histopathology, augmenting the training dataset with diverse images becomes pivotal in enabling DL models to discern known and underlying morphological features, consequently bolstering their generalization and robustness (Campanella et al., 2019). However, the paucity of training data has stymied the development of universally applicable DL models in the context of liver cancer histopathology. To mitigate this challenge, data augmentation techniques have emerged as an approach, widely adopted across many studies. Typical data augmentation methods encompass geometric transformation (e.g., flips, rotations, and translations) and color transformation (e.g., color space translation, color augmentation, and blur). These methods, while not entirely eliminating this challenge, serve to partially enhance the generalization of DL models.

4.2. Classification

When addressing classification problems, DL models are directed towards the discernment of target categories within histopathological images, notably distinguishing between characteristics indicative of benign and malignant samples. The exhaustive setting of the classification problem is provided in Section 2.1. As shown in Fig. 4, we identify two main learning paradigms behind these DL models: supervised learning (Section 4.2.1) and weakly supervised learning (Section 4.2.2). Within the ambit of the first paradigm, a trifecta of core learning methodologies has been devised to effectively tackle the specific challenges inherent to liver cancer histopathology diagnosis, including basic CNNs, transfer learning-based networks, and attention-based networks. Finally, the second paradigm primarily addresses the constraints arising from a dearth of high-quality labeled images. Within this paradigm, DL models are closely related to Multiple Instance Learning (MIL) approaches (Dietterich et al., 1997). An overview of representative DL models, including supervised learning and weakly supervised learning approaches, is provided in Table 3.

4.2.1. Supervised learning

Basic CNNs. In recent years, DL models, particularly CNNs, have been widely applied in the domain of liver cancer HIA. A common practice involves the segmentation of histopathological images into numerous patches, with the model rendering predictions for each patch to classify them into their respective target categories, thus establishing a patch-level classification. For example, Li et al. (2017) combined the basic CNN with 12 fully connected layers, enabling the extraction of multi-form feature vectors. Subsequently, they integrated the Extreme Learning Machine (ELM) to effectively grade individual HCC nuclei into three distinct malignant degrees.

As the field has witnessed the emergence of more advanced and deeper models, such as VGG (Simonyan and Zisserman, 2014), Inception (Szegedy et al., 2015), and ResNet (He et al., 2016), researchers have progressively adapted these models to the realm of liver cancer HIA. For example, Kiani et al. (2020) harnessed DenseNet (Huang et al., 2017) to classify image patches from tumor regions as either HCC or ICC within WSIs. They then aggregated individual patch-level probabilities to formulate slide-level outcomes. Employing Class Activation Maps (CAMs) (Zhou et al., 2016), the heatmaps were generated to enhance the interpretability of DL models by highlighting the tissue regions closely associated with the diagnostic outcomes. Additionally, Liao et al. (2020) introduced residual conception into their DL model to distinguish normal or HCC image patches within WSIs and Tissue Microarrays (TMAs). Subsequently, they utilized two methods to derive slide-level results, including the average of each patch from the corresponding WSI and the summary of the percentage of classified positive patches from the corresponding WSI (≥0.5). To further discover the correlation between outcomes obtained by their model at varying magnifications, they used the model trained on patches at both 5× and 20× magnifications. Experimental results revealed a robust correlation between outcomes across these two magnifications, with the model achieving an exceptionally high level of performance (e.g., patch-level AUC = 0.975 at 20× magnification). These studies have indicated that these advanced models and groundbreaking structure designs have stronger visual feature representation capabilities and can be better applied to liver cancer histopathological diagnosis.

Tumor mutation prediction constitutes a pivotal application within the realm of liver cancer diagnosis, offering profound insights into the intricate landscape of genomic mutation types that explain cancer development and progression. For example, Chen et al. (2020b) employed Incption-V3, pre-trained on ImageNet and fine-tuned on TCGA-LIHC, to predict HCC gene mutations, effectively distinguishing patches and WSIs into wild-type or mutated states. This model was trained to discern mutations within the ten most frequently encountered genes in HCC. Remarkably, their findings unveiled the significant identification of four mutated genes (CTNNB1, FMN2, TP53, and ZFX4) at the slide-level. Similarly, Liao et al. (2020) proposed a CNN with the residual structure to predict genomic mutation on both WSIs and TMAs. Their model was also trained on ten mutated genes, yielding experimental results that underscored its impressive performance in predicting mutated genes associated with HCC. Notably, this model expended less memory and time, alongside heightened generalization capabilities

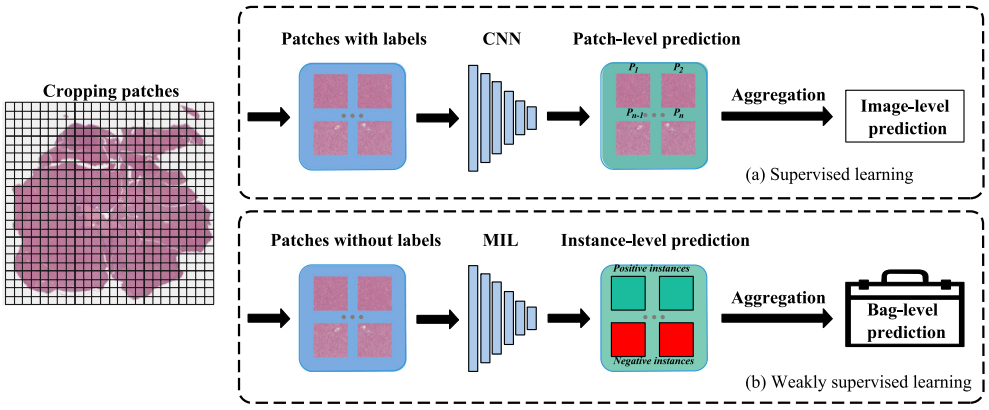

Fig. 4. Overall frameworks for liver cancer histopathology classification models. The WSI is cropped into abundant patches before the inference process. (a) represents a supervised learning paradigm, where image patches, along with their corresponding labels, are input into a CNN for feature extraction. Subsequently, predictions are made for each individual patch. The final image-level result is obtained through aggregation approaches. (b) illustrates the weakly supervised learning principle, with MIL-based models being the typical approach. In this context, the cropped image patches, lacking corresponding labels, form instances that are fed into the MIL-based model. This model can autonomously predict whether each instance (patch) and bag (image) is associated with positive or negative conditions.

when contrasted with Inception-V3. Particularly noteworthy was the model's proficiency in predicting CTNNB1 mutations, achieving an impressive AUC of 0.903 at the slide-level.

In addition to the prediction of genomic mutation types directly from HCC histopathological images, Zhang et al. (2019) embarked on an exploration aimed at predicting Tumor Mutational Burden (TMB) from these very images. They proposed the CNN model to prognosticate TMB levels, classifying them into high-level or low-level groups. It is imperative to highlight their noteworthy observation that they encountered a significant issue concerning over-fitting when applying several well-established DL models, including but not limited to AlexNet, VGG, and ResNet. The main reason behind this issue lies in the fact that these networks are predominantly designed for the processing of natural images, boasting larger receptive fields optimized for capturing global features. However, there is a need to consider the extraction of detailed local features (e.g., cellular structure) of histopathological images. In response to this need, they developed a novel CNN model comprising four pairs of convolution layer and max pooling layer followed by a fully connected layer with 256 neurons. Their meticulous experimentation revealed a pivotal insight: the size of the receptive field has a profound impact on the performance of DL models in the classification of TMB. Consequently, they achieved their most promising results when utilizing a 48×48 size for the receptive field, while the kernel size was configured at 5×5 for the initial convolutional layer and 3×3 for the subsequent three convolutional layers within their model.

Transfer learning-based networks. The dearth of a large-scale dataset poses a formidable challenge in the field of liver cancer HIA, prompting the exploration of an alternative approach involving the utilization of the pre-trained network. The objective revolves around leveraging knowledge acquired from a source domain and applying it to a target domain. Typically, the model is pre-trained on an extensive dataset such as ImageNet (Krizhevsky et al., 2017), with the resulting trained weights of all layers serving as the initialization for the target task.

In the context of liver cancer HIA, a similar strategy has been adopted across most studies. This strategy involves the utilization of a model pre-trained on ImageNet and subsequently applying it to the target dataset, such as TCGA-LIHC. For example, Chen et al. (2020b) used pre-trained Inception-V3 to implement binary-classification (normal *verse* tumor) and HCC multi-classification (well *verse* moderate *verse* poor), achieving promising performance. Another noteworthy study, proposed by Dong et al. (2022), delved into the fusion of several popular pre-trained models, including ResNet-50, VGG-16, DenseNet-201, and InceptionResNet-V2, to enhance the classification performance for multi-differentiated liver cancer. The study involved a series of carefully designed experiments encompassing the selection of fusion strategies and the choice of efficient classifiers. Experimental results indicated that the fusion of ResNet-50, VGG-16, and DenseNet-201 has the best generalization performance, particularly when integrated with an attention mechanism and a Stacking-based ensemble learning classifier, which outperformed other classifiers with an impressive Acc of 0.7246. However, it is essential to note that the generalization of this fusion model exhibited a slight decrement when applied to unknown data, potentially attributable to the limited training samples.

Attention-based networks. Attention mechanisms have proven to be highly effective in enhancing model performance by enabling the extraction of relevant feature information, particularly for the identification of tumor sub-types. For instance, inspired by BreastNet (Toğaçar et al., 2020), Aatresh et al. (2021) integrated the Convolutional Block Attention Module (CBAM) (Woo et al., 2018) and introduced a novel residual block into their model to capture the relevant features (e.g., tissue regions) and to filter irrelevant information (e.g., background). Moreover, they leveraged the Atrous Spatial Pyramid Pooling (ASPP) block (Chen et al., 2017a) with depth-wise separable convolutions (Howard et al., 2017), along with the hypercolumn technique, to effectively extract multi-scale features. The resulting model, named "liverNet", demonstrated remarkable performance, achieving an average Acc of 0.9772 on the TCGA-LIHC dataset for the classification of four categories of HCC image patches (normal, low sub-type HCC tumor, medium sub-type HCC tumor, and high sub-type HCC tumor), while expending a handful of calculation parameters.

To ascertain the impact of attention mechanisms on the generalization of DL models, Chen et al. (2022a) conducted a comparative study. They evaluated several popular DL models used for distinguishing between three differentially graded liver cancer histopathological images, including VGG-16, ResNet-50, ResNet-50_CBAM, SKNet (Li et al., 2019), and SENet (Hu et al., 2018) - the selected network in their study. Experimental results demonstrated that VGG-16 showed the poorest performance due to its relatively simple structure, which struggled to capture complex histological features. In contrast, ResNet-50, with its deeper architecture, outperformed VGG-16. Furthermore, the addition of CBAM to ResNet-50 yielded a slight improvement, particularly in the prediction of moderately differentiated images. SENet and SKNet also demonstrated remarkable performance, with SENet achieving the highest overall classification performance (Acc = 0.9527), particularly excelling in the prediction of poorly differentiated images. These results indicated that SENet is particularly well-suited for learning complex histology features and extracting fine-grained information.

Table 3
Overview of deep learning-based classification models for liver cancer histopathology.

