Automatic origin prediction of liver metastases via hierarchical artificial-intelligence system trained on multiphasic CT data: a retrospective, multicentre study

Hongjie Xin,[a,i] Yiwen Zhang,[b,i] Qianwei Lai,[a] Naying Liao,[a] Jing Zhang,[c] Yanping Liu,[a,d] Zhihua Chen,[e] Pengyuan He,[f] Jian He,[a] Junwei Liu,[g] Yuchen Zhou,[h] Wei Yang,[b,**] and Yuanping Zhou[a,*]

[a]Department of Gastroenterology, Nanfang Hospital, Southern Medical University, Guangzhou, China
[b]School of Biomedical Engineering, Southern Medical University, Guangzhou, China
[c]Department of Medical Imaging Center, Nanfang Hospital, Southern Medical University, Guangzhou, China
[d]Department of Gastroenterology, The Second Affiliated Hospital, University of South China, Hengyang, China
[e]Department of Radiology, The Second Affiliated Hospital, University of South China, Hengyang, China
[f]Department of Infectious Diseases, The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, China
[g]Department of Infectious Diseases and Hepatology Unit, Nanfang Hospital, Southern Medical University, Guangzhou, China
[h]Department of General Surgery, Cancer Center, Integrated Hospital of Traditional Chinese Medicine, Southern Medical University, Guangzhou, China

eClinicalMedicine 2024;69: 102464
Published Online xxx
https://doi.org/10.1016/j.eclinm.2024.102464

Summary

Background Currently, the diagnostic testing for the primary origin of liver metastases (LMs) can be laborious, complicating clinical decision-making. Directly classifying the primary origin of LMs at computed tomography (CT) images has proven to be challenging, despite its potential to streamline the entire diagnostic workflow.

Methods We developed ALMSS, an artificial intelligence (AI)-based LMs screening system, to provide automated liver contrast-enhanced CT analysis for distinguishing LMs from hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC), as well as subtyping primary origin of LMs as six organ systems. We processed a CECT dataset between January 1, 2013 and June 30, 2022 (n = 3105: 840 HCC, 354 ICC, and 1911 LMs) for training and internally testing ALMSS, and two additional cohorts (n = 622) for external validation of its diagnostic performance. The performance of radiologists with and without the assistance of ALMSS in diagnosing and subtyping LMs was assessed.

Findings ALMSS achieved average area under the curve (AUC) of 0.917 (95% confidence interval [CI]: 0.899–0.931) and 0.923 (95% [CI]: 0.905–0.937) for differentiating LMs, HCC and ICC on both the internal testing set and external testing set, outperformed that of two radiologists. Moreover, ALMSS yielded average AUC of 0.815 (95% [CI]: 0.794–0.836) and 0.818 (95% [CI]: 0.790–0.842) for predicting six primary origins on both two testing sets. Interestingly, ALMSS assigned origin diagnoses for LMs with pathological phenotypes beyond the training categories with average AUC of 0.761 (95% [CI]: 0.657–0.842), which verify the model's diagnostic expandability.

Interpretation Our study established an AI-based diagnostic system that effectively identifies and characterizes LMs directly from multiphasic CT images.

Funding National Natural Science Foundation of China, Guangdong Provincial Key Laboratory of Medical Image Processing.

Keywords: Liver metastases; Primary liver cancer; Artificial intelligence; Computed tomography; Primary origin

*Corresponding author.
**Corresponding author.
E-mail addresses: ypzhou@smu.edu.cn (Y. Zhou), weiyanggm@gmail.com (W. Yang).
[i]These authors contributed equally as first authors.

Research in context

Evidence before this study

We searched PubMed up to March 30, 2023, for research articles containing the terms of "(artificial intelligence)" OR "(deep learning)" OR "(convolutional neural network)" AND "(liver metastases)" OR "(liver cancer)" AND "(CT)", without any date or language restrictions. All the previous studies focused on developing screening systems at CT scans for liver lesion classification that restricted to differentiation between benign and malignant lesions, or primary and metastatic lesions. We found no screening system could directly output the site of origin for liver metastases and apply to new categories beyond training samples.

Added value of this study

Here, we present a deep learning system (ALMSS) that adopts a fully-automated design and has the advantage of not requiring annotations from radiologists, which could be particularly important in a clinical setting where expertise in liver tumor imaging is limited. To our knowledge, this system is the first to predict the primary origin of liver metastases based on conventional imaging approach (CECT). The two-hierarchical design of this system addresses the uncertainty in determining a liver tumor as either a new-emerging primary liver cancer or a relapse of an antecedent malignancy for individuals with a history of extrahepatic malignancies. On the other hand, the entire pipeline can be activated to streamline the origin diagnosis for complex cases where the primary tumors' site is not readily apparent, enabling clinicians to focus their efforts on confirming the most likely diagnostic sites rather than adopting a broad, scatter-shot diagnostic approach.

Implications of all the available evidence

ALMSS could potentially assist less experienced radiologists, including those in low resource areas and help determine the subsequent site-specific examinations even before invasive histology test results are available. This allows for the rapid localization of the origin of metastatic tumors at low cost, enabling timely referral of patients to specialized departments for further vital diagnostic work-up and treatment planning.

Introduction

The liver stands out as the organ most susceptible to metastatic diseases due to its unique cellular and architectural composition, which creates a favorable environment for cancer cells.[1] Liver metastases (LMs) can originate from various primary tumor sites, with the most common sources being solid tumors in the colorectum, pancreas, stomach, breast, or lung.[2] LMs are a leading cause of cancer-related mortality and pose significant challenges for clinical intervention.[3–5] As current therapeutic options predominantly target the primary tumor, a variety of tailored and site-specific treatment approaches have been adopted to improve clinical outcomes in patients with LMs, including target drugs and immune checkpoint inhibitors.[6] However, in cases where the primary origin cannot be identified, patients are often treated with empirical combination chemotherapy, resulting in a low median overall survival of 6–15 months.[7–9]

Liver cancers presenting as metastases with an unknown primary tumor are not uncommon in clinical practice. Due to the diverse localization and histology of primary tumors, patients often undergo whole-body positron emission tomography (PET) examinations and repeated histology tests to identify the origin.[10] However, the former can be expensive and may lack clinical penetration in low-resource settings, while the latter still involves invasive sampling and complex slides preparation. Computed tomography (CT) imaging is indeed a critical noninvasive technique in the initial diagnostic pathway of LMs.[11] Radiologists leverage explicit features selected from CT images to differentiate between benign and malignant liver lesions and to determine whether a malignant hepatic tumor is primary or metastatic.[12,13] However, these features, known as semantic features, may indicate the presence of a metastatic lesion, but they do not provide sufficient diagnostic significance about the primary site from which the cancer originated.

With recent advancements in integrating artificial intelligence (AI) and medicine, there is the potential for AI classifiers to address the unmet need in the conventional imaging approach. Deep learning (DL), a core technique of AI, has significantly enhanced the accuracy and efficiency of segmentation and classification for medical images, surpassing the human capabilities.[14] Early but promising evidence suggests that machine learning algorithms can effectively identify metastatic cancer types by extracting computational features from CT images, revealing local or global differences that may be unperceived to the human observer.[15] Based on these findings, we hypothesized that CT scan empowered by DL technology could unlock untapped potential for developing an alternative diagnostic pattern. This pattern would allow us to infer the primary origin of LMs directly from CT images, enabling subsequent site-specific examinations even before histological test results are available.

To test our hypothesis, we developed and validated an AI-based LMs screening system (ALMSS), which automatically identifies liver tumor as primary or metastatic and predicts the primary origin across different types of LMs. This study is expected to streamline the diagnostic process for complex cases with occult primary tumors by automating the analysis of multiphasic CT images.