Type	Reference	Application	Pre-processing	Dataset	Result	Methods, contributions, and limitations
Supervised learning						
	Li et al. (2017)	HCC nuclei grading	Patch with an individual nucleus extraction using Center-proliferation Segmentation (CPS)	Private dataset: 127 WSIs	Acc = 0.967	(1) Proposing a CNN with 12 fully connected layers to extract multi-form feature vectors and utilizing Extreme Learning Machine (ELM) as the classifier; (2) Employing a novel up-sample method contained in backpropagation algorithm; (3) But this study is not ideal for the segmentation of overlapping or adhesive cell nucleus.
	Zhang et al. (2019)	HCC TMB prediction	Patch extraction using the sliding window + Data augmentation by flexibly adjusting the step size	TCGA-LIHC: 380 WSIs from 362 patients	Patch-level: Acc = 0.9486, Auc = 0.9488; patient-level: Acc = 0.9971	(1) Proposing a CNN with modified receptive fields; (2) Only predicting the classification of TMB patients, rather than the TMB score.
	Liao et al. (2020)	Binary classification (benign and HCC) and mutation prediction	Patch extraction using OpenSlide library + Data augmentation using color transformation	TCGA-LIHC: 402 WSIs with 89 matched ones of adjacent normal tissue; private dataset: 455 HCC TMAs with 265 matched normal tissues	TCGA-LIHC: patch-level AUC = 0.9419, image-level AUC = 0.9937 (binary classification); patch-level AUC = 0.7172, image-level AUC = 0.8122 (mutation prediction)	(1) Developing a CNN with the residual structure to predict WSIs and TMAs; (2) But the predicted results of TMAs under two magnifications were significantly different. (3) The prediction of somatic mutation needs to be further improved.
	Chen et al. (2020b)	Binary classification (benign and HCC) + HCC multi-classification (well, moderate, and poor tumor differentiation) + mutation prediction	Patch extraction	TCGA-LIHC: 491 WSIs (402 normal and 89 HCC slides), 383 WSIs of HCC with available histopathological grade (55 well, 187 moderate, and 141 poor slides), and 387 WSIs of HCC with corresponded gene mutation information; private dataset: 67 HCC and 34 normal WSIs	Varying	(1) Utilizing the pre-trained Inception-V3 to deal with binary classification, multiclass classification, and gene mutation prediction; (2) Requiring further validating their DL model on other large-scale datasets.
	Kiani et al. (2020)	Binary classification (HCC and ICC)	Patch extraction by the reference GI pathologist	TCGA-LIHC; TCGA-CC; private dataset: 40 HCC and 40 ICC WSIs	Validation set: Acc = 0.885; independent test set: Acc = 0.842	(1) Employing DenseNet to assist pathologists in the real clinical setting; (2) Using Class Activation Maps (CAMs) to highlight the tissue regions closely associated with the diagnostic outcomes and improve the interpretability of the DL model; (3) Only focusing on the classification of HCC and ICC, and it is imperative to collect the data of other subtypes, thereby further enhancing clinical application.
	Lin et al. (2021)	Binary classification (HCC and normal)	Patch extraction	Private dataset: 29 WSIs	Acc = 0.9137, Sens = 0.9216, Spec = 0.9057	(1) Adopting Inception-V1 to classify HCC histopathology images; (2) Conducting the inverse power law function-based model to estimate the minimum number of annotated training images for satisfying the ideal performance of DL models; (3) Seeking further investigation into more efficient DL models and labeling methods, with an emphasis on scaling datasets.
	Aatresh et al. (2021)	HCC multi-classification (normal, low sub-type, medium sub-type, high sub-type tumor)	Patch extraction + Data augmentation using geometric transformation + Image normalization	TCGA-LIHC: 141 WSIs (60 WSIs with sub-type 0, 50 WSIs with sub-type 1, and 31 WSIs with sub-type 2); KMC	TCGA-LIHC: Prec, Re, F1, Acc = 0.9772, IOU = 0.9561; KMC: Prec, Re, F1, Acc = 0.9093, IOU = 0.836	(1) Proposing a novel architecture, called LiverNet, integrating with CNAM and ASPP; (2) Providing a novel publicly dataset, called KMC; (3) Seeking further validation of their DL model on additional large-scale datasets.
	Dong et al. (2022)	Liver cancer multi-classification (poorly-differentiated, moderately-differentiated, and well-differentiated tumor)	Stain normalization using Adaptive Histogram Equalization (AHE) algorithm + Gaussian filtering using to mitigate noise + Data augmentation using geometric transformation	Private dataset: 73 WSIs (24 poorly differentiated, 27 moderately differentiated, and 22 well differentiated WSIs)	Acc = 0.7246	(1) Fusing several pre-trained models (ResNet-50, VGG-16, and DenseNet-201) with both channel attention and spatial attention (2) Employing the Stacking-based ensemble learning classifier; (3) This DL model showcased the slightly poor performance in the test set.

(*continued on next page*)

Table 3 (*continued*).

Weakly supervised learning					
Chen et al. (2022a)	Liver cancer multi-classification (poorly-differentiated, moderately-differentiated, and well-differentiated tumor)	Data augmentation using geometric transformation	Private dataset: 74 WSIs (24 poorly differentiated WSIs, 28 moderately differentiated WSIs, and 22 highly differentiated WSIs)	VGG-16: Acc = 0.7695; ReNet-50: Acc = 0.9422; ResNet_CBAM: Acc = 0.9473; SENet: Acc = 0.9527; SKNet: Acc = 0.9515	(1) Comparing with several popular CNNs (VGG-16, ResNet-50, ResNet_CBAM, SKNet, and SENet); (2) The dataset used in this study is small and single, requiring the further validation on other large-scale datasets.
Sun et al. (2019)	Binary classification (HCC and normal)	Color transformation + Tissue extraction using Otsu's method + Color normalization using a Histogram Equalization Algorithm (HEA) + Patch extraction	TCGA-LIHC: 462 WSIs (79 normal WSIs and 383 HCC WSIs)	Pr = 1, Re and AUC over 0.95	(1) Employing the pre-trained ResNet-50 and an MLP classifier; (2) Applying the weakly supervised learning training paradigm to alleviate the insufficient training data; (3) There exists the issue of data imbalance due to the lack of normal WSIs and the prediction of normal images is relatively poor.
Tan et al. (2021)	ICC subtype classification (peripheral small duct type and perihilar large duct type)	Patch extraction	Private dataset: 235 WSIs from 119 patients	Acc = 0.7006	(1) Proposing a novel label smoothing using EfficientNet for the hidden class detection; (2) Utilizing MIL framework encompassing VGG-16 integrating with an MLP classifier for image-level classification; (3) The performance of this DL model should be further enhanced by incorporating gene mutation information without relying on manual selection based on the gene mutation data.
Zeng et al. (2022)	HCC immune and inflammatory gene signatures prediction	Tissue extraction + Patch extraction + Stain conversion + Color normalization + Data augmentation using geometric transformation	TCGA-LIHC: 349 slides from 336 tumors	Varying	(1) Applying the Ward2 algorithm and Euclidean distance to conduct hierarchical clustering of samples. (2) Comparing with three DL-based approaches (the patch-based CNN employing ShuffleNet, the classical MIL, and CLAM); (3) Requiring further improving the performance of their DL model.
Tan et al. (2023)	ICC subtype classification (peripheral small duct type and perihilar large duct type)	Patch extraction with three magnifications + Data augmentation using geometric transformation	Private dataset: 332 WSIs from 168 patients	Acc = 0.7715 (under tumor image patches)	(1) Introducing a multi-scale CNN built upon a Siamese contrastive learning network, featuring a VGG-16 backbone and a projection head. (2) Employing multi-scale attention for the integration of the meaningful features based on the MIL framework; (3) Only concentrating solely on the fusion of features from two scales.

4.2.2. *Weakly supervised learning*

Considering the deficiency of labeled histopathological images, the weakly supervised learning algorithm emerges as a promising direction. While it is relatively easy to acquire coarse-grained labels for entire liver cancer histopathology images, obtaining fine-grained (patch/pixel-level) annotations is exceedingly uncommon and expensive. This constraint hampers the improvement of the performance of DL models. Weakly supervised learning offers a solution to this issue by alleviating the need for fine-grained annotations (Xu et al., 2014).

Herein, MIL is the main approach of weakly supervised learning for liver cancer histopathology study and needs only a set of weakly labeled data. Specifically, MIL operates by utilizing coarse image-level histopathological images as "bags", wherein the model automatically predicts individual unknown patches as either positive instances (e.g., cancerous patches) or negative instances (e.g., normal patches). These instance-level predictions can be aggregated to form either positive bag-level outcomes, containing at least one positive instance, or negative bag-level outcomes, comprising exclusively negative instances, as shown in Fig. 4(b). For instance, Sun et al. (2019) first utilized coarse image-level annotated images to train ResNet-50, pre-trained on ImageNet, to generate feature vectors for each patch. These feature vectors then were fed into the MIL model, responsible for selecting the highest-scoring positive instances and the lowest-scoring negative instances. In this specific case, a total of 100 highest-scoring positive instances and 100 negative instances were chosen and utilized as input for a Multilayer Perceptron (MLP). The MLP was responsible for generating binary bag-level outcomes, classifying the images as either cancerous or normal.

In typical MIL scenarios, performance is often limited. Notably, Clustering-constrained Attention Multiple Instance Learning (CLAM) (Lu et al., 2021) is an efficient algorithm that incorporates an attention mechanism into the MIL framework, thereby further improving the generalization capabilities of DL models. For instance, Zeng et al. (2022) aimed to investigate the effectiveness of the basic patch-based CNN model based on ShuffleNet, the classical MIL, and CLAM for predicting six immune gene signatures from HCC histopathological images. In their approach, the authors first used the Ward2 algorithm (Murtagh and Legendre, 2014) and Euclidean distance to perform hierarchical clustering of samples, resulting in three distinct sample clusters: high, median, and low, for each gene signature. These clusters were subsequently labeled as either "cluster high" or "cluster median/low" to train three DL models. Among the three approaches, CLAM exhibited the most favorable performance. This was attributed to two key advantages of CLAM: firstly, the integrated attention mechanism allowed the model to focus on diagnostic tissue regions more effectively, particularly through adaptive weighting for each patch; secondly, CLAM generated visualized heatmaps that enhanced model interpretation, aiding pathologists in assessing the entire histopathological image.

In a different approach proposed by Tan et al. (2021), the authors optimized the generalization of MIL-based models for classifying two sub-types (small duct type and large duct type) of ICC through a two-stage approach. In the first stage, they introduced an additional class, referred to as the "hidden class", based on the hypothesis that patches from both sub-types shared similar features. EfficientNet (Tan and Le, 2019), pre-trained on ImageNet, was utilized to iteratively refine per-patch labels and assign patches to three reasonable classes, including the hidden class. In the second stage, the study aimed to validate the

performance of the patch filtering approach introduced in the first stage. This involved discarding patches from the hidden class to enable the subsequent model to learn more discriminative features. The MIL framework in this scenario employed pre-trained VGG-16 as the feature extractor, with an MLP serving as the classifier to implement binary image-level classification.

4.3. Localization

In localization networks, the primary objective is to pinpoint lesion regions, such as cancerous regions, often accomplished through the utilization of heatmaps or segmentation techniques. A detailed setting of localization problems is provided in Section 2.1. Among the domain of liver cancer histopathology, existing localization approaches can be categorized into several paradigms, including multi-scale networks, the multi-task ensemble learning-based network, the three-stage cascaded network, and multi-class segmentation networks within the supervised learning paradigm. In the context of weakly supervised learning, representative approaches encompass the multi-scale attention network and the weak complementary label-learning network. The systematic discussion of these methods is presented in the following sections, and an overview of representative DL models is presented in Table 4.

4.3.1. Supervised learning

Multi-scale networks. At a specific magnification, individual patches exhibit constrained histological features, and the extensive contextual information is often compromised due to successive feature map reductions. To mitigate this limitation, multi-scale networks offer an effective approach by amalgamating both local and global information, thereby enhancing the efficacy of DL models. As shown in Fig. 5, multi-scale approaches reported in liver cancer histopathology can be further subcategorized into two distinct implementations: multi-magnification networks and multi-resolution networks, each with its unique strategies for addressing this problem.

Multi-magnification networks were designed to extract features from images at different magnifications, effectively merging local and global features, as illustrated in Fig. 5(a2). In this approach, image patches maintain the same size while exhibiting distinct spatial structures. For example, a single-scale CNN captures features from patches at the specific magnification, whereas the multi-scale CNN, as proposed by Huang et al. (2019), employed two parallel CNNs to extract features from patches in WSIs at both 5× and 20× magnification, with fusing the features from these two magnifications to produce the final outcome. Similarly, Feng et al. (2021) developed a seven-level Gaussian pyramid multi-scale model for HCC tumor region segmentation, using U-Net (Ronneberger et al., 2015) as the backbone. Each level's histopathological image size is reduced to one-fourth of the preceding level, with the bottom level representing the original 20× magnification image without color normalization. Challenges emerged during patch stitching, including seams and artifacts at patch boundaries, which were addressed by refining the weight map calculation using an overlapped patch extraction and assembling method proposed by Cui et al. (2019). A voting mechanism was employed to combine the multi-scale features, and the segmentation results at each level were mapped back to the original image size before fusion. Experimental results on the PAIP 2019 dataset demonstrated that their model achieved state-of-the-art performance, with a Jacard score of 0.7964.

However, the typical multi-scale CNN, as demonstrated by Huang et al. (2019), encounters a spatial alignment challenge during integration. This issue arises because features extracted from various magnifications (e.g., 2.5×, 5×, 10×, or 20×) are not co-located on the original images. To address this limitation, Yang et al. (2022) developed a feature-aligned multi-scale CNN. Their model also employed two parallel CNNs, similar to VGG-16, with one model focusing on capturing micro-level cell details and the other on macro-level tissue structure features. To mitigate spatial structure disruption during integration, they introduce the feature map spatially-constrained integration block, which consists of two main operations: central cropping and feature map resizing. Central cropping was employed to select the region on low-magnification patches corresponding to the region on high-magnification patches. Feature map resizing utilized bilinear interpolation to upsample the patches to match the size of high-magnification feature maps. Additionally, an ASPP block was incorporated to enable the low-magnification model to extract features from a broader perspective. Compared to the basic single-scale CNN and the conventional multi-scale CNN, the proposed model exhibited superior performance, achieving a sensitivity of 0.96 on images at 20×_5× magnification and an average IOU of 0.89 on images at 5×_2.5× magnification.

Multi-resolution networks represent another efficient approach, focusing on the extraction of features from patches at different resolutions obtained from WSIs. In this scenario, patches exhibit varying resolutions while maintaining identical spatial structures, enabling the capture of both local and global features through the receptive fields of CNNs. For instance, Yan et al. (2021) proposed the Hierarchical Attention Guided (HAG) framework. Their approach began with pre-training three distinct input-resolution segmentation networks, utilizing the Linknet (Chaurasia and Culurciello, 2017) architecture with a ResNet-18 backbone, resembling a U-Net model. Different from the straightforward approach of directly upsampling images to match the size of the original image for multi-scale feature fusion, the key innovation in this algorithm lies in the creation of sparse-constrained hierarchical attention maps generated by the macro branch - responsible for global information extraction. These attention maps guided the other two branches in selecting fine-grained local tissue features. The quadtree method (Finkel and Bentley, 1974) was then used in the meso and the micro branches to further pinpoint the significant sub-regions. The final prediction map was obtained through hierarchical fusion. What is noteworthiness is that the images at the macro resolution were directly downsampled from the original image to ensure global information extraction. In contrast, the images of the meso and the micro resolutions were cropped into the patches to facilitate the local information extraction. This model effectively reduced additional computational costs, accelerating inference without sacrificing accuracy, and the authors demonstrated its promising performance on an HCC dataset.

The multi-task ensemble learning-based network. As shown in Fig. 5(b), Wang et al. (2021a) proposed a three-branch CNN model that leverages multi-task learning and ensemble learning techniques to address both segmentation and classification tasks. This model adopted an encoder–decoder architecture, with SE-ResNet-101 (Hu et al., 2018) serving as the chosen encoder. The model features three distinct decoders: the tumor segmentation branch concentrates on the segmentation of tumor regions of HCC WSIs but ignores normal regions; the whole tissue region segmentation branch takes both tumor and normal regions into account to reduce the number of false positives; the classification branch undertakes an auxiliary binary classification task. All three decoders share the same encoder, with the first two decoders sharing five decoding layers. Additionally, they applied the deep supervision strategy (Xie and Tu, 2015) to enhance the representation ability of their model. In their ensemble strategy, they respectively embedded Selective Kernel Modules (SKM) from SKNet and Spatial and Channel-wise Squeeze-and-Excitation Modules (scSEMs) (Roy et al., 2018) into their base model and trained these two models separately. They trained these two models independently, computed their respective outcomes, and averaged them to produce the final segmentation probability map. The binary segmentation mask was generated using an appropriate threshold. Ablation experiments demonstrated that this hybrid network employing the multi-task ensemble learning strategy, significantly improved performance compared to single models. It also outperformed

Table 4
Overview of deep learning-based localization models for liver cancer histopathology.

Type	Reference	Application	Pre-processing	Dataset	Result	Methods, contributions, and limitations
Supervised learning						
	Lal et al. (2021)	HCC nuclei segmentation	–	KMC	F1 = 0.8359, Jaccard = 0.7206	(1) Proposing a CNN integrated with attention mechanism, an encoder and decoder structure; (2) Providing a novel dataset of annotated liver nuclei with 80 H&E stained liver cancer histopathology images; (3) Without pre-processing, the performance of DL models may be influenced by the quality of images like color inconsistency.
	Feng et al. (2021)	HCC segmentation	Partial color normalization + Data augmentation using geometric transformation	PAIP 2019	Jaccard = 0.7964	(1) Utilizing U-Net with the Gaussian pyramid representation and a voting mechanism to fuse multi-scale feature; (2) Providing a novel pre-processing method, partial color normalization for normalizing tissue regions and disregarding blank regions; (3) Applying shifted cropping and weighted overlapping to overcome the discontinuous issue of blocks; (4) Requiring further validating their DL model on other large-scale datasets.
	Chen et al. (2021)	MVI detection	Patch extraction using a sliding window	Private dataset: 190 WSIs	None weighted: Re = 0.512, Pr = 0.767, F1 = 0.614; area weighted: Re = 0.616, Pr = 0.825, F1 = 0.705	(1) Introducing a three-stage cascaded network that comprises a U-Net for cancer tissue segmentation, handcrafted features combined with SE-ResNet for region feature extraction, and a GCNN for region classification; (2) However, the training process still needs many manual operations. (3) False positive predictions are relatively severe.
	Yan et al. (2021)	HCC segmentation	Patch extraction using a sliding window + Background filtering	Private dataset: 90 WSIs	Acc = 0.888, Dice = 0.903	(1) Adopting Linknet with ResNet-18 as the backbone and utilizing attention mechanism to effectively fuse multi-scale features (2) Using the quadtree-based method to select patches; (3) But the scales are empirically selected.
	Zhu et al. (2022)	Liver cancer segmentation	Patch extraction using a sliding window	Private dataset: 56672 pathology image patches	Dice = 0.863, IOU = 0.852, Acc = 0.981	(1) Proposing a CNN with an encoder–decoder structure, where ResNet is used as an encoder; (2) Adopting dilated convolution module to extract multi-scale features; (3) Only focusing on the performance of their DL model in the in-house dataset and lacking external validation.
	Zhai et al. (2022)	Liver cancer segmentation	Patch extraction	Private dataset: 95 WSIs	Acc = 0.981, IOU = 0.865, Dice = 0.845	(1) Utilizing U-Net as a backbone; (2) Introducing a novel global attention mechanism that merges the pyramid sampling strategy with the global attention module, aiming to integrate contextual information and local detailed features. (3) Lacking the validation of their DL model on public datasets.
	Wang et al. (2021a)	HCC viable tumor area segmentation	Tissue extraction using Otsu method + Patch extraction using a sliding window + Data augmentation using geometric transformation and color transformation	PAIP 2019	Jaccard = 0.797	(1) Proposing a three-branch CNN building upon multi-task learning for both segmentation and classification; (2) Adopting a novel ensemble learning strategy to integrate their predictions; (3) Experiencing limitations with a high model size and prolonged computation time.
	Huang et al. (2019)	HCC detection	Patch extraction with a sliding window	Private dataset: 79 WSIs (48 HCC-positive and 31 HCC-negative cases)	mIOU = 0.868, Sens = 0.985	(1) Utilizing two parallel backbones to extract features under two magnifications, respectively adopting VGG-16 and Inception V4; (2) Requiring further improvement by training the DL model with more tumor morphologies.
	Yang et al. (2022)	HCC detection	Patch extraction	Private dataset: 46 HCC-positive WSIs and 1 HCC-negative WSI	5x_2.5x magnification: Sens = 0.94, IOU = 0.89	(1) Proposing a multi-magnification CNN for extracting features under two magnifications; (2) Proposing a feature map spatially-constrained block for the integration of features of low-magnification and high-magnification image patches on the same location of WSIs; (3) The false-positive prediction is slightly high due to the confusion of analogous tissue morphologies.

(*continued on next page*)

other baseline models in pixel-wise segmentation of HCC viable regions on the PAIP dataset, achieving the Jaccard score of 0.797.

The three-stage cascaded network. Chen et al. (2021) introduced a three-stage cascaded network to detect Microvascular Invasion (MVI),

Table 4 (*continued*).

	Qu et al. (2023)	HCC multi-class classification and segmentation	Patch extraction	Private dataset: 495 WSIs from 380 patients	Macro-average, micro-average AUC = 0.97	(1) Employing ResNet-50 with squeeze-and-excitation module for multiclass classification; (2) Requiring further validating the DL model on other datasets from multiple institutions.
	Liang et al. (2023)	HCC multi-class classification and segmentation	Tissue extraction using Otsu method + Patch extraction	TCGA-LIHC: 342 WSIs from 330 patients; PAIP 2019: 100 WSIs; private dataset: 1182 WSIs from 83 patients	Private dataset: Acc = 0.948, AUC = 0.998; TCGA-LIHC: Acc = 0.956, AUC = 0.9984; 2019 PAIP: Acc = 0.941, AUC = 0.9974	(1) Proposing PaSegNet building upon pre-trained ResNet-50 for multiclass classification; (2) Providing a novel labeling method, called mate-annotation dataset; (3) Requiring further validating the DL model on other large-scale datasets.
Weakly supervised learning						
	Diao et al. (2022)	HCC detection	Patch extraction + Noisy images added during the generation of patches + Data augmentation using geometric transformation	TCGA-LIHC: 100 WSIs	Classification: Acc = 0.921, AUC = 0.920, Sen = 0.933, Spec = 0.906; localization: Dice = 0.861	(1) Utilizing two distinct VGG-19 networks as backbones for the extraction of multi-resolution features; (2) Incorporating an attention mechanism to efficiently fuse features; (3) Requiring further advancement by training the DL model with multiple tumor types.
	Hägele et al. (2023)	HCC and ICC semantic segmentation	Patch extraction + Data augmentation using geometric transformation + Stain normalization using color space augmentations	Private dataset: 262 patients (124 ICC and 138 HCC)	Case-level: Acc = 0.91	(1) U-Net with a ResNet-18 backbone trained with a set of weak complementary labels; (2) Proposing a novel complementary loss; (3) Seeking further improvement through the learning of morphologies associated with other rare subtypes.

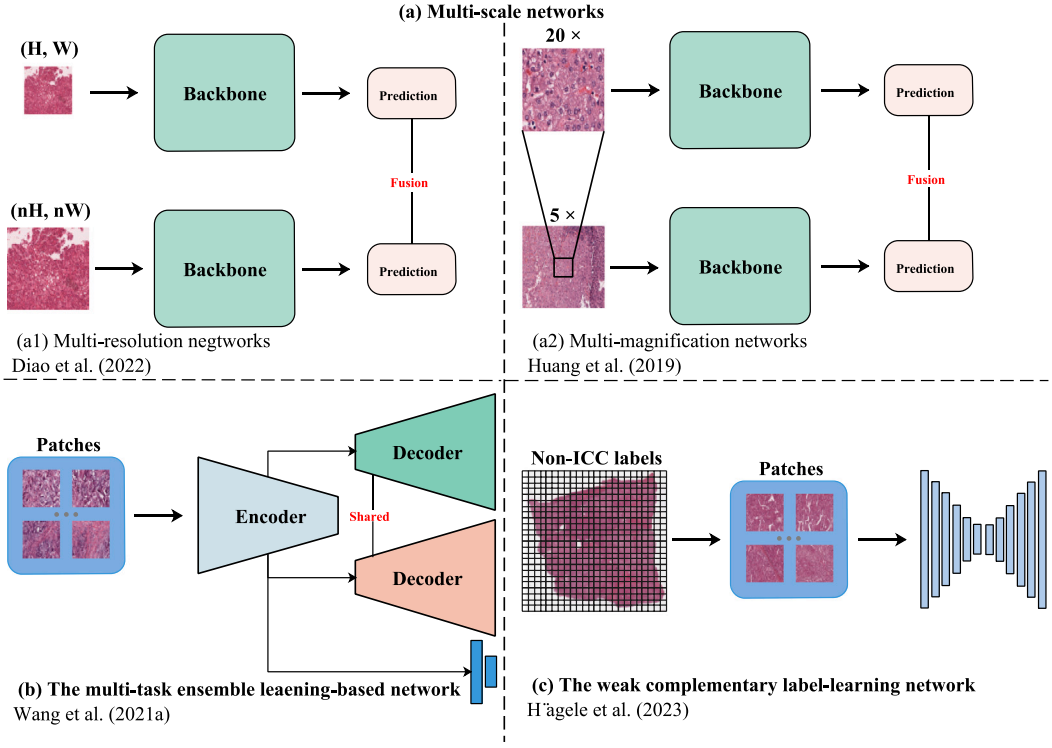

Fig. 5. The representative liver cancer histopathology localization models. (a) demonstrates two distinct multi-scale approaches: (a1) employs two different input-resolution image patches to train two backbones, extracting global and local features through the receptive fields of CNNs. (a2) utilizes two input-magnification image patches of the same size to train two backbones and subsequently fuses these features. (b) represents a multi-task ensemble learning-based network. In this approach, three branches share the same encoder, while two decoders share five decoding layers. Additionally, the auxiliary classification branch remains fixed. (c) illustrates the weak complementary label-learning network. This model takes a set of image patches with weak complementary image-level labels, namely the opposite diagnosis. Note: (a2) and (b) are categorized as supervised learning approaches, while (a1) and (c) fall under the category of weakly supervised learning approaches.

a common histological feature in HCC. In the first stage, the authors employed U-Net, with a notable modification being the replacement of the encoder with SENet, to implement cancer tissue segmentation. It is worth noting that they combined MVI, cancer tissue, and uncertain tissue types as diagnosed by pathologists to train this model in this stage because of the shared characteristics among the tumor cells in these three tissue types. In the second stage, the connected component algorithm was used to generate tissue instances, which included cancer tissue, MVI, and normal tissue instances. Each instance was assigned a specific label. Additionally, the authors manually extracted morphological features from these tissue regions, encompassing properties such as area, perimeter, perimeter-to-radius ratio, and more. SENet was trained using both the segmentation masks of tissue instances and the corresponding raw images. The final prediction outcome from SENet

was concatenated with six hand-crafted features, and this combined feature set was utilized for the classification of MVI. In the third stage, the authors introduced modifications to the Graph Convolutional Neural Network (GCNN). Specifically, they incorporated a shortcut connection to alleviate the over-smoothing problem and added a MLP at the network's end. This stage aimed to generate image-level results, where the vertices of the graph corresponded to the instances generated in the previous stage.