Methods

Study design

The patient-level classification task focused on six extrahepatic organ systems: breast, intestine (duodenum, jejunum, ileum, colon, or rectum), respiratory system (nasopharynx or lung), esophagogastric system (esophagus, gastroesophageal junction, or stomach), pancreatobiliary system (pancreas, common bile duct, gallbladder, or Vater's ampulla), and reproductive system (ovary or uterus). These six classes were defined based on several criteria, including the adjacency of the organ systems, their common embryological origin, and the utilization of shared pathways, such as the biliary tract and pancreas. Importantly, these grouped organ systems are often examined together in clinical practice due to their proximity. As these six origin sites account for over 90% of LMs diagnoses,[2] our objective was to demonstrate that this AI system (ALMSS) can effectively handle the majority of LMs diagnoses.

Training and testing datasets generation

We retrospectively identified 3727 eligible cases with a clinical diagnosis of either primary or metastatic liver tumors [980 with hepatocellular carcinoma (HCC) and 474 with intrahepatic cholangiocarcinoma (ICC), and 2273 with LMs (2083 cases pathologically confirmed as adenocarcinoma or squamous cell carcinoma for model training and testing, 190 cases pathologically diagnosed with other rare pathological types for model's diagnostic expandability testing)] from three independent centers (Fig. 1). The training cohort comprised 2061 patients who were consecutively diagnosed at Nanfang Hospital of Southern Medical University (NFHSMU, Guangzhou, China) from January 1, 2013, to December 31, 2019, and the internal testing cohort comprised 884 patients who were consecutively diagnosed at the same hospital from January 1, 2020, to June 30, 2022. Besides, 355 patients diagnosed at the Second Affiliated Hospital of University of South China (SAHUSC, Hengyang, China) and 237 patients at the Fifth Affiliated Hospital of Sun Yat-sen University (FAHSYU, Zhuhai, China) from July 1, 2020, to June 30, 2022 were used for external testing. The inclusion criteria were as follows: (1) being aged 18 years or older, (2) have a confirmed pathological diagnosis of a liver tumor either through surgery or biopsy, (3) underwent a liver CECT scan at the initial diagnosis of liver lesions, and (4) have a definite diagnosis of primary tumor site, determined by biopsy and/or formal clinical follow-up. Exclusion criteria were applied to ensure data quality, and included: (1) CT scans with incomplete images or inadequate image quality that could compromise accurate analysis, and (2) cases where liver tumors were undetectable in CT scan (referred to as CT-negative), but were diagnosed using other imaging modalities. All enrolled patients underwent four-phase abdominal CECT scans. More information regarding the acquisition and processing procedure of CT images can be found in Supplementary Materials. An overview of the study workflow is provided in Fig. 2. This study followed the reporting guidelines prescribed by the Standards for Reporting of Diagnostic Accuracy (STARD) checklist.[16]

Ethics

This study was approved by the Ethics Committee of Nanfang Hospital, Southern Medical University (NFEC-2022-119). All procedures performed in this study were accordance with the Declaration of Helsinki and its amendments. Informed consent was exempted by the Ethics Committee for the retrospective dataset.

Overview of ALMSS

As shown in Fig. 3, ALMSS comprises two modules: the segmentation module and the classification module. Two separate 3D U-Net models were trained for the automatic segmentation of the region of interest (ROI) for liver and tumor, using nnU-Net[17] as the framework. And then, the ROIs of the liver and tumor generated by the segmentation module were fed into the classification module for further diagnostic tasks. The core component of ALMSS was built based on two sequential neural networks. When a test sample is provided, the first neural network performs a differential diagnosis of three major malignant liver tumors: HCC, ICC, and LMs. If a metastatic lesion is identified in the initial step, the second neural network is activated to predict the site of origin. Both neural networks share a common structure consisting of three essential components: a feature extraction module, a feature fusion module, and a classifier. Initially, we employed four 3D Residual Connected Network (3D-ResNet) backbones to extract global features and local features from the ROI of the liver and tumor, respectively. Meanwhile, a Multi-Layer Perceptron (MLP) was utilized to extract relevant clinical features. The clinical information and image information from each phase were then integrated through addition operation at the feature level. Furthermore, the information from the four phases was combined using a Long Short-Term Memory (LSTM[18]) model for fusion purposes. Finally, a fully connected layer, followed by a softmax function, served as the classifier to generate the model's predicted probabilities for the corresponding task. Based on the trained models, we selected the cut-off values that yields best F1 scores for each classification task. More information regarding these networks and their training process are provided in Supplementary Materials.

Comparison with benchmark models

To evaluate the advantage of our established framework, the performance of ALMSS was compared with traditional models, including a classic Convolutional Neural Network (CNN),[13] a novel hierarchical CNN (H-CNN),[19] and a SpatialExtractor-TemporalEncoder-Integration

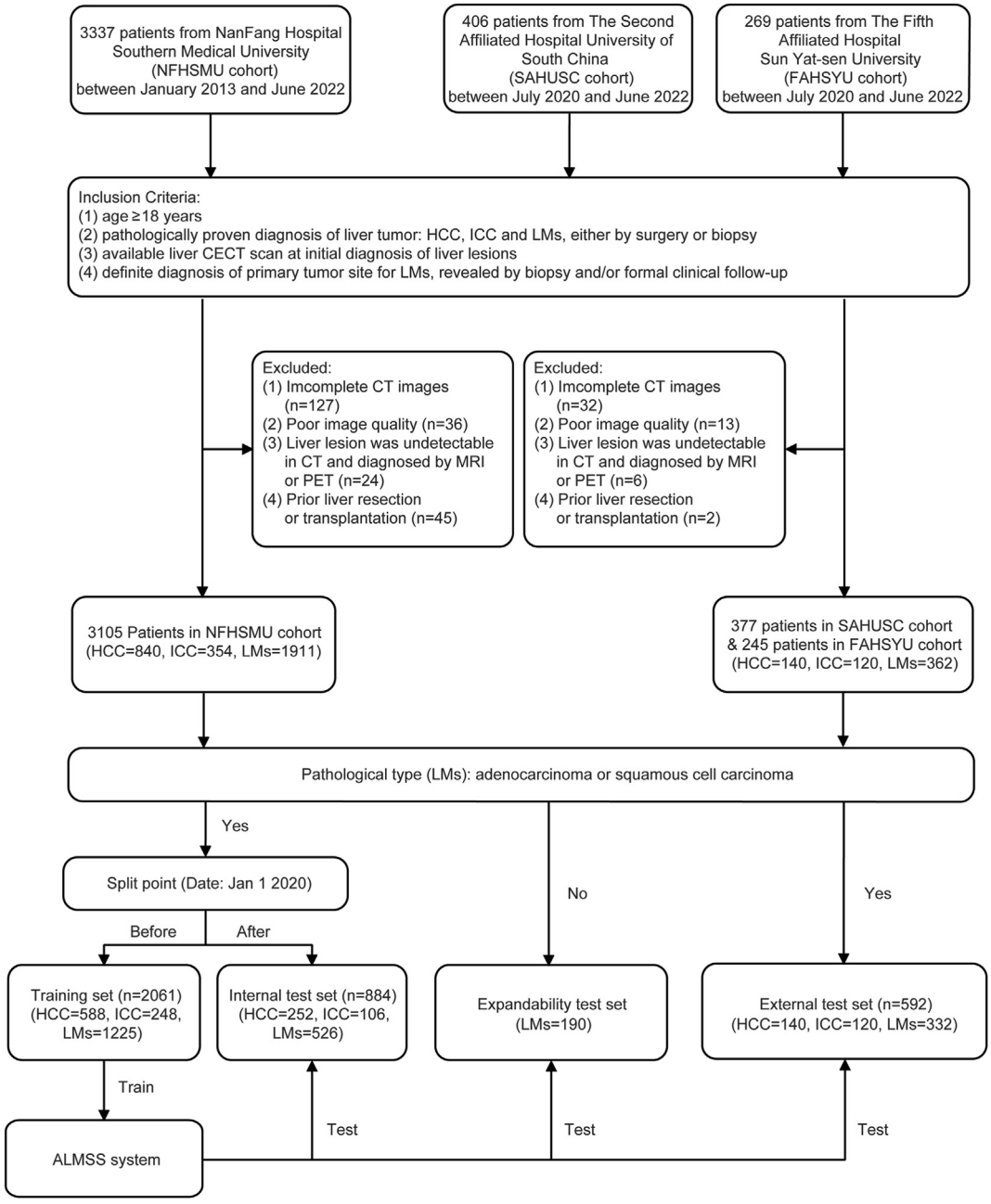


Fig. 1: **The workflow of study enrollment**.