Multi-class segmentation networks. Existing DL models in the realm of liver cancer HIA have predominantly been on binary segmentation, distinguishing between cancerous and normal regions. However, this approach has certain limitations as it constrains the diversity of features that these models can learn. In contrast, multi-class segmentation networks have emerged as a promising avenue to address these limitations. These networks are capable of learning a wider array of histological features, thereby not only bolstering their generalization but also furnishing invaluable pathological insights for prognostic analysis. For example, Qu et al. (2023) harnessed the power of ResNet-50, augmented with a squeeze-and-excitation module, to undertake the task of classifying HCC WSIs into six distinct categories. The generated segmentation maps enable pathologists to visually discern and analyze the targeted tissue regions. Similarly, Liang et al. (2023) developed a high-generalization CNN named PaSegNet. Notably, this model trained on the high-quality and high-diversity mate-annotation dataset (for further elaboration, refer to Section 6.2. PaSegNet exhibited exceptional performance across three diverse datasets, encompassing TCGA-LIHC, PAIP, and a private dataset. Importantly, this study underscored that even when a substantial-scale dataset is lacking, models trained on high-quality, diverse small-scale datasets can still exhibit remarkable robustness and generalization capabilities.

4.3.2. Weakly supervised learning

Supervised learning localization models heavily depend on an extensive dataset comprising high-quality annotated images, often requiring pixel-level labels. Nonetheless, the acquisition of such labeled images at scale presents formidable challenges. In response to this issue, several studies have explored the application of weakly supervised learning-based approaches. In the subsequent sections, we will specifically highlight two prominent approaches: the multi-scale attention network and the weak complementary label-learning network.

The multi-scale attention network. Diao et al. (2022) exploited a weakly-supervised multi-scale attention model, comprising two parallel VGG-19 networks as the backbone. This model operated efficiently with a set of weak labels, namely image-level labels, while patch labels corresponded to the labels of large tissue regions extracted from WSIs. As illustrated in Fig. 5(a1), this model can be treated as a multi-resolution model due to the distinct resolutions of the two input patches, yet maintaining the same spatial structure. The fusion mechanism employed in this model is the attention mechanism. The attention weights were generated by the low-resolution sub-model. The feature maps from the high-resolution sub-model were downsampled to match the dimensions of the low-resolution feature maps and were then concatenated with the attention maps to yield the model's output. Combining the outcomes from the two sub-models resulted in the final classification output. Experimental results demonstrated that this model achieved remarkable performance in HCC tumor region detection, with a Dice score of 0.807.

The weak complementary label-learning network. As illustrated in Fig. 5(c), Hägele et al. (2023) discovered that the additional diagnostic outcome can lead to improved performance of DL models. They utilized weak complementary labels to train their DL model for segmenting HCC and ICC histopathological images. Weak complementary labels, in this context, refer to labels that represent the opposite diagnosis of histopathological images, such as labeling HCC histopathological images as non-ICC, and vice versa. To guide the DL model to closely approximate the ground truth classes, they introduced a novel complementary loss function. The overall loss of their model was formulated as a weighted combination of the cross-entropy loss and the complementary loss. Furthermore, they incorporated an extended focal loss (Lin et al., 2017) by introducing a multiplicative factor into the loss calculation. Their approach involved training a U-Net architecture with a ResNet-18 backbone using a subset of pixel-level labeled images alongside other complementary labeled images. The experimental results showcased the strong performance of the model trained on complementary labeled images, achieving an Acc of 0.905 in case-level prediction.

4.4. Discussion

In the realm of classification tasks pertaining to liver cancer histopathology, the utilization of CNNs has witnessed significant development. Initially, conventional CNNs demonstrated promising performance in early studies. However, owing to the inherent limitations in representation learning of these traditional CNNs, subsequent research endeavors have increasingly turned to more advanced and deeper models, including ResNet, VGG, and DenseNet. To ease the challenge of limited training data, the adoption of transfer learning has emerged as a viable solution. In this approach, models pre-trained on large-scale datasets like ImageNet are repurposed for the target dataset, facilitating knowledge transfer. Another effective approach to mitigate the lack of high-quality labeled data is weakly supervised learning. These MIL models require only a set of coarse labels, namely the image-level labels, making them particularly valuable in scenarios with limited annotated data. Furthermore, to augment the classification performance of these models, several studies have introduced modifications, such as the incorporation of attention mechanisms.

Among localization tasks, the main objective revolves around the precise identification of lesion regions, often accomplished through the generation of heatmaps or segmentation, along with the classification of the target class. Multi-scale networks have emerged as the prevailing paradigm for addressing these challenges. These networks typically employ two or more branches dedicated to extracting both global and local information from high-magnification and low-magnification images, which are subsequently fused to generate the final output. A widely recognized segmentation network used in this context is U-Net, characterized by its encoder–decoder architecture. However, traditional U-Net models have relatively shallow layers, prompting several research efforts to enhance feature extraction capabilities. One common approach involves replacing the U-Net's encoder with deeper networks like SE-ResNet, resulting in improved feature representation. Moreover, it is worth noting that the landscape of localization networks has seen the incorporation of various innovative strategies. These include multi-task ensemble learning-based networks, three-stage cascade networks, and notable weakly supervised networks, each contributing to the diversity of approaches employed in liver cancer histopathology studies.

While existing models have demonstrated commendable performance, they are not without their limitations. The foremost challenge is the insufficiency of training datasets, particularly those with high-quality labels, such as patch-level or pixel-level annotations. Furthermore, these datasets are often derived from single-source origins, which imposes constraints on the advancement of DL models. Despite attempts to address this limitation through various methods, including the utilization of learning-efficient algorithms like transfer learning or weakly supervised learning, as well as data augmentation techniques aimed at expanding the pool of available images, the resulting improvements in model performance have been somewhat limited (Wang et al., 2021c). Consequently, there is a pressing need to amass a more extensive collection of histopathological images sourced from multiple institutions to bolster the training of DL models in future endeavors. The second limitation involves the lack of interpretability. The majority of existing DL models employed in liver cancer histopathology are

often perceived as the "black box", rendering the underlying decision-making processes inscrutable, especially for pathologists. Bridging this gap is essential for the eventual integration of DL models into clinical diagnostic workflows. Therefore, the development of interpretable, human-centered Artificial Intelligence (AI) models holds promise for facilitating meaningful interactions between clinical practitioners and automated systems.

5. Deep survival learning for prognosis

5.1. Difference with diagnostic models

In Section 4, we provided an extensive overview of representative DL models that have been applied in routine liver cancer histopathology research. These models are mainly designed to extract and learn histological features directly from histopathological images, enabling them to predict specific disease types, such as HCC and ICC. Consequently, they can be categorized as clinical diagnostic models (Li et al., 2022). Prognosis entails the prediction of the likelihood of a specific event occurring, such as patient mortality or disease recurrence, within a predefined timeframe. This prediction is made for patients who have undergone a period of treatment, such as liver transplantation or resection, subsequent to their initial diagnosis. While both diagnostic and prognostic models share certain similarities, they also exhibit notable distinctions. One commonality lies in the dichotomized output produced by these two types of models. As such, survival models can effectively be regarded as a classification problem. Nonetheless, the data required for training these models differs significantly. Diagnostic models typically rely on a dataset composed only of image data training. In contrast, the development of prognostic models necessitates the integration of survival data, such as time-to-event information, with histopathological images to create a training dataset.

5.2. Methodologies

In recent years, there has been a notable surge in the utilization of DL survival models within the domain of liver cancer histopathology, primarily for prognostic analysis and the identification of significant biomarkers. This section provides an in-depth exploration of select DL models that directly harness histological features derived from histopathological images to facilitate liver cancer prognosis. One pivotal aspect of survival models in this context is risk stratification, which involves the assignment of risk scores and the determination of optimal cutoff values. Finally, three distinct approaches have emerged as main avenues for conducting prognostic analysis in liver cancer, namely: deep learning integrated with Cox survival models, two-stage deep prognostic networks, and the end-to-end deep prognostic network. An overview of these representative approaches is shown in Table 5.

Deep learning integrated with Cox survival models. This approach is centered on the utilization of DL models for categorizing patches into distinct groups, with each group characterized by specific histologic patterns. These groups serve as covariates within the framework of the Cox proportional hazards model, a widely employed approach for modeling individual survival outcomes. For example, Muhammad et al. (2019) proposed an unsupervised convolutional autoencoder-based model to subtype ICC. Their model adeptly classified WSIs into five distinct clusters, drawing upon cellular and structural morphologies. Subsequently, these five clusters were leveraged as covariates for training five separate univariate Cox models, facilitating the examination of survival correlations among these diverse histological subtypes.

In contrast to unsupervised techniques that seek to discern unknown features within WSIs to form distinct clusters, weakly supervised models in this context are tailored to identify predefined histology features with direct relevance to survival outcomes. For example, Qu et al. (2022) used weakly supervised Inception-V3, trained on multi-class annotation WSIs, to classify the patches into six categories, including tumor regions, normal liver regions, and portal areas, among others. The resulting classification maps underwent optimization through morphological processing. Pathological signatures were subsequently extracted from the top ten patches with the highest probability within each category, forming a matrix that served as input for a Least Absolute Shrinkage and Selection Operator (LASSO) Cox regression model (Tibshirani, 1997). The histological score based on pure histological features, and the combined score taking together other clinical markers were the outcome for further stratification.

Two-stage deep prognostic networks. Compared with the aforementioned approach, DL-based survival models exhibit enhanced capabilities for handling complex, non-linear data (Katzman et al., 2018). Within this approach, the initial DL model is typically used as a diagnostic model, with the main objective of discerning histopathological images and classifying them into specific target categories. Subsequently, the second DL prognostic model leverages the prior knowledge derived from the initial diagnostic model to capture relational histological features, independently of other survival-related factors. Consequently, the risk score generated by DL prognostic models relies exclusively on pure histopathological information.

The DeepSurv network (Katzman et al., 2018) serves as a typical Cox-based DL survival model. A recent study proposed by Qu et al. (2023) modified the DeepSurv for analyzing pathological signatures pertaining to six distinct tissue categories associated with HCC. This adaptation yielded state-of-the-art prognostic performance. Other studies (Liang et al., 2023; Yamashita et al., 2021) employed a loss function similar to the DeepSurv loss, namely the negative Cox partial log-likelihood, for their DL prognostic models. These models also yielded considerable prognostic performance.

Differing from the conventional approach of training a DL prognostic network using individual patches, Liang et al. (2023) adopted an innovative strategy. They harnessed an eight-dimensional macro mode derived from the entire WSI generated through the prior classification network. These feature vectors served as input for a prognostic network, implemented using ResNet-50 alongside MLP with batch normalization layer (Xie et al., 2017), named MacroNet. This model exhibited exceptional performance in the context of survival regression for HCC patients, surpassing the prognostic network trained on a set of individual patches. The core innovation of their prognostic network is the feature extraction of global spatial distributions of eight tissue categories of HCC, and the authors assumed that the spatial distribution of a certain tissue type in WSIs can be correlated with the patient's survival.

Limited interpretability presents a fundamental challenge in the domain of DL prognostic networks. To address this issue, several studies have endeavored to demystify the "black box" nature of these networks, thereby augmenting the reliability of clinical decision-making. For instance, Saillard et al. (2020) devised an interpretable-by-design prognostic network capable of identifying pertinent tumor invasion features associated with patient survival in the context of HCC histopathological images. More specifically, expert pathologists analyzed the 400 most predictive patches produced by the DL prognostic model, which comprised 200 high-risk patches and 200 low-risk patches. Their analysis revealed that the presence of vascular spaces, a macro trabecular architectural pattern, a high degree of cytological atypia, and nuclear hyperchromasia within tumor areas exhibited significant associations with patient survival. In another study by Liang et al. (2023), an attribution method was adopted to bolster the interpretability of their prognostic network. This approach generated saliency maps (Simonyan et al., 2013) that were subsequently overlaid with segmentation maps. The findings from this study highlighted the spatial distribution of necrosis and the fraction of necrotic tissue significantly correlated with patient survival. These investigations demonstrated that DL prognostic networks possess the capacity to predict both known and underlying morphological features that are closely tied to patient survival outcomes.

The end-to-end deep prognostic network. Building upon the principles of End-to-end Part Learning (EPL) introduced by Xie et al.

Table 5
Overview of deep learning-based survival models for liver cancer prognosis.

Type	Reference	Application	Pre-processing	Dataset	Result	Methods, contributions, and limitations
Deep learning integrated with Cox survival models						
	Muhammad et al. (2019)	Evaluating histological clusters for survival prediction of ICC	Tissue extraction using Otsu method + Patch extraction	Private dataset: 246 cases	–	(1) Proposing an unsupervised convolutional autoencoder for ICC subtype clustering based on similar morphologies; (2) Utilizing univariate Cox survival models to assess the performance of this DL model; (3) Lacking the detailed comparative analysis with other approaches and needing further validation on a much broader range of patients.
	Qu et al. (2022)	Dichotomized recurrence-related histological score of early-stage HCC and the correlation between immune microenvironment and histological score	Tissue extraction using Otsu method + Patch extraction + Data augmentation using geometric transformation + Stain standardization by modifying the traditional Reinhard algorithm	Private dataset: 416 WSIs from 387 patients; TCGA-LIHC: 154 WSIs from 147 patients	Private dataset: C-Index = 0.739; TCGA-LIHC: C-Index = 0.708	(1) Employing Inception-V3 to implement multiclass classification; (2) Applying LASSO Cox regression for the generalization of histological score (3) Using CAM to visualize the importance of the local regions; (4) Requiring further validating this prognostic network on the external validation dataset.
	Hou et al. (2022)	Integrative histology-genomic multi-modality information for survival prediction of HCC	Patch extraction + Data augmentation using geometric transformation	TCGA-LIHC: 346 cases	C-Index = 0.746	(1) Utilizing pre-trained VGG-19 and K-means to cluster patches; (2) Adopting MIL based on generalized Siamese architecture to analyze the patches from each category; (3) Using Weighted Gene Co-expression Network Analysis (WGCNA) and LASSO for the acquisition of hub genes (4) Employing Cox survival model receiving the risk score and hub genes for survival analysis; (5) Lacking analysis between gene and phenotype.
Two-stage deep prognostic networks						
	Saillard et al. (2020)	Dichotomized survival-related risk score for HCC	Tissue extraction using U-Net + Patch extraction	Private dataset: 390 WSIs from 206 tumors; TCGA-LIHC: 342 WSIs from 328 patients	Private dataset: C-index = 0.78; TCGA-LIHC: C-Index = 0.70	(1) Employing pre-trained ResNet-50 used as feature extractor; (2) Presenting CHOWDER, a model designed for weakly supervised learning with inherent interpretability, and SCHMOWDER, which incorporates a supervised branch featuring an attention mechanism alongside a weakly supervised branch for prognostic analysis; (3) Requiring further improving inherent interpretability of this prognostic network.
	Yamashita et al. (2021)	Dichotomized recurrence-related risk score for HCC	Patch extraction + Data augmentation using geometric transformation	Private dataset: 198 WSIs; TCGA-LIHC	Private dataset: C-index = 0.683; TCGA-LIHC: C-Index = 0.724	(1) Utilizing PathCNN to differentiate between patches classified as cancerous or normal; (2) Incorporating a pre-trained MobileNet-V2 as a prognostic network to process the top 100 tumor patches with the highest predicted probabilities, subsequently generating the risk score; (3) Requiring additional validation on diverse, large-scale datasets for comprehensive verification. (4) Absence of interpretability, hindering the ability to gain insights into their DL model.
	Liu et al. (2022)	Dichotomized recurrence-related risk score for HCC	Patch extraction + Stain normalization	Private dataset: 120 nucleus samples + 552 WSIs (surgical resection) + 144 WSIs (liver transplantation); TCGA-LIHC: 302 cases	Varying	(1) Adopting U-Net to capture the nuclear architecture; (2) Employing MobileNet-V2 based on MIL as a prognostic network to process the four-channel images, which include segmentation masks combined with raw images; (3) Lacking interpretability of their DL prognostic model.

(*continued on next page*)

(2020), Muhammad et al. (2021) devised a novel end-to-end deep prognostic network referred to as EPIC-Survival, bridging the gap of the above-mentioned two-stage deep prognostic networks. In this model, the pre-trained ResNet-50 was used as the feature extractor. Subsequently, the generated feature vectors underwent an assignment process into several logically constructed histology feature clusters based on initial global centroids. These feature vectors were subjected to iterative reassignment, guided by local slide-level centroids. Consequently, the patches from each cluster that were closest to the specific local centroid were designated as part representations of the WSI. These segmented parts of the WSI were concatenated and jointly trained with survival data, ultimately yielding the final risk score for the WSI. Additionally, the authors presented a stratification loss to enhance the model's capability for risk grouping. Experimental results underscored the substantial performance improvement achieved by incorporating the stratification loss, particularly in the prognosis of ICC patients.

5.3. Discussion

DL models play a pivotal role in the domain of liver cancer survival analysis, undergoing rapid development. In the context of integrating

Table 5 (*continued*).

	Shi et al. (2021)	Dichotomized recurrence- and survival-related risk score for HCC	Tissue extraction using Otsu method + Patch extraction + Stain standardization using CycleGAN	Private dataset: median C-Index = 0.731; TCGA-LIHC: median C-Index = 0.713	Private dataset: 2191 WSIs from 1029 patients; TCGA-LIHC: 320 WSIs from 320 patients	(1) Utilizing a classification network, specifically the Neural Conditional Random Field (NCRF) built upon ResNet-18 architecture for multi-class classification; (2) Integrating Conditional Random Field (CRF) to model the spatial correlation of patches; (3) Implementing a prognostic network utilizing a multi-scale architecture based on ResNet-50; (4) Adapted the CAM to create a Risk Activation Mapping (RAM) for displaying heatmaps on the patches; (5) Requiring further validating this prognostic network on other datasets.
	Liang et al. (2023)	Dichotomized recurrence- and survival-related risk score and novel biomarkers exploration for HCC	Patch extraction	PAIP 2019; TCGA-LIHC: 342 WSIs from 330 patients; private dataset: 1182 WSIs from 83 patients	TCGA-LIHC: C-Index = 0.708, Survival AUC = 0.732; private dataset: C-Index = 0.754, Survival AUC = 0.796	(1) Proposing PathFinder by employing the macro mode of WSIs to achieve high-precision prognosis and identifies the spatial distribution of necrosis as a significant indicator for clinical prognosis. (2) Providing a novel labeling method, a mate-annotation dataset; (3) Utilizing a classification network named PaSegNet, which is based on ResNet-50, to acquire the macro mode of WSIs; (4) Implementing a prognostic network named MacroNet, built upon ResNet-50 architecture, followed by an MLP with a batch normalization layer. This configuration is designed for learning the macro mode of WSIs and generating the risk score. (5) Utilizing attribution methods that involve saliency maps to verify hypotheses in the discovery of new biomarkers; (6) The identification of novel biomarkers necessitates the formulation of empirical hypotheses for pathologists.
	Qu et al. (2023)	Dichotomized recurrence-related deep pathomics score and immune score for HCC	Tissue extraction using Otsu method + Patch extraction + Data augmentation using geometric transformation	Private dataset: 380 cases	Deep pathomics score: C-Index = 0.794; immune score: C-Index = 0.768	(1) Utilizing ResNet-50 with a squeeze-and-excitation module as classification network for implementing multiclass classification; (2) Modifying DeepSurv network as prognostic network for analyzing the pathological signatures of six tissue categories; (3) Incorporating an attention mechanism to focus on critical regions associated with the model prediction; (4) Future endeavors should be focused on the correlation between pathological signatures and multi-omics sequencing data.
The end-to-end deep prognostic network						
	Muhammad et al. (2021)	Boosting risk stratification for survival and recurrence prediction of ICC	Patch extraction	Private dataset: 265 WSIs	C-Index = 0.88	(1) Applying pre-trained ResNet-34 to generate feature vectors of each patch. Assigning patches to specific clusters with respect to the histological characteristic based on initial global centroids. Nearest patches according to local centroids as part representations of the WSI. (2) Introducing a novel stratification loss, integrated with the negative log partial likelihood loss, aimed at improving risk grouping in prognostic analysis; (3) Requiring further validating this DL model on other large-scale datasets.

deep learning with Cox survival models, DL models are frequently utilized as diagnostic tools for distinguishing between cancerous and normal images. However, the inherent nature of these approaches falls short of realizing their full potential in uncovering the underlying characteristics of DL models. DL-based prognostic networks exhibit significant advantages, particularly in handling non-linear data. A notable example is the DeepSurv model, surpassing many traditional survival models. Furthermore, DL-based prognostic models can identify critical underlying characteristics that significantly influence prognostic outcomes, often challenging for human observation. Nevertheless, a major limitation of these methods is their lack of interpretability. While numerous studies have employed diverse approaches to enhance the interpretability of DL prognostic networks, ongoing efforts are essential to improve transparency and dependability until they achieve clinical acceptability. Additionally, two-stage deep prognostic networks commonly employ decoupling training strategies, rendering DL prognostic models insufficiently capable of understanding histological features and their impact on prognostic outcomes. Therefore, an end-to-end deep prognostic network is emerging as the prevailing trend.

6. Open issues and future trends

As elucidated in Sections 4 and 5, DL models have achieved gratifying performance in the domain of liver cancer histopathology, encompassing both diagnostic and prognostic applications. Nevertheless, it is imperative to acknowledge that DL models employed within this domain are still at an incipient stage, beset by a multitude of uncertainties on the horizon. In light of the extant DL approaches, we have deliberated upon various pervasive challenges and delineated viable strategies to address them. By way of outlook, we have proffered several prospective avenues for future development.

6.1. Limited training data and high-quality labeled images

As previously highlighted, the training of a high-performance DL model hinges upon the availability of sufficient high-quality labeled data. Unfortunately, the dearth of such data remains a persistent challenge in the domain of liver cancer histopathology. It is worth noting, however, that TCGA-LIHC can only furnish image-level annotations. The PAIP dataset, while providing pixel-level labeled images, is rather limited in scale, comprising a mere 100 histopathological images, thereby constraining the performance potential of DL models on this specific dataset. Unlike natural images, labeling the histopathological image necessitates the expertise of skilled pathologists, rendering this endeavor arduous and costly. Researchers have continually explored a range of efficacious strategies to mitigate this quandary, encompassing the adoption of learning-efficient algorithms and the introduction of labeling-efficient approaches. Subsequently, we systematically reviewed the existing approaches for alleviating this challenge and explored several feasible countermeasures to overcome this open problem.

Learning-efficient paradigms. Learning-efficient paradigms have assumed prominence in liver cancer HIA. While supervised learning serves as the prevailing approach, it indispensably demands a copious reservoir of high-quality annotated data (Lin et al., 2021). In a concerted effort to alleviate the reliance on high-quality labeled images, recent years have witnessed the proliferation of **weakly supervised learning** methods. Notably, MIL has emerged as the primary approach within the domain of weakly supervised learning, leveraging only a set of coarsely labeled images, namely image-level labels, thereby rendering it eminently suitable for HIA. It is imperative to underscore that, although MIL offers an efficacious solution to mitigate the reliance on high-quality labeled data, the training of a MIL model with high generalization capabilities still necessitates an extensive dataset (Liang et al., 2023; Campanella et al., 2019).

In an endeavor to circumvent the challenges posed by the scarcity of histopathological images, researchers have explored **transfer learning** approaches. This entails the utilization of a DL model that has been pre-trained on ImageNet, subsequently fine-tuning it on the target dataset, such as TCGA-LIHC, to expedite convergence. For instance, Sun et al. (2019) leveraged a ResNet-50 model pre-trained on ImageNet, utilizing the patch-level feature vector as input for the ensuing MIL framework, responsible for selecting the most discriminative patches and culminating in a slide-level outcome through an MLP classifier. This learning strategy exhibited promise when applied to a small-scale dataset derived from TCGA-LIHC. Nevertheless, it is worth noting that most pre-trained models rely on ImageNet, a large-scale dataset comprised of natural images, which may not be ideally suited for precise HIA tasks.

Furthermore, the potential of **unsupervised learning, semi-supervised**, and **self-supervised learning** remains to be further explored. An example of this is the study by Roy et al. (2021) which harnessed an unsupervised convolutional autoencoder featuring a bespoke reconstruction loss function to effect the segmentation of HCC viable tumor regions. This investigation underscores the untapped utility of unsupervised learning techniques in the realm of liver cancer HIA. Additionally, a recent study proposed by Guo et al. (2023) leveraged the self-supervised learning-based approach (Ciga et al., 2022) for the precise segmentation of entire tumor areas in HCC WSIs. This approach harnessed ResNet-18 within the SimCLR framework (Chen et al., 2020a) for contrastive learning. The chosen model served as a feature extractor, complemented by two Transformer encoders that effectively amalgamated both local and global feature information, enabling accurate localization of whole tumor regions within the WSIs. We anticipate that future research endeavors will introduce increasingly learning-efficient paradigms aimed at reducing reliance on high-quality annotations, all while preserving optimal performance levels.

On the other hand, bolstering the generalization capabilities of DL models across diverse datasets necessitates access to a rich tapestry of training data sourced from multiple institutions. Unfortunately, the imperative of safeguarding privacy and upholding data confidentiality precludes many institutions from making their datasets publicly available (Rieke et al., 2020). To surmount this challenge, privacy-preserving learning, notably federated learning, presents a viable solution. This learning paradigm revolves around the sharing of training outcomes alone, as opposed to the raw training data, thereby ensuring that privacy-sensitive data remains under the custodianship of its respective owners, while DL models are trained locally. Several inspiring studies (Lu et al., 2022; Hosseini et al., 2023) harnessed federated learning within the domain of histopathology, albeit primarily in the context of other cancer types. These endeavors have yielded DL models of notable proficiency, delivering remarkable performance on specific tasks within the field.