Classifier (STIC).[20] All the models were constructed in the training set and tested in the internal and external test sets.

Evaluation on the auxiliary performance of ALMSS in clinical scenario

A junior radiologist (JR) with 5 years of liver imaging experience, and a senior radiologist (SR) with more than 10 years of liver imaging experience, were invited to review the images on both the internal and external testing sets. To ensure unbiased assessments, the radiologists were blinded to the pathological results and reviewed CT images encompassing the entire liver using ITK-SNAP software (version 4.0.0) (www.itksnap.org). During their first assessment, the radiologists were required to provide judgment on the classification of HCC, ICC, and LMs. Once their initial decision was made, it was recorded. Following a washout period of

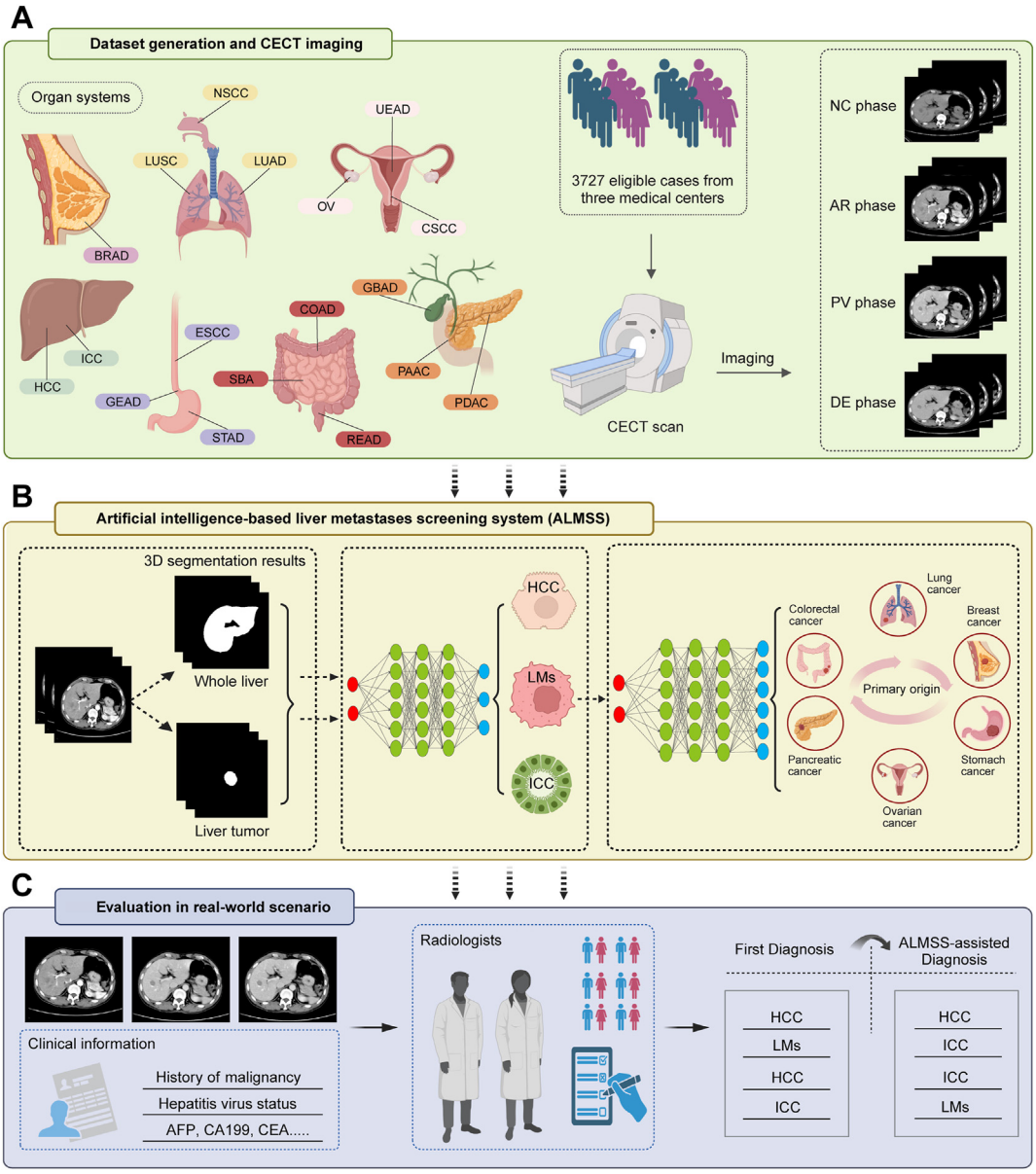


Fig. 2: **The roadmap of overall study**. **(A)** The study included subjects representing primary and metastatic liver tumors of known primary origins with their assigned name acronyms (refer to Table 1). For model development and testing, we collected, in total, 3727 patient cases with confirmed diagnosis and their corresponding CECT scans from three medical centers. **(B)** The screening system consisted of three parts: (1) the segmentation network locating the liver and tumor, and segmenting the corresponding regions of interest; (2) the first screening network was constructed to provide automatic referral for patients with LMs; (3) the second screening network predicting the primary sites of LMs and making a final decision. **(C)** The auxiliary performance of the system was evaluated in a real-world scenario. LMs = liver metastases.

one month, they were invited to re-evaluate the tumor diagnosis with the assistance of ALMSS. The dominant predicted diagnosis generated by ALMSS were presented to them, giving them the option to either maintain their initial decision or adjust it. The performance of ALMSS was compared with that of the radiologists. The Chi-Squared test was used to assess the performance between radiologists with and without the assistance of ALMSS. The accuracy, sensitivity and specificity were calculated for the evaluation.