Labeling-efficient approaches. In addition to learning-efficient paradigms, an alternative strategy for mitigating the constraints imposed by limited training data involves the development of an efficient labeling approach. Recently, Liang et al. (2023) proposed a **meta-annotation approach**, which bridges the gap between pathological annotations and the training requisites of DL models. To diversify the morphological features learned by DL models across various classes (e.g., tumor, normal, fibrosis, steatosis, etc.) while addressing data imbalances, they initially employed rectangular boxes to encompass the target tissue regions. Subsequently, they selectively extracted only 100 patches each from the tumor and normal regions, recognizing the larger proportion of these two tissue types relative to others and aiming to achieve data balance. The remaining tissue classes were derived from full patches. Random sampling techniques were then harnessed to assemble the ultimate balanced meta-annotation training dataset. Although this dataset is relatively modest in size, the DL model trained on it exhibited exceptional performance, benefitting from the high-quality and diverse labeled images. However, it is worth noting that this annotation method still hinges on the expertise of senior pathologists, incurring substantial costs for research endeavors in the field. Hopefully, the future will witness the exploration of more efficient labeling methods.

6.2. Poor generalization and robustness

Presently, the majority of existing DL models have demonstrated noteworthy performance within their specific in-house datasets, with some even surpassing the diagnostic capabilities of certain pathologists within clinical contexts (Kiani et al., 2020; Diao et al., 2022). However, their performance may falter when transitioning to real-world clinical applications. This discrepancy arises from the inherent heterogeneity among histopathological images originating from diverse sources (Marini et al., 2021). Furthermore, WSIs from distinct patients exhibit varying features due to differences in staining procedures and imaging protocols, thus presenting a formidable challenge in HIA for the development of robust and generalized models (Chen et al., 2017b). These challenges bring the hurdle for the improvement of DL models. In the subsequent discussion, we explore several viable strategies to address these challenges, along with advanced methodologies that hold

the potential to advance the development of robust and generalized DL models.

Domain adaptation. The presence of color inconsistencies among datasets collected from multiple centers poses a significant challenge in liver cancer histopathology research. While some studies have integrated stain normalization techniques into their pre-processing pipelines, the resulting improvements in model performance have been limited (Sun et al., 2019). An effective strategy to address this issue is domain adaptation, which involves transferring knowledge from a source dataset to a target dataset with the aim of enhancing the generalization capabilities of DL models. This approach is particularly advantageous in alleviating the inherent heterogeneity in the feature distribution of WSIs across datasets originating from various institutions. Although the application of domain adaptation in liver cancer histopathology studies is still relatively unexplored, it holds substantial promise. Several noteworthy studies (Ren et al., 2018; Alirezazadeh et al., 2018; Ren et al., 2019) have paved the way for the utilization of domain adaptation techniques in the field of histopathology.

Multi-modal feature fusion. The integration of complementary information with histopathological images for training DL models holds the potential to enhance clinical diagnostic and prognostic capabilities. This strategy can be effectively realized through multi-modal approaches, which have found application in the field of histopathology. For example, Hou et al. (2022) combined liver cancer histopathological images with mRNA expression data to predict patient survival. They utilized deep CNNs to generate risk scores for each WSI, followed by the application of Weighted Gene Co-expression Network Analysis (WGCNA) and LASSO for the identification and selection of hub genes. Subsequently, the multi-modal fusion features were incorporated into a Cox survival model. In the context of liver cancer histopathology studies, there remains a need for further exploration and development of more effective multi-modal approaches.

Foundation models. Recently, a new wave of research has been sparked by foundation models such as ChatGPT, CLIP (Radford et al., 2021), and SAM (Kirillov et al., 2023), garnering increased attention in various domains. Among liver cancer HIA, there is a need for further exploration of the capabilities of these foundation models. For example, CLIP, an innovative zero-learning algorithm, holds substantial promise for image classification, and its potential application in liver cancer HIA warrants investigation. Furthermore, a study conducted by Huang et al. (2023) has highlighted the remarkable performance of SAM in diverse medical image segmentation tasks, including histopathology. More recently, Ma and Wang (2023) introduced MedSAM, a universal foundation model tailored for medical image segmentation, which demonstrated superior performance across multiple medical image domains when compared to SAM. Inspired by these two versatile segmentation foundation models, future research endeavors can delve into their application in the realm of liver cancer WSI segmentation. Additionally, GMAI (Moor et al., 2023) a foundation model specifically designed for medical applications, presents a notable development. GMAI possesses the capability to seamlessly integrate multi-modal features encompassing medical images, textual data, experimental reports, and more, enabling the generation of accurate clinical decisions. As a prospective avenue, foundation models hold tremendous potential, and future investigations can explore their integration into liver cancer HIA, ushering in new horizons for this field.

Adversarial attacks. Adversarial attacks with imperceptible perturbations pose a significant concern for DL models in histopathology, as they can lead to erroneous diagnoses. This issue not only raises potential safety hazards but also hinders the deployment of DL models in clinical settings. Ma et al. (2021) has documented instances of adversarial attacks in various medical scenarios, including Chest X-ray, Fundoscopy, and Dermoscopy, highlighting the vulnerability of DL models in medical image analysis. Interestingly, these adversarial attacks were found to be more easily detectable in medical images than in natural images. While the presence of adversarial attacks in the domain of liver cancer histopathology remains relatively unexplored, it represents an area where misleading features, such as those related to cancer sub-types, could be crafted to deceive and evaluate the robustness of DL models. Beyond the detection of adversarial examples, several defense methods to prevent adversarial attacks, such as adversarial training (Wang and Zhang, 2019; Wang et al., 2021b) and the regularization method (Ross and Doshi-Velez, 2018), may be also developed for liver cancer histopathology. These advancements in adversarial attack detection and defense methods are vital for ensuring the reliability and safety of DL-based diagnostic tools in clinical practice.

6.3. Lack of interpretability

A pivotal challenge inherent to DL models lies in their characterization as the "black box" (Van der Velden et al., 2022). This concern holds particular significance within the prognostic models, where the internal decision-making process remains shrouded in obscurity from the perspective of pathologists. Consequently, instilling trust among clinical practitioners regarding the relevance of predicted outcomes to patient survival proves to be a formidable task. It is worth noting that this issue may bear less weight in the context of clinical diagnostic models, as pathologists possess the visual means to inspect lesion regions and independently arrive at conclusive results. Within the extensive literature we have surveyed, a spectrum of viable countermeasures has been explored to address this challenge. These countermeasures predominantly pertain to the deployment of visualization techniques, intended to bolster the interpretability of DL models. Additionally, prospective avenues for enhancing interpretability through inspiring approaches have been postulated and are elucidated as follows.

Class activation maps. The concept behind CAMs revolves around the visualization of feature maps derived from the last convolutional layer. This visualization technique allows for the observation of CNNs as they focus on discerning regions that hold greater discriminative value for specific target classes. For instance, Kiani et al. (2020) incorporated CAM into their DL diagnostic model, assisting pathologists in making conclusive clinical decisions. Furthermore, in the context of prognostic models, Shi et al. (2021) introduced a modification of CAMs known as Risk Activation Mapping (RAM) to accentuate histological ROIs linked to patients' survival. However, the drawback of CAM is that there is a need to modify the structure of CNNs and to retrain the CNN when using it. To address these limitations, more efficient CAM techniques such as Grad-CAM (Selvaraju et al., 2017) and Grad-CAM++ (Chattopadhay et al., 2018) hold potential for further exploration.

Attribution methods. The implementation of attribution methods is various. As a reference, Deng et al. (2023) have systematically categorized 14 attribution methods, encompassing gradient-based, back-propagation, and perturbation-based techniques. For instance, Liang et al. (2023) employed saliency maps generated by computing gradients of the loss function for risk score with respect to input pixels. They overlaid these maps with segmentation maps, allowing them to validate hypotheses and identify new biomarkers associated with the prognosis of HCC patients based on histopathological images. Despite the existence of numerous attribution methods offering diverse perspectives, there remains a lack of a unified conceptual framework to guide their application. Within the domain of liver cancer histopathology research, there is a need for further exploration and investigation into various attribution maps and their potential contributions to the field.

Non-attribution methods. Attribution methods traditionally center on calculating the confidence score for each pixel as a means of explaining image classification, providing localized insights into model decision-making. However, these methods tend to operate at low-level conceptual levels that are less intuitive for human comprehension. In contrast, the Testing with Concept Activation Vectors (TCAV) approach,

introduced by Kim et al. (2018), offers a distinct advantage by providing high-level concepts that facilitate a deeper understanding of DL models from a human perspective. TCAV generates a global explanation of DL models using concepts that are readily interpretable by humans. For example, in the context of liver cancer histopathology, TCAV can leverage Concept Activation Vectors (CAVs) to establish concepts related to specific morphological characteristics associated with the survival of liver cancer patients. It then quantifies the sensitivity of these concepts within the DL model. In future research within the field of liver cancer histopathology, TCAV has the potential to significantly enhance model transparency and enable precise quantification of the contributions made by particular histological features.

6.4. Cancer subtyping and multi-class classification

Liver cancer histopathology classification predominantly relies on DL models to differentiate image patches or WSIs as benign or malignant. However, to enhance the precision of treatment planning by assisting pathologists in distinguishing the degree of malignancy, it is imperative to extract more refined and fine-grained features. For instance, liver cancer can be categorized into three distinct malignant degrees: poorly-differentiated, moderately-differentiated, and highly-differentiated, where high differentiation closely resembles normal tissue cells and low differentiation approaches malignancy with a higher degree of severity (Chen et al., 2022a; Dong et al., 2022). Additionally, Aatresh et al. (2021) have proposed a DL histological grading system for HCC, classifying it into three sub-types: low sub-type, medium sub-type, and high sub-type, in accordance with the Edmondson and Steiner's standard (Edmondson and Steiner, 1954). While these studies have initiated the exploration of cancer subtyping, there remains ample room for further investigation in this domain. Future research efforts can delve into various aspects of cancer subtyping, including the intra-specific diversity within early and advanced cancer subtypes, as well as the differentiation of HCC or ICC from combined hepatocellular-cholangiocarcinoma WSIs within the context of liver cancer histopathology. Fascinatingly, a recent study by Calderaro et al. (2023) utilized a self-supervised DL model to reclassify combined hepatocellular-cholangiocarcinoma samples into HCC and ICC, achieving remarkable performance.

Additionally, the trajectory of future research may increasingly gravitate towards more complex multi-class classification scenarios. For example, Liang et al. (2023) engineered a high-performance CNN capable of classifying seven distinct tissue types in HCC: tumor, normal, fibrosis, inflammation, necrosis, bile duct reaction, and steatosis. This accomplishment was made possible by leveraging a meticulously curated high-quality and diverse mate-annotation dataset, enabling the model to acquire a more comprehensive understanding of morphological features. The incorporation of multi-class classification and segmentation tasks within DL models holds significant promise. Such endeavors are poised to provide valuable diagnostic and prognostic insights that can greatly enhance patient treatment strategies and offer enhanced support to clinical practitioners.

6.5. Translating AI models towards to clinical setting

While current DL models have demonstrated impressive capabilities in liver cancer HIA, often achieving proficiency comparable to experienced pathologists within their respective in-house datasets (Chen et al., 2020b), their practical clinical deployment remains a challenging endeavor. The translation of these AI models into real clinical applications necessitates further exploration and validation. Many studies primarily emphasize the performance of their models within specific datasets, often relying on fixed evaluation metrics such as Acc and AUC for classification models, and IOU and Dice for localization models. However, these models disengaged from the real clinical setting, which lacks trustworthy clinical applicability.

The initial investigation conducted (Kiani et al., 2020) was undertaken to assess the impact of DL models when employed as aids to pathologists in real clinical scenarios, specifically in distinguishing between HCC and ICC image patches. In their experimental design, they assembled a cohort of 11 pathologists, categorizing them into four distinct groups based on their levels of expertise, comprising three GI pathologists, three non-GI subspecialty pathologists, three anatomic pathology trainees, and two pathologists with a non-specific specialization. To evaluate the interactive effect on diagnostic accuracy, the authors constructed a mixed-effect multivariate logistic regression model, considering various factors such as the pathologists' experience levels, tumor differentiation grades, and whether they received assistance from DL models during the diagnosis process. The findings derived from their study revealed a dual impact of DL models on the diagnostic process: correct model predictions positively influenced diagnostic accuracy, whereas incorrect model predictions exerted a negative impact, sometimes resulting in diagnostic accuracy inferior to that achieved without DL model assistance. This pioneering study serves as a significant milestone in demonstrating the potential translation of AI models into clinical settings. It is hoped that future research endeavors will focus on evaluating the complementary role of DL models in aiding pathologists within the clinical context, rather than simply assessing their performance on in-house datasets. Such investigations can provide valuable insights into the practical utility of AI assistance in real-world medical diagnoses.

In summary, CADs based on DL techniques in the domain of liver cancer histopathology have witnessed rapid advancements. While encountering various technical challenges and potential pitfalls, CAD systems rooted in DL methods continue to exhibit significant promise. Given the insights discussed above, the realization of this potential relies on researchers devising increasingly effective strategies to surmount these challenges and create high-performance DL models designed to better collaborate with human experts. Such advancements hold the potential to gain acceptance among clinicians and address concerns related to automation bias (Skitka et al., 1999; Schemmer et al., 2022), thus fostering the adoption of CAD systems founded on DL techniques. The principal limitation of this survey lies in our endeavor to encompass a wide spectrum of methodologies to generalize the methods employed in this domain comprehensively. This work may inadvertently overlook some valuable contributions due to the extensive studies conducted in this field. Consequently, we are steadfast in our commitment to focusing on this domain and conducting enlightening research to advance AI in digital pathology.

7. Conclusion

In this survey, we have provided a comprehensive overview of the evolution of CAD systems utilizing DL models within the domain of liver cancer HIA. Within these approaches, we have identified and delineated two primary applications, namely diagnosis and prognosis. To facilitate readers' understanding of the research landscape concerning DL techniques in liver cancer histology, we have offered a structured taxonomy. Within the realm of diagnosis, we have highlighted the utilization of supervised learning and weakly supervised learning techniques for a diverse array of liver cancer histopathology tasks, including disease classification and lesion localization. Moreover, we have explored the rapid advancements in DL survival models for liver cancer prognosis, presenting an in-depth analysis of three distinct implementation approaches. Finally, our survey has discussed various outstanding challenges and proposed viable strategies to mitigate them while outlining potential future trends. As a valuable reference, our work has showcased numerous DL models applied in liver cancer histopathology. We anticipate that the insights provided herein will serve as a guide for future research endeavors, inspiring the development of increasingly efficient approaches in the field of computational histopathology for liver cancer and other cancer types. Future work will explore studies

encompassing other cancer types, expanding beyond the scope of liver cancer HIA. We firmly believe that the cross-application of these approaches across various cancer types will significantly contribute to the advancement of AI in digital pathology.

CRediT authorship contribution statement

Haoyang Jiang: Conceptualization, Data curation, Investigation, Methodology, Validation, Visualization, Writing – original draft. **Yimin Yin:** Funding acquisition, Project administration, Supervision. **Jinghua Zhang:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Wanxia Deng:** Formal analysis, Funding acquisition, Methodology, Supervision, Writing – review & editing. **Chen Li:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62306323, the China Scholarship Council under Grant 202206110005, and the "Scientific Research Project of Education Department of Hunan Province" (No. 21C0839).

References

Aatresh, A.A., Alabhya, K., Lal, S., Kini, J., Saxena, P.P., 2021. 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. 16, 1549–1563.

Alirezazadeh, P., Hejrati, B., Monsef-Esfahani, A., Fathi, A., 2018. Representation learning-based unsupervised domain adaptation for classification of breast cancer histopathology images. Biocybern. Biomed. Eng. 38 (3), 671–683.

BenTaieb, A., Hamarneh, G., 2017. Adversarial stain transfer for histopathology image analysis. IEEE Trans. Med. Imaging 37 (3), 792–802.

Blanche, P., Kattan, M.W., Gerds, T.A., 2019. The c-index is not proper for the evaluation of-year predicted risks. Biostatistics 20 (2), 347–357.

Calderaro, J., Ghaffari Laleh, N., Zeng, Q., Maille, P., Favre, L., Pujals, A., Klein, C., Bazille, C., Heij, L.R., Uguen, A., et al., 2023. Deep learning-based phenotyping reclassifies combined hepatocellular-cholangiocarcinoma. Nat. Commun. 14 (1), 8290.

Calderaro, J., Kather, J.N., 2021. Artificial intelligence-based pathology for gastrointestinal and hepatobiliary cancers. Gut 70 (6), 1183–1193.

Calderaro, J., Seraphin, T.P., Luedde, T., Simon, T.G., 2022. Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma. J. Hepatol. 76 (6), 1348–1361.

Campanella, G., Hanna, M.G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K.J., Brogi, E., Reuter, V.E., Klimstra, D.S., Fuchs, T.J., 2019. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat. Med. 25 (8), 1301–1309.

Chattopadhay, A., Sarkar, A., Howlader, P., Balasubramanian, V.N., 2018. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In: 2018 IEEE Winter Conference on Applications of Computer Vision. WACV, IEEE, pp. 839–847.

Chaurasia, A., Culurciello, E., 2017. Linknet: Exploiting encoder representations for efficient semantic segmentation. In: 2017 IEEE Visual Communications and Image Processing. VCIP, IEEE, pp. 1–4.

Chen, C., Chen, C., Ma, M., Ma, X., Lv, X., Dong, X., Yan, Z., Zhu, M., Chen, J., 2022a. Classification of multi-differentiated liver cancer pathological images based on deep learning attention mechanism. BMC Med. Inform. Decis. Mak. 22 (1), 1–13.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020a. A simple framework for contrastive learning of visual representations. In: ICML. PMLR, pp. 1597–1607.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017a. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell. 40 (4), 834–848.

Chen, H., Qi, X., Yu, L., Dou, Q., Qin, J., Heng, P.-A., 2017b. DCAN: Deep contour-aware networks for object instance segmentation from histology images. Med. Image Anal. 36, 135–146.

Chen, X., Wang, X., Zhang, K., Fung, K.-M., Thai, T.C., Moore, K., Mannel, R.S., Liu, H., Zheng, B., Qiu, Y., 2022b. Recent advances and clinical applications of deep learning in medical image analysis. Med. Image Anal. 102444.

Chen, H., Wang, K., Zhu, Y., Yan, J., Ji, Y., Li, J., Xie, D., Huang, J., Cheng, S., Yao, J., 2021. From pixel to whole slide: automatic detection of microvascular invasion in hepatocellular carcinoma on histopathological image via cascaded networks. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24. Springer, pp. 196–205.

Chen, M., Zhang, B., Topatana, W., Cao, J., Zhu, H., Juengpanich, S., Mao, Q., Yu, H., Cai, X., 2020b. Classification and mutation prediction based on histopathology h&e images in liver cancer using deep learning. NPJ Precis. Oncol. 4 (1), 1–7.

Chhikara, B.S., Parang, K., 2023. Global cancer statistics 2022: the trends projection analysis. Chem. Biol. Lett. 10 (1), 451.

Ciga, O., Xu, T., Martel, A.L., 2022. Self supervised contrastive learning for digital histopathology. Mach. Learn. Appl. 7, 100198.

Courtiol, P., Tramel, E.W., Sanselme, M., Wainrib, G., 2018. Classification and disease localization in histopathology using only global labels: A weakly-supervised approach. arXiv preprint arXiv:1802.02212.

Cui, Y., Zhang, G., Liu, Z., Xiong, Z., Hu, J., 2019. A deep learning algorithm for one-step contour aware nuclei segmentation of histopathology images. Med. Biol. Eng. Comput. 57, 2027–2043.

Das, K., Conjeti, S., Chatterjee, J., Sheet, D., 2020. Detection of breast cancer from whole slide histopathological images using deep multiple instance CNN. IEEE Access 8, 213502–213511.

Deng, H., Zou, N., Du, M., Chen, W., Feng, G., Yang, Z., Li, Z., Zhang, Q., 2023. Understanding and unifying fourteen attribution methods with taylor interactions. arXiv preprint arXiv:2303.01506.

Diao, S., Tian, Y., Hu, W., Hou, J., Lambo, R., Zhang, Z., Xie, Y., Nie, X., Zhang, F., Racoceanu, D., et al., 2022. 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. 192 (3), 553–563.

Dietterich, T.G., Lathrop, R.H., Lozano-Pérez, T., 1997. Solving the multiple instance problem with axis-parallel rectangles. Artif. Intell. 89 (1–2), 31–71.

Dong, X., Li, M., Zhou, P., Deng, X., Li, S., Zhao, X., Wu, Y., Qin, J., Guo, W., 2022. Fusing pre-trained convolutional neural networks features for multi-differentiated subtypes of liver cancer on histopathological images. BMC Med. Inform. Decis. Mak. 22 (1), 1–27.

Edmondson, H.A., Steiner, P.E., 1954. Primary carcinoma of the liver. A study of 100 cases among 48,900 necropsies. Cancer 7 (3), 462–503.

Epstein, J.I., Egevad, L., Amin, M.B., Delahunt, B., Srigley, J.R., Humphrey, P.A., 2016. The 2014 international society of urological pathology (ISUP) consensus conference on gleason grading of prostatic carcinoma. Am. J. Surg. Pathol. 40 (2), 244–252.

Erickson, B., Kirk, S., Lee, Y., Bathe, O., Kearns, M., Gerdes, C., Rieger-Christ, K., Lemerman, J., 2016. Radiology data from the cancer genome atlas liver hepatocellular carcinoma [TCGA-LIHC] collection. Cancer Imag. Arch 10, K9.

Feng, Y., Hafiane, A., Laurent, H., 2021. A deep learning based multiscale approach to segment the areas of interest in whole slide images. Comput. Med. Imaging Graph. 90, 101923.

Finkel, R.A., Bentley, J.L., 1974. Quad trees a data structure for retrieval on composite keys. Acta Inform. 4, 1–9.

Goceri, E., Shah, Z.K., Layman, R., Jiang, X., Gurcan, M.N., 2016. Quantification of liver fat: A comprehensive review. Comput. Biol. Med. 71, 174–189.

Goode, A., Gilbert, B., Harkes, J., Jukic, D., Satyanarayanan, M., 2013. OpenSlide: A vendor-neutral software foundation for digital pathology. J. Pathol. Inform. 4 (1), 27.

Guo, Z., Wang, Q., Müller, H., Palpanas, T., Loménie, N., Kurtz, C., 2023. A hierarchical transformer encoder to improve entire neoplasm segmentation on whole slide images of hepatocellular carcinoma. In: 2023 IEEE 20th International Symposium on Biomedical Imaging. ISBI, IEEE, pp. 1–5.

Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M., Yener, B., 2009. Histopathological image analysis: A review. IEEE Rev. Biomed. Eng. 2, 147–171.

Hägele, M., Eschrich, J., Ruff, L., Alber, M., Schallenberg, S., Guillot, A., Roderburg, C., Tacke, F., Klauschen, F., 2023. Leveraging weak complementary labels to improve semantic segmentation of hepatocellular carcinoma and cholangiocarcinoma in H&E-stained slides. arXiv preprint arXiv:2302.01813.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: CVPR. pp. 770–778.

Hosseini, S.M., Sikaroudi, M., Babaie, M., Tizhoosh, H., 2023. Proportionally fair hospital collaborations in federated learning of histopathology images. IEEE Trans. Med. Imaging.

Hou, J., Jia, X., Xie, Y., Qin, W., 2022. Integrative histology-genomic analysis predicts hepatocellular carcinoma prognosis using deep learning. Genes 13 (10), 1770.

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: CVPR. pp. 7132–7141.

Huang, W.-C., Chung, P.-C., Tsai, H.-W., Chow, N.-H., Juang, Y.-Z., Tsai, H.-H., Lin, S.-H., Wang, C.-H., 2019. Automatic HCC detection using convolutional network with multi-magnification input images. In: 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems. AICAS, IEEE, pp. 194–198.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: CVPR. pp. 4700–4708.

Huang, Y., Yang, X., Liu, L., Zhou, H., Chang, A., Zhou, X., Chen, R., Yu, J., Chen, J., Chen, C., et al., 2023. Segment anything model for medical images? arXiv preprint arXiv:2304.14660.

Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., Kluger, Y., 2018. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med. Res. Methodol. 18 (1), 1–12.

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., 2020. Impact of a deep learning assistant on the histopathologic classification of liver cancer. NPJ Digit. Med. 3 (1), 1–8.

Kim, Y.J., Jang, H., Lee, K., Park, S., Min, S.-G., Hong, C., Park, J.H., Lee, K., Kim, J., Hong, W., et al., 2021. PAIP 2019: Liver cancer segmentation challenge. Med. Image Anal. 67, 101854.

Kim, B., Wattenberg, M., Gilmer, J., Cai, C., Wexler, J., Viegas, F., et al., 2018. Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (tcav). In: ICML. PMLR, pp. 2668–2677.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., et al., 2023. Segment anything. arXiv preprint arXiv:2304.02643.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. Commun. ACM 60 (6), 84–90.

Lal, S., Das, D., Alabhya, K., Kanfade, A., Kumar, A., Kini, J., 2021. NucleiSegNet: robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images. Comput. Biol. Med. 128, 104075.

Li, S., Jiang, H., Pang, W., 2017. Joint multiple fully connected convolutional neural network with extreme learning machine for hepatocellular carcinoma nuclei grading. Comput. Biol. Med. 84, 156–167.

Li, L., Li, X., Li, W., Ding, X., Zhang, Y., Chen, J., Li, W., 2022. Prognostic models for outcome prediction in patients with advanced hepatocellular carcinoma treated by systemic therapy: a systematic review and critical appraisal. BMC Cancer 22 (1), 750.

Li, X., Wang, W., Hu, X., Yang, J., 2019. Selective kernel networks. In: CVPR. pp. 510–519.

Liang, J., Zhang, W., Yang, J., Wu, M., Dai, Q., Yin, H., Xiao, Y., Kong, L., 2023. Deep learning supported discovery of biomarkers for clinical prognosis of liver cancer. Nat. Mach. Intell. 1–13.

Liao, H., Long, Y., Han, R., Wang, W., Xu, L., Liao, M., Zhang, Z., Wu, Z., Shang, X., Li, X., et al., 2020. Deep learning-based classification and mutation prediction from histopathological images of hepatocellular carcinoma. Clin. Transl. Med. 10 (2).

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection. In: CVPR. pp. 2980–2988.