Evaluation on challenging cases of liver metastases

We identified a subset of challenging LMs cases from two test sets, including patients with liver tumors could

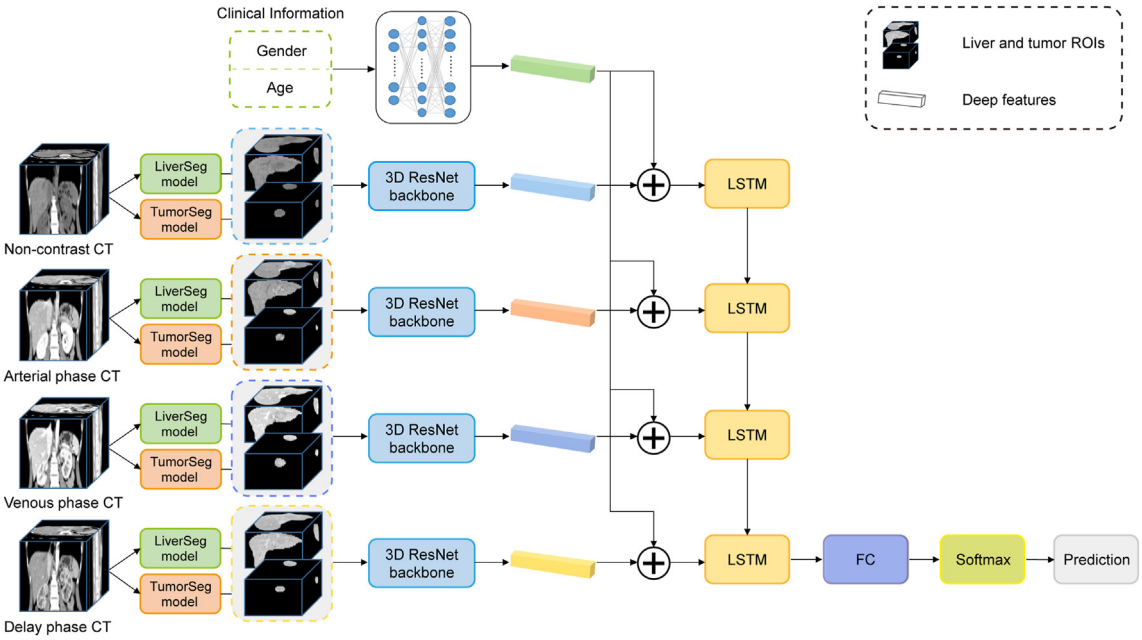


Fig. 3: **Overview pipeline of artificial intelligence-based liver metastases screening system (ALMSS)**. Two separate segmentation models were specifically designed to perform liver and tumor segmentation of abdominal CT images in pre-contrast phase and three enhanced phases. The segmentation results were input into the 3D ResNet backbone in the classification module to generate CT image features from each phase. The gender and age of patient were input into a MLP to output relevant clinical features. The clinical features and image features from each phase were then integrated using an addition operation at the feature level. The information from four phases was combined using an LSTM model for fusion purposes and input into a fully connected layer followed by a softmax function to get respective predicted probabilities in the screening system. CT = computed tomography. MLP = multi-layer perceptron. LSTM = long short-term memory. FC = fully connected layer.

not be diagnosed definitely with conventional pathological analysis. Owing to the difficulty of assessing cases whose tumors were characterized as poorly differentiated, pathologists often assign a spectrum of potential primary origins in their pathology reports. For these cases, further clinical correlation or PET confirmation is required. Two radiologists reviewed CT images in combination with electronic medical records (EMR), comprising laboratory biochemical profiles, family history, treatment history, and pathology reports to make their preliminary judgment for primary origins. Subsequently, under the guidance of ALMSS, they revisited their judgments for a second evaluation. Overall accuracy was used as the primary metric for the evaluation.

Statistics

Continuous variables are presented as medians along with their corresponding standard deviation (SD). Categorical variables are reported as counts (%). Statistical significance was determined with a 2-tailed p-value of less than 0.05.

For segmentation, the Dice metrics and Jaccard coefficient were used for model evaluation. For classification, the receiver operating characteristic (ROC) curves were plotted for each classification task, using the true-positive rate (sensitivity) vs. the false-positive rate (1-specificity). The 95% confidence interval (CI) for these metrics was calculated for model evaluation. Details regarding model evaluation are provided in Supplementary Materials. Heat maps were produced using gradient-weighted class activation mapping (Grad-CAM) for model interpretability.[21]

Model development and computation of various classification metrics were performed using Python (version 3.7.7). Descriptive analysis was conducted in SPSS (version 22.0, Chicago, IL, USA) and R software (version 4.1.0).

Role of the funding source

The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Results

Participants' characteristics

ALMSS was trained and validated on a processed liver cancer dataset, consisting of 3727 patient cases representing with primary and metastatic liver tumors across eight cancer groups. The sample size for each cancer group and their corresponding demographic

distribution, highlighting the potential diagnostic value of age and gender, can be found in Fig. 4A and B. Each type of LMs encompassed both common and rare pathological types (Table 1 and Supplementary Table S1). The training set consisted of 2061 cases, with 642 (31.1%) being female. The mean age for this set was 55.75 years, with a standard deviation (SD) of 11.48. Model evaluation was conducted on the internal test set, which included 884 cases (33.2% female) with a mean (SD) age of 56.02 (11.80) years. Additionally, the pooled external test set comprised 592 cases (35.3% female) with a mean (SD) age of 56.10 (11.27) years. Lastly, the expandability test set consisted of 190 cases (24.7% female) with a mean (SD) age of 57.22 (12.59) years. Detailed information of sample distribution within each dataset is also shown in Table 2.

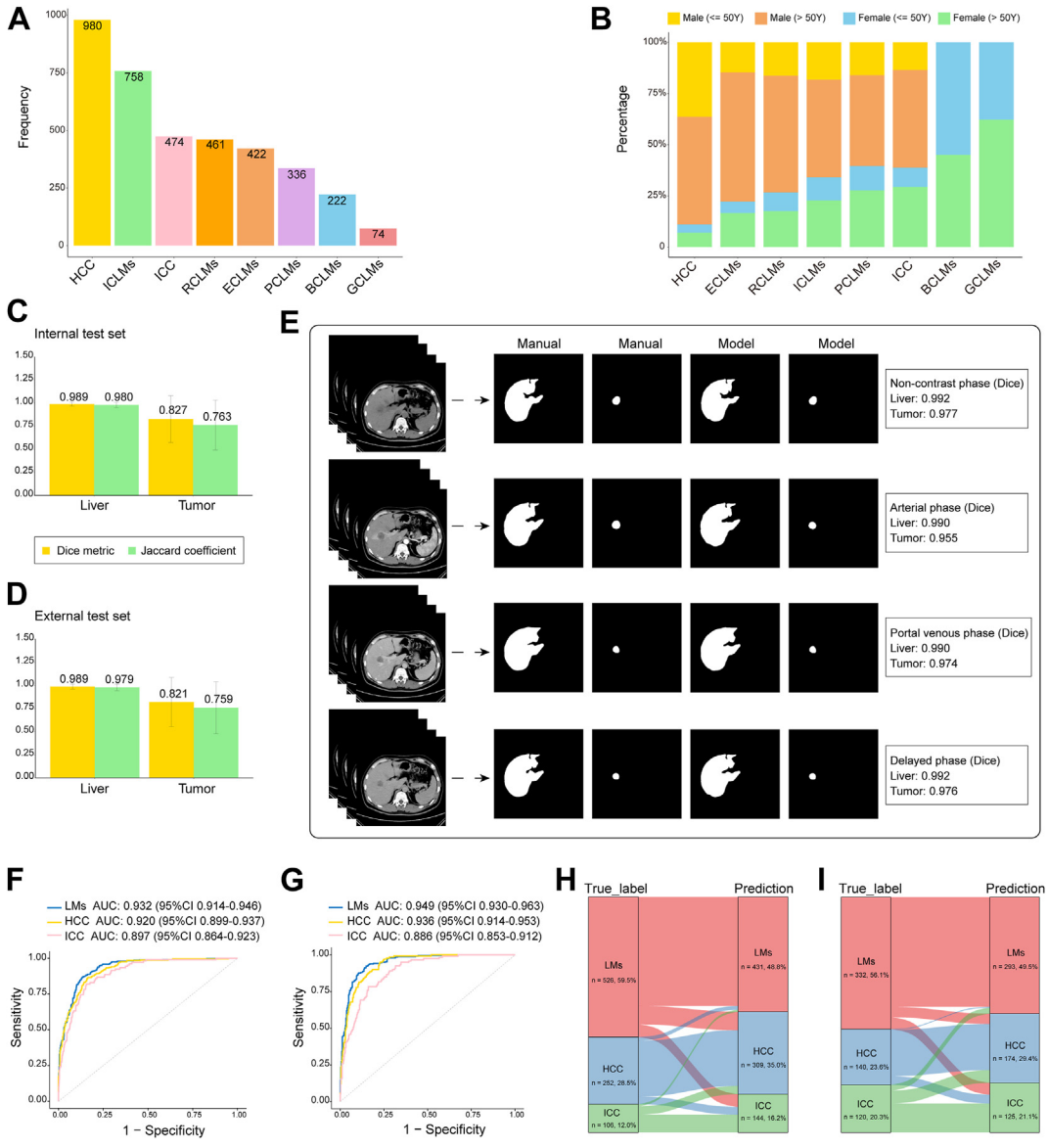


Fig. 4: **Segmentation accuracy and diagnostic performance of artificial intelligence-based liver metastases screening system (ALMSS).** **(A** and **B)** The sample size for each cancer group and their corresponding demographics distribution. **(C** and **D)** Dice metric, Jaccard coefficient of the segmentation model for the liver and tumor in the internal and external test sets. The error bars indicate standard deviation of 0.023, 0.253, 0.030, 0.270, 0.027, 0.266, 0.036, and 0.282, respectively. **(E)** Visualization of segmentation results. **(F** and **G)** Receiver operating characteristic curves of ALMSS for identifying HCC, ICC, and LMs in the internal and external test sets. **(H** and **I)** Sankey diagrams depicting the flow of the assessments of tumor types by ALMSS in the internal and external test sets. The proportions of the diagnoses by pathologists and ALMSS are shown on the left and right sides of each diagram, respectively.

Taxonomy	Organ system	Tumor type	Abbreviation	Total cases
PLC	Liver	Hepatocellular carcinoma	HCC	980
	Liver	Intrahepatic cholangiocarcinoma	ICC	474
ICLMs	Colon	Colon adenocarcinoma	COAD	494
	Rectum	Rectal adenocarcinoma	READ	219
	Small Intestine	Small bowel adenocarcinoma[a]	SBA	16
PCLMs	Pancreas	Pancreatic ductal adenocarcinoma	PDAC	224
	Vater's Ampulla	Periampullary adenocarcinoma[b]	PAAC	35
	Common Bile Duct	Cholangiocarcinoma	CHOL	23
	Gallbladder	Gallbladder adenocarcinoma	GBAD	29
ECLMs	Stomach	Stomach adenocarcinoma	STAD	309
	Gastroesophageal Junction	Gastroesophageal adenocarcinoma	GEAD	15
	Esophagus	Esophageal squamous cell carcinoma	ESCC	59
GCLMs	Ovary	Ovarian serous cystadenocarcinoma	OV	39
	Uterus	Cervical squamous cell carcinoma	CSCC	15
	Uterus	Uterine endometrial adenocarcinoma	UEAD	13
RCLMs	Lung	Lung adenocarcinoma	LUAD	280
	Lung	Lung squamous cell carcinoma	LUSC	36
	Nasopharynx	Nasopharyngeal squamous cell carcinoma	NSCC	55
BCLMs	Breast	Breast adenocarcinoma	BRAD	222

PLC = primary liver cancer. ICLMs = intestinal cancer liver metastases. PCLMs = pancreatobiliary cancer liver metastases. ECLMs = esophagogastric cancer liver metastases. GCLMs = gynecologic cancer liver metastases. RCLMs = respiratory system cancer liver metastases. BCLMs = breast cancer liver metastases. [a]SBA are composed of duodenal (excluding papilla), jejunal, ileal, and appendiceal adenocarcinoma. [b]PAAC arise from the duodenal papilla, distal common bile and pancreatic duct or the structures of the ampullary complex.

Table 1: **Distribution of tumor types in training and testing sets.**

Performance evaluation of the sub-network in ALMSS

The segmentation network demonstrated high accuracy in locating the volumes of the liver and tumor in two independent test sets (Fig. 4C and D), with average Dice metrics of 0.827 (SD 0.253) and 0.821 (SD 0.266), Jaccard coefficients of 0.763 (SD 0.270) and 0.759 (SD 0.282) for tumor segmentation. The corresponding values for the liver segmentation were average Dice metrics of 0.989 (0.023) and 0.989 (0.027), Jaccard coefficients of 0.980 (0.030) and 0.979 (0.036). The visualization representation of the four-phase segmentation results based on an example, demonstrating the superior performance of the sub-network in ALMSS, is shown in Fig. 4E. Furthermore, exploratory studies were conducted to investigate the benefit of using gender and age at metastasis for identifying LMs and predicting its primary origin, as well as the effect of adding global information as an input covariate. The findings are discussed in Supplementary Fig. S1, Supplementary Tables S4, and S5. ALMSS exhibited the best performance in terms of AUC, accuracy, F1 score, sensitivity, specificity, and Brier score when utilizing a combination of global information (liver), local information (tumor), age, and gender as input data.

Application of ALMSS on identifying and subtyping metastatic tumors

On the internal test set, ALMSS achieved an average AUC, average accuracy, F1 score, sensitivity, specificity, and Brier score of 0.917 (95% [CI]: 0.892–0.935), 0.852, 0.732, 0.838, 0.853, and 0.107 for differentiating HCC, ICC, and LMs. The AUC for diagnosing HCC, ICC, and LMs were 0.920 (95% [CI]: 0.899–0.937), 0.897 (95% [CI]: 0.864–0.923), and 0.932 (95% [CI]: 0.914–0.946), respectively (Fig. 4F). Further analysis on the external testing set revealed an average AUC of 0.923 (95% [CI]: 0.900–0.942), average accuracy of 0.858, and F1 score, sensitivity, specificity, and Brier score of 0.761, 0.836, 0.861, and 0.103, respectively. ALMSS yielded AUC of 0.936 (95% [CI]: 0.914–0.953), 0.886 (95% [CI]: 0.853–0.912), and 0.949 (95% [CI]: 0.930–0.963) in HCC, ICC, and LMs, respectively (Fig. 4G). Detailed description and explanation of the discrepant results observed are provided in Fig. 4H and I. The summary of per-cancer type assessment can be found in Supplementary Table S6.

We further validated the capabilities of ALMSS to predict the six predominant origins of metastatic tumors. Our model achieved an average AUC of 0.815 (95% [CI]: 0.763–0.857), with an average accuracy, F1 score, sensitivity, specificity, and Brier score on the internal test set measuring 0.771, 0.529, 0.712, 0.771, and 0.112, respectively. The diagnostic performance for each individual primary origin, with AUC ranging from 0.720 (95% [CI]: 0.673–0.765) to 0.955 (95% [CI]: 0.900–0.986) is displayed in Fig. 5A. Notably, gynecologic cancer exhibited the highest performance. On the external test set, the model attained an average AUC of 0.818 (95% [CI]: 0.753–0.867), alongside average accuracy, F1 score, sensitivity, specificity, and Brier

Characteristics	Training dataset	Internal test dataset	External test dataset	Expandability test dataset
Total. Number of Individuals	2061	884	592	190
Age (years, Mean ± SD)	55.75 ± 11.48	56.02 ± 11.80	56.10 ± 11.27	57.22 ± 12.59
Number of Female (%)	642 (31.1%)	294 (33.2%)	209 (35.3%)	47 (24.7%)
Subtype				
HCC	588	252	140	0
Age (years, Mean ± SD)	53.36 ± 11.30	54.11 ± 11.42	52.74 ± 10.66	NA
Number of Female (%)	61 (10.4%)	32 (12.7%)	15 (10.7%)	NA
ICC	248	106	120	0
Age (years, Mean ± SD)	57.08 ± 10.69	59.00 ± 11.36	59.05 ± 10.08	NA
Number of Female (%)	89 (24.6%)	45 (42.5%)	49 (40.8%)	NA
ICLMs	419	180	130	29
Age (years, Mean ± SD)	56.81 ± 11.73	56.68 ± 12.30	57.06 ± 12.62	55.17 ± 16.51
Number of Female (%)	138 (32.9%)	63 (35.0%)	45 (34.6%)	10 (34.5%)
PCLMs	197	84	30	25
Age (years, Mean ± SD)	57.56 ± 11.62	57.12 ± 11.14	56.83 ± 11.86	51.60 ± 13.07
Number of Female (%)	79 (40.1%)	30 (35.7%)	13 (43.3%)	11 (44.0%)
ECLMs	233	100	50	39
Age (years, Mean ± SD)	59.91 ± 9.95	59.37 ± 10.52	57.14 ± 11.28	57.00 ± 12.23
Number of Female (%)	45 (19.3%)	26 (26.0%)	13 (26.0%)	10 (25.6%)
GCLMs	38	17	12	7
Age (years, Mean ± SD)	52.89 ± 9.35	49.53 ± 12.25	52.92 ± 9.56	51.43 ± 5.00
Number of Female (%)	38 (100%)	17 (100%)	12 (100%)	7 (100%)
RCLMs	204	87	80	90
Age (years, Mean ± SD)	57.05 ± 11.66	56.71 ± 12.96	57.71 ± 10.09	59.98 ± 10.89
Number of Female (%)	58 (28.4%)	23 (26.4%)	32 (40.0%)	9 (10.0%)
BCLMs	134	58	30	0
Age (years, Mean ± SD)	49.43 ± 10.79	50.34 ± 9.89	50.33 ± 8.03	NA
Number of Female (%)	134 (100%)	58 (100%)	30 (100%)	NA

ICLMs = intestinal cancer liver metastases. PCLMs = pancreatobiliary cancer liver metastases. SD = standard deviation. NA = not available. HCC = hepatocellular carcinoma. ECLMs = esophagogastric cancer liver metastases. GCLMs = gynecologic cancer liver metastases. ICC = intrahepatic cholangiocarcinoma. RCLMs = respiratory system cancer liver metastases. BCLMs = breast cancer liver metastases.

***Table 2*: The datasets for training and testing of ALMSS.**

score measuring 0.768, 0.518, 0.713, 0.774, and 0.114, respectively, for the same multi-classification task. Individual performance for each primary origin, with AUC ranging from 0.685 (95% [CI]: 0.613–0.750) to 0.967 (95% [CI]: 0.929–0.988), is shown in Fig. 5B. In addition, a total of 190 LMs cases with 13 pathological types that were not incorporated into model training phase (Supplementary Table S1), comprising 29 small intestine cancers, 90 lung cancers, 39 gastric cancers, 25 pancreatobiliary cancers and 7 gynecologic cancers, were included to evaluate the diagnostic expandability of ALMSS for primary origin (Breast cancer cases were not included, as there was only breast adenocarcinoma in our dataset). The results showed that ALMSS exhibited reliable ability in assigning origin diagnoses for LMs with pathological types beyond adenocarcinoma and squamous cell carcinoma, with an average AUC of 0.761 (95% [CI]: 0.657–0.842). Model predictions for each individual primary origin, with AUC ranging from 0.618 (95% [CI]: 0.512–0.715) to 0.945 (95% [CI]: 0.846–0.997), are illustrated in Fig. 5C, and the consistent results further supported our model's strongest performance in identifying gynecologic cancer. Detailed description and explanation of these discrepant results are provided in Fig. 5D–F. The summary of performance metrics in three test sets is shown in Table 3.

Comparison with benchmark models

On the internal test set and external test set, the average accuracy of ALMSS (0.852 and 0.858) were higher than that of CNN (0.740 and 0.728), HCNN (0.803 and 0.794), and STIC (0.816 and 0.809) for differentiating HCC, ICC and LMs. In terms of predicting origin on metastatic tumors, ALMSS also showed better performance than other models with average accuracy of 0.771 and 0.768, 0.615 and 0.625, 0.656 and 0.658, 0.664 and 0.667 for ALMSS, CNN, HCNN, and STIC on both two test sets, respectively, which approved the capability of our DL framework. The summary of performance metrics of four models in two diagnostic tasks is presented in Table 4.

Radiologists' performance with the assistance of ALMSS

As shown in Supplementary Table S7, ALMSS exhibited greater performance compared to JR and SR, with

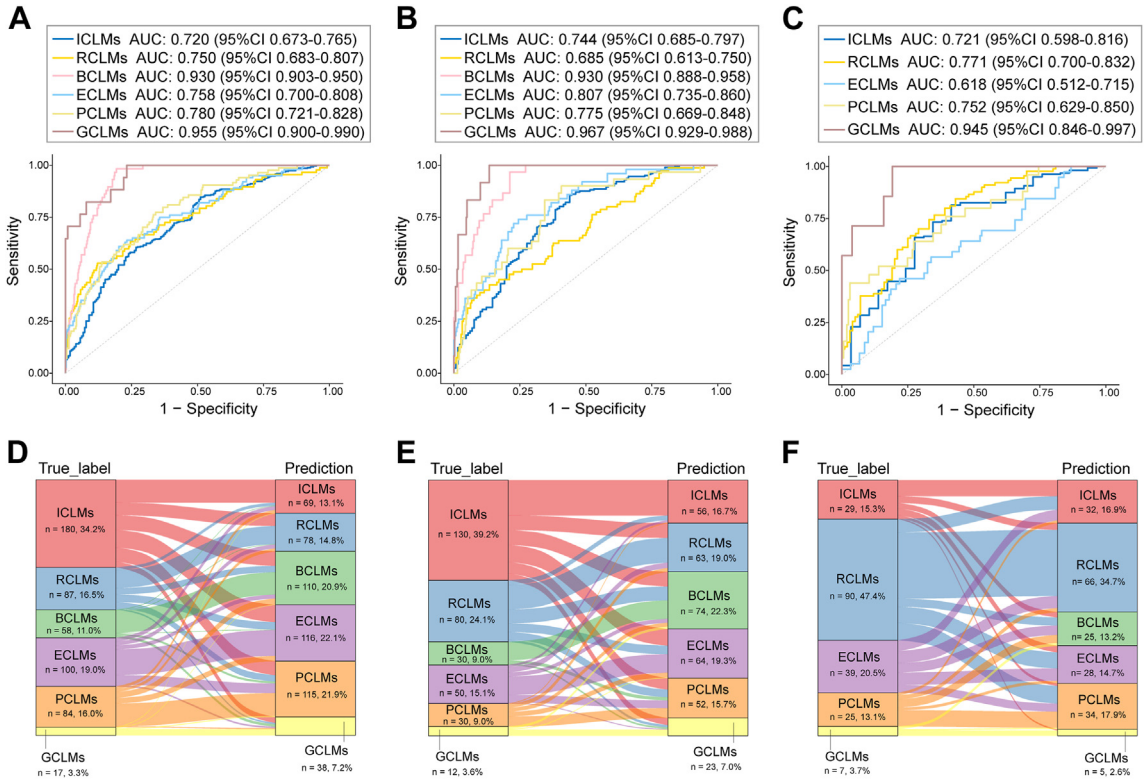


Fig. 5: **The performance of predicting primary origins on liver metastases. (A–C)** Receiver operating characteristic curves of artificial intelligence-based liver metastases screening system (ALMSS) for predicting the primary origins of liver metastases, including intestinal cancer (ICLMs), respiratory system cancer (RCLMs), breast cancer (BCLMs), esophagogastric cancer (ECLMs), pancreatobiliary cancer (PCLMs), and gynecologic cancer (GCLMs) in the internal, external, and expandability test sets. **(D–F)** Sankey diagrams depicting the flow of the assessments of primary origins by ALMSS in the internal, external, and expandability test sets. The proportions of the diagnoses by pathologists and ALMSS are shown on the left and right sides of each diagram, respectively.

average accuracy of 0.852 (95% [CI]: 0.827–0.873), 0.723 (95% [CI]: 0.692–0.751) and 0.817 (95% [CI]: 0.789–0.841), sensitivity of 0.838 (95% [CI]: 0.782–0.884), 0.517 (95% [CI]: 0.454–0.583) and 0.693 (95% [CI]: 0.629–0.752), specificity of 0.853 (95% [CI]: 0.822–0.881), 0.793 (95% [CI]: 0.758–0.826) and 0.859 (95% [CI]: 0.828–0.887) for identifying LMs, HCC, and ICC on the internal test set. In the external test set, ALMSS also reached higher accuracy than JR (0.887 vs. 0.697, $p = 0.0022$) and SR (0.887 vs. 0.777, $p < 0.0001$) for identifying LMs. However, the findings indicated that the accuracy of SR was comparable to that of JR in diagnosing LMs and ICC, which highlights the difficulty in clinically diagnosing these two types of liver tumors via reviewing CT images, often necessitating further confirmation through pathological analysis. The sensitivity and specificity points of two radiologists for LMs, HCC, and ICC in the two test sets are drawn on the same ROC curve in Fig. 6A–C and Supplementary Fig. S4. The diagnostic performance of JR was significantly improved with the assistance of ALMSS, achieving an increase of 9.4% in accuracy, 12.2% in sensitivity, and 8.5% in specificity for diagnosing LMs on the internal test set. The performance of SR for identifying LMs also exhibited improvement under the guidance of ALMSS, with an increase of 6.9% in accuracy, 3.4% in sensitivity, and 12.1% in specificity. Similar results were observed in the external test set. Subsequently, we focused on six cases where prediction errors occurred in the first round of reading by radiologists (two cases pathologically confirmed as LMs, two ICC cases, and two HCC cases), as presented in Fig. 6D. When provided with the results from ALMSS, radiologists maintained their initial judgments for LMs case 1 and ICC case 2, which were incorrectly identified as HCC. Importantly, the predicted probability for each tumor type output by ALMSS effectively assisted radiologists in reconsidering their initial judgments and making a correct diagnosis for the remaining four cases.

Finally, we analyzed the model's auxiliary performance on a subset of challenging LMs cases (n = 76), in which 68.4% (52/76) tumors' primary origins were predicted correctly by ALMSS. For each case, two radiologists reviewed CT images in combination with the laboratory

Origin	Dataset	AUC (95% CI)	Cut-off value	ACC (95% CI)	F1 score (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Brier score (95% CI)
ICLMs	Internal test	0.720	0.20	0.646	0.585	0.728	0.604	0.221
		0.672–0.765		0.605–0.684	0.529–0.638	0.659–0.790	0.553–0.655	0.200–0.245
	External test	0.744		0.693	0.636	0.685	0.698	0.240
		0.685–0.797		0.639–0.738	0.567–0.698	0.600–0.760	0.631–0.759	0.212–0.271
	Expandability test	0.721		0.711	0.409	0.655	0.721	0.122
		0.598–0.816		0.637–0.768	0.282–0.538	0.455–0.815	0.648–0.786	0.099–0.152
RCLMs	Internal test	0.750	0.24	0.709	0.423	0.644	0.722	0.121
		0.683–0.807		0.667–0.745	0.348–0.698	0.538–0.738	0.678–0.762	0.105–0.139
	External test	0.685		0.615	0.439	0.625	0.611	0.168
		0.613–0.750		0.557–0.663	0.357–0.519	0.513–0.726	0.550–0.668	0.144–0.196
	Expandability test	0.771		0.700	0.708	0.767	0.640	0.256
		0.700–0.832		0.626–0.758	0.631–0.777	0.667–0.845	0.537–0.729	0.223–0.291
BCLMs	Internal test	0.929	0.25	0.863	0.571	0.828	0.868	0.066
		0.903–0.950		0.829–0.888	0.480–0.659	0.709–0.911	0.834–0.896	0.054–0.079
	External test	0.930		0.849	0.500	0.833	0.851	0.056
		0.888–0.958		0.804–0.883	0.370–0.614	0.651–0.936	0.808–0.888	0.045–0.071
	Expandability test	NA		NA	NA	NA	NA	NA
ECLMs	Internal test	0.758	0.26	0.736	0.479	0.640	0.758	0.133
		0.700–0.808		0.694–0.770	0.405–0.554	0.543–0.734	0.714–0.796	0.117–0.152
	External test	0.807		0.765	0.480	0.720	0.773	0.109
		0.735–0.860		0.714–0.804	0.379–0.575	0.578–0.830	0.720–0.818	0.092–0.129
	Expandability test	0.618		0.632	0.386	0.564	0.649	0.159
		0.512–0.715		0.558–0.695	0.275–0.504	0.400–0.717	0.571–0.723	0.129–0.194
PCLMs	Internal test	0.780	0.24	0.690	0.428	0.726	0.683	0.117
		0.721–0.828		0.648–0.728	0.357–0.502	0.620–0.815	0.638–0.726	0.103–0.133
	External test	0.775		0.730	0.418	0.667	0.738	0.088
		0.669–0.848		0.672–0.768	0.304–0.535	0.462–0.808	0.686–0.752	0.073–0.106
	Expandability test	0.752		0.695	0.356	0.640	0.703	0.105
		0.629–0.850		0.621–0.753	0.233–0.489	0.429–0.818	0.631–0.762	0.084–0.133
GCLMs	Internal test	0.955	0.40	0.979	0.686	0.706	0.988	0.017
		0.900–0.985		0.962–0.987	0.476–0.842	0.444–0.900	0.975–0.996	0.012–0.027
	External test	0.967		0.961	0.640	0.750	0.972	0.025
		0.929–0.988		0.931–0.976	0.375–0.833	0.375–0.929	0.947–0.985	0.016–0.040
	Expandability test	0.945		0.937	0.455	0.714	0.945	0.019
		0.846–0.997		0.889–0.963	0.182–0.714	0.167–1.000	0.903–0.972	0.008–0.043

ICLMs = intestinal cancer liver metastases. PCLMs = pancreatobiliary cancer liver metastases. BCLMs = breast cancer liver metastases. ECLMs = esophagogastric cancer liver metastases. GCLMs = gynecologic cancer liver metastases. RCLMs = respiratory system cancer liver metastases. NA = not available. CI = confidence interval. AUC = area under the receiver operating characteristic curve. ACC = accuracy.

Table 3: **Diagnostic performance of ALMSS for each origin on liver metastases.**

biochemical results, and pathology reports. In the first round of assessment, the primary origins of 30.3% (23/76) cases were assigned correctly by the radiologists' consensus. When provided with predicted results and the heat maps of ALMSS (Supplementary Fig. S5), we found that the primary origins in 56 out of 76 cases (73.7%) were correctly predicted by the ALMSS-assisted radiologists' consensus (Supplementary Table S8).

Discussion

Currently, available clinical approaches to ascertain the origin of metastatic tumor can be laborious and typically rely on histopathological characteristics or metabolic patterns observed in PET scans. With the rapid integration of machine learning in medical research, several computational algorithms have been developed to evaluate the possible tissue of origin for metastatic tumors, such as CUPLR,[22] OncoNPC,[23] and CUP-AI-Dx.[24] These algorithms represent a paradigm shift in the diagnosis of the origin of cancer. Despite their promising capabilities, it remains uncertain whether these computational methods based on pan-cancer analysis can be effectively applied and fine-tuned for habitat-specific metastatic tumors. More importantly, access to these advanced diagnostic tests (genome or transcriptome analysis) remains uneven, particularly for patients in low-resource settings.

In this study, we presented a 'proof-of-concept' strategy that combined computer vision with

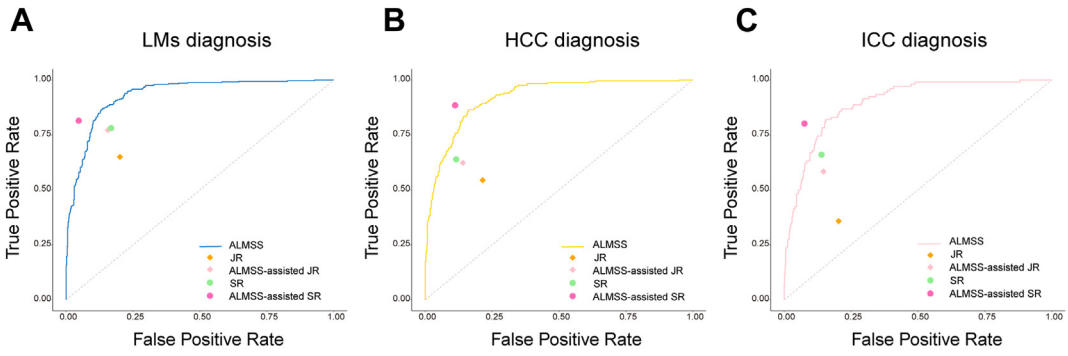
Model	Test set	Liver cancer discrimination						Origin prediction					
		AUC (95% CI)	ACC (95% CI)	F1 score (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Brier score	AUC (95% CI)	ACC (95% CI)	F1 score (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Brier score
CNN	Internal	0.823	0.740	0.600	0.745	0.741	0.162	0.648	0.615	0.318	0.618	0.616	0.132
		0.786–0.854	0.710–0.768	0.556–0.644	0.683–0.800	0.703–0.776	0.153–0.172	0.588–0.704	0.572–0.653	0.253–0.389	0.488–0.725	0.570–0.660	0.115–0.151
	External	0.799	0.728	0.603	0.714	0.727	0.167	0.647	0.625	0.298	0.573	0.628	0.133
		0.757–0.835	0.690–0.761	0.547–0.654	0.642–0.778	0.680–0.770	0.156–0.180	0.566–0.720	0.570–0.672	0.218–0.389	0.397–0.723	0.571–0.683	0.112–0.157
HCNN	Internal	0.871	0.803	0.665	0.795	0.805	0.160	0.708	0.656	0.348	0.638	0.659	0.127
		0.841–0.896	0.775–0.827	0.621–0.707	0.736–0.846	0.770–0.836	0.150–0.171	0.650–0.762	0.613–0.694	0.280–0.421	0.510–0.749	0.613–0.702	0.112–0.146
	External	0.855	0.794	0.686	0.793	0.794	0.159	0.695	0.658	0.337	0.638	0.664	0.130
		0.820–0.884	0.759–0.823	0.635–0.734	0.727–0.849	0.751–0.831	0.147–0.173	0.610–0.766	0.604–0.704	0.250–0.432	0.454–0.771	0.607–0.717	0.110–0.153
STIC	Internal	0.899	0.816	0.688	0.823	0.817	0.122	0.722	0.664	0.363	0.658	0.663	0.125
		0.871–0.922	0.789–0.839	0.646–0.729	0.766–0.870	0.783–0.846	0.113–0.132	0.667–0.772	0.622–0.701	0.293–0.437	0.529–0.763	0.618–0.705	0.109–0.143
	External	0.880	0.809	0.703	0.798	0.808	0.128	0.722	0.667	0.359	0.685	0.666	0.125
		0.848–0.906	0.776–0.837	0.653–0.751	0.733–0.854	0.767–0.845	0.116–0.141	0.644–0.789	0.612–0.712	0.273–0.452	0.501–0.811	0.609–0.720	0.105–0.148
ALMSS	Internal	0.917	0.852	0.732	0.838	0.853	0.107	0.815	0.771	0.529	0.712	0.771	0.112
		0.892–0.935	0.827–0.873	0.687–0.773	0.782–0.883	0.822–0.881	0.097–0.119	0.763–0.857	0.734–0.800	0.432–0.616	0.586–0.815	0.732–0.805	0.098–0.129
	External	0.923	0.858	0.761	0.836	0.861	0.103	0.818	0.768	0.519	0.713	0.774	0.114
		0.899–0.942	0.826–0.882	0.710–0.805	0.773–0.886	0.823–0.893	0.092–0.117	0.753–0.867	0.719–0.805	0.392–0.629	0.530–0.831	0.724–0.817	0.097–0.135

CNN = Convolutional Neural Network. HCNN = Hierarchical CNN. STIC = SpatialExtractor-TemporalEncoder-Integration-Classifier. AUC = area under the receiver operating characteristic curve. ACC = accuracy. CI = confidence interval.

Table 4: **The performance of ALMSS comparison with existing benchmark models.**

conventional imaging approach to evaluate the possible origin of LMs. Our model can be particularly valuable in multidisciplinary management settings, as it provides information and helps formulate a targeted search pattern. For example, clinicians may perform an upper endoscopy to evaluate the foregut, including the esophagus, stomach, and duodenum. Additionally, a bronchoscopy can be utilized to inspect the distant parts of the lung, while a pelvic magnetic resonance imaging (MRI) can aid in identifying the primary tumors in the reproductive system. Such an AI assistance system may pose opportunity for applying site-specific therapies and bring survival benefits to patients with difficult-to-diagnose LMs as clinical responses to certain treatments exhibit variability based on patients' tumor types. Notably, patients with LMs originating from lung cancer demonstrate a more favorable response to immune checkpoint inhibitors, while LMs derived from primary tumors in the digestive tract have been established as insensitive to such treatment.[5,6,25] Additionally, a substantial subset of LMs arising from colorectal cancer may benefit from surgical resection to mitigate hepatic tumor burden, whereas LMs from other sources are considered unsuitable for surgical intervention, necessitating the implementation of molecular-targeted therapies, such as HER-2 targeting in breast cancer.[26] On the other hand, distinguishing newly-emerging HCC or ICC from metastases of an antecedent malignancy is of paramount importance. For small isolated lesions, local therapeutic interventions, such as thermal ablation, may be recommended for HCC or ICC, but not applicable to LMs.[27] Furthermore, precise diagnosis of ICC can also guide the judicious selection of portal lymphadenectomy, which is not advocated in HCC during surgery.[27] The American College of Radiology has introduced LI-RADS in the past few years, to standardize the reporting process of CT imaging, primarily for the diagnosis of HCC without pathological confirmation.[28] While LI-RADS designated an LR-M category for observations that are definitely malignant (except HCC), it falls short in specifically differentiating ICC and LMs, which restricts its clinical applicability. The integration of ALMSS into LI-RADS workflow can be particularly valuable for the noninvasive subtyping of liver cancer.

ALMSS incorporated multi-scale features to accurately predict the origin of distinct liver tumors. Local features derived from tumors themselves, like morphology or vascular enhancement pattern, are traditionally associated with their respective origins. Beyond that, doctors typically review liver tumors within their broader context, recognizing that certain global features can offer valuable insights. For instance, liver cirrhosis of various etiologies stands as the primary risk factor for HCC development, whereas biliary duct cysts, hepatolithiasis, and parasitic biliary infection are prevalent in ICC.[29,30] In this study, we emphasize