Lin, Y.-S., Huang, P.-H., Chen, Y.-Y., 2021. Deep learning-based hepatocellular carcinoma histopathology image classification: accuracy versus training dataset size. IEEE Access 9, 33144–33157.

Liu, Z., Liu, Y., Zhang, W., Hong, Y., Meng, J., Wang, J., Zheng, S., Xu, X., 2022. Deep learning for prediction of hepatocellular carcinoma recurrence after resection or liver transplantation: a discovery and validation study. Hepatol. Int. 16 (3), 577–589.

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., Pietikäinen, M., 2020. Deep learning for generic object detection: A survey. Int. J. Comput. Vis. 128, 261–318.

Liu, G.-J., Wang, W., Lu, M.-D., Xie, X.-Y., Xu, H.-X., Xu, Z.-F., Chen, L.-D., Wang, Z., Liang, J.-Y., Huang, Y., et al., 2015. Contrast-enhanced ultrasound for the characterization of hepatocellular carcinoma and intrahepatic cholangiocarcinoma. Liver Cancer 4 (4), 241–252.

Lu, M.Y., Chen, R.J., Kong, D., Lipkova, J., Singh, R., Williamson, D.F., Chen, T.Y., Mahmood, F., 2022. Federated learning for computational pathology on gigapixel whole slide images. Med. Image Anal. 76, 102298.

Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F., 2021. Data-efficient and weakly supervised computational pathology on whole-slide images. Nat. Biomed. Eng. 5 (6), 555–570.

Ma, X., Niu, Y., Gu, L., Wang, Y., Zhao, Y., Bailey, J., Lu, F., 2021. Understanding adversarial attacks on deep learning based medical image analysis systems. Pattern Recognit. 110, 107332.

Ma, J., Wang, B., 2023. Segment anything in medical images. arXiv preprint arXiv:2304.12306.

Malaguarnera, G., Paladina, I., Giordano, M., Malaguarnera, M., Bertino, G., Berretta, M., 2013. Serum markers of intrahepatic cholangiocarcinoma. Dis. Markers 34 (4), 219–228.

Marini, N., Otálora, S., Müller, H., Atzori, M., 2021. Semi-supervised training of deep convolutional neural networks with heterogeneous data and few local annotations: An experiment on prostate histopathology image classification. Med. Image Anal. 73, 102165.

Massarweh, N.N., El-Serag, H.B., 2017. Epidemiology of hepatocellular carcinoma and intrahepatic cholangiocarcinoma. Cancer Control 24 (3), 1073274817729245.

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D., 2021. Image segmentation using deep learning: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 44 (7), 3523–3542.

Moor, M., Banerjee, O., Abad, Z.S.H., Krumholz, H.M., Leskovec, J., Topol, E.J., Rajpurkar, P., 2023. Foundation models for generalist medical artificial intelligence. Nature 616 (7956), 259–265.

Muhammad, H., Sigel, C.S., Campanella, G., Boerner, T., Pak, L.M., Büttner, S., IJzermans, J.N., Koerkamp, B.G., Doukas, M., Jarnagin, W.R., et al., 2019. Unsupervised subtyping of cholangiocarcinoma using a deep clustering convolutional autoencoder. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22. Springer, pp. 604–612.

Muhammad, H., Xie, C., Sigel, C.S., Doukas, M., Alpert, L., Simpson, A.L., Fuchs, T.J., 2021. EPIC-survival: End-to-end part inferred clustering for survival analysis, with prognostic stratification boosting. In: Medical Imaging with Deep Learning.

Murtagh, F., Legendre, P., 2014. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? J. Classification 31, 274–295.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. 9 (1), 62–66.

Qu, W.-F., Tian, M.-X., Lu, H.-W., Zhou, Y.-F., Liu, W.-R., Tang, Z., Yao, Z., Huang, R., Zhu, G.-Q., Jiang, X.-F., et al., 2023. Development of a deep pathomics score for predicting hepatocellular carcinoma recurrence after liver transplantation. Hepatol. Int. 1–15.

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., 2022. Exploring pathological signatures for predicting the recurrence of early-stage hepatocellular carcinoma based on deep learning. Front. Oncol. 12, 968202.

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al., 2021. Learning transferable visual models from natural language supervision. In: ICML. PMLR, pp. 8748–8763.

Ren, J., Hacihaliloglu, I., Singer, E.A., Foran, D.J., Qi, X., 2018. Adversarial domain adaptation for classification of prostate histopathology whole-slide images. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11. Springer, pp. 201–209.

Ren, J., Hacihaliloglu, I., Singer, E.A., Foran, D.J., Qi, X., 2019. Unsupervised domain adaptation for classification of histopathology whole-slide images. Front. Bioeng. Biotechnol. 7, 102.

Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H.R., Albarqouni, S., Bakas, S., Galtier, M.N., Landman, B.A., Maier-Hein, K., et al., 2020. The future of digital health with federated learning. NPJ Digit. Med. 3 (1), 119.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

Ross, A., Doshi-Velez, F., 2018. Improving the adversarial robustness and interpretability of deep neural networks by regularizing their input gradients. In: AAAI. Vol. 32.

Roy, M., Kong, J., Kashyap, S., Pastore, V.P., Wang, F., Wong, K.C., Mukherjee, V., 2021. Convolutional autoencoder based model HistoCAE for segmentation of viable tumor regions in liver whole-slide images. Sci. Rep. 11 (1), 1–10.

Roy, A.G., Navab, N., Wachinger, C., 2018. Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I. Springer, pp. 421–429.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323 (6088), 533–536.

Saillard, C., Schmauch, B., Laifa, O., Moarii, M., Toldo, S., Zaslavskiy, M., Pronier, E., Laurent, A., Amaddeo, G., Regnault, H., et al., 2020. Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides. Hepatology 72 (6), 2000–2013.

Schemmer, M., Kühl, N., Benz, C., Satzger, G., 2022. On the influence of explainable AI on automation bias. arXiv preprint arXiv:2204.08859.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: CVPR. pp. 618–626.

Shi, J.-Y., Wang, X., Ding, G.-Y., Dong, Z., Han, J., Guan, Z., Ma, L.-J., Zheng, Y., Zhang, L., Yu, G.-Z., et al., 2021. Exploring prognostic indicators in the pathological images of hepatocellular carcinoma based on deep learning. Gut 70 (5), 951–961.

Simonyan, K., Vedaldi, A., Zisserman, A., 2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Singal, A.G., Pillai, A., Tiro, J., 2014. Early detection, curative treatment, and survival rates for hepatocellular carcinoma surveillance in patients with cirrhosis: a meta-analysis. PLoS Med. 11 (4), e1001624.

Skitka, L.J., Mosier, K.L., Burdick, M., 1999. Does automation bias decision-making? Int. J. Human-Comput. Stud. 51 (5), 991–1006.

Srinidhi, C.L., Ciga, O., Martel, A.L., 2021. Deep neural network models for computational histopathology: A survey. Med. Image Anal. 67, 101813.

Steck, H., Krishnapuram, B., Dehing-Oberije, C., Lambin, P., Raykar, V.C., 2007. On ranking in survival analysis: Bounds on the concordance index. Adv. Neural Inf. Process. Syst. 20.

Sun, C., Xu, A., Liu, D., Xiong, Z., Zhao, F., Ding, W., 2019. Deep learning-based classification of liver cancer histopathology images using only global labels. IEEE J. Biomed. Health Inform. 24 (6), 1643–1651.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: CVPR. pp. 1–9.

Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. PMLR, pp. 6105–6114.

Tan, J.W., Lee, K., Lee, K., Jeong, W.-K., 2021. Improving the accuracy of intrahepatic cholangiocarcinoma subtype classification by hidden class detection via label smoothing. In: 2021 IEEE 18th International Symposium on Biomedical Imaging. ISBI, IEEE, pp. 1944–1948.

Tan, J.W., Nguyen, K.T., Lee, K., Jeong, W.-K., 2023. Multi-scale contrastive learning with attention for histopathology image classification. In: Medical Imaging 2023: Digital and Computational Pathology. Vol. 12471, SPIE, pp. 294–301.

Tibshirani, R., 1997. The lasso method for variable selection in the Cox model. Statist. Med. 16 (4), 385–395.

Toğaçar, M., Özkurt, K.B., Ergen, B., Cömert, Z., 2020. BreastNet: A novel convolutional neural network model through histopathological images for the diagnosis of breast cancer. Physica A 545, 123592.

Van der Velden, B.H., Kuijf, H.J., Gilhuijs, K.G., Viergever, M.A., 2022. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. Med. Image Anal. 79, 102470.

Wang, X., Fang, Y., Yang, S., Zhu, D., Wang, M., Zhang, J., Tong, K.-y., Han, X., 2021a. A hybrid network for automatic hepatocellular carcinoma segmentation in H&E-stained whole slide images. Med. Image Anal. 68, 101914.

Wang, Y., Ma, X., Bailey, J., Yi, J., Zhou, B., Gu, Q., 2021b. On the convergence and robustness of adversarial training. arXiv preprint arXiv:2112.08304.

Wang, J., Xu, Z., Pang, Z.-F., Huo, Z., Luo, J., 2021c. Tumor detection for whole slide image of liver based on patch-based convolutional neural network. Multimedia Tools Appl. 80, 17429–17440.

Wang, J., Zhang, H., 2019. Bilateral adversarial training: Towards fast training of more robust models against adversarial attacks. In: CVPR. pp. 6629–6638.

Woo, S., Park, J., Lee, J.-Y., Kweon, I.S., 2018. Cbam: Convolutional block attention module. In: ECCV. pp. 3–19.

Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., 2017. Aggregated residual transformations for deep neural networks. In: CVPR. pp. 1492–1500.

Xie, C., Muhammad, H., Vanderbilt, C.M., Caso, R., Yarlagadda, D.V.K., Campanella, G., Fuchs, T.J., 2020. Beyond classification: Whole slide tissue histopathology analysis by end-to-end part learning. In: Medical Imaging with Deep Learning. PMLR, pp. 843–856.

Xie, S., Tu, Z., 2015. Holistically-nested edge detection. In: CVPR. pp. 1395–1403.

Xu, Y., Jia, Z., Wang, L.-B., Ai, Y., Zhang, F., Lai, M., Chang, E.I., et al., 2017. Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. BMC Bioinform. 18 (1), 1–17.

Xu, Y., Zhu, J.-Y., Eric, I., Chang, C., Lai, M., Tu, Z., 2014. Weakly supervised histopathology cancer image segmentation and classification. Med. Image Anal. 18 (3), 591–604.

Yamamoto, K., Imamura, H., Matsuyama, Y., Kume, Y., Ikeda, H., Norman, G.L., Shums, Z., Aoki, T., Hasegawa, K., Beck, Y., et al., 2010. AFP, AFP-L3, DCP, and GP73 as markers for monitoring treatment response and recurrence and as surrogate markers of clinicopathological variables of HCC. J. Gastroenterol. 45, 1272–1282.

Yamashita, R., Long, J., Saleem, A., Rubin, D.L., Shen, J., 2021. Deep learning predicts postsurgical recurrence of hepatocellular carcinoma from digital histopathologic images. Sci. Rep. 11 (1), 1–14.

Yan, J., Chen, H., Wang, K., Ji, Y., Zhu, Y., Li, J., Xie, D., Xu, Z., Huang, J., Cheng, S., et al., 2021. Hierarchical attention guided framework for multi-resolution collaborative whole slide image segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24. Springer, pp. 153–163.

Yang, T.-L., Tsai, H.-W., Huang, W.-C., Lin, J.-C., Liao, J.-B., Chow, N.-H., Chung, P.-C., 2022. Pathologic liver tumor detection using feature aligned multi-scale convolutional network. Artif. Intell. Med. 125, 102244.

Zanjani, F.G., Zinger, S., Bejnordi, B.E., van der Laak, J.A., de With, P.H., 2018. Stain normalization of histopathology images using generative adversarial networks. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, pp. 573–577.

Zeng, Q., Klein, C., Caruso, S., Maille, P., Laleh, N.G., Sommacale, D., Laurent, A., Amaddeo, G., Gentien, D., Rapinat, A., et al., 2022. Artificial intelligence predicts immune and inflammatory gene signatures directly from hepatocellular carcinoma histology. J. Hepatol. 77 (1), 116–127.

Zhai, Z., Wang, C., Sun, Z., Cheng, S., Wang, K., 2022. Deep neural network guided by attention mechanism for segmentation of liver pathology image. In: Proceedings of 2021 Chinese Intelligent Systems Conference: Volume III. Springer, pp. 425–435.

Zhang, H., Ren, F., Wang, Z., Rao, X., Li, L., Hao, J., Yan, R., Luo, J., Du, M., Zhang, F., 2019. Predicting tumor mutational burden from liver cancer pathological images using convolutional neural network. In: 2019 IEEE International Conference on Bioinformatics and Biomedicine. BIBM, IEEE, pp. 920–925.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., 2016. Learning deep features for discriminative localization. In: CVPR. pp. 2921–2929.

Zhu, P., Wang, C., Sun, Z., Cheng, S., Wang, K., 2022. Segmentation of liver cancer pathology images based on multi-scale feature fusion. In: Proceedings of 2021 Chinese Intelligent Systems Conference: Volume III. Springer, pp. 596–605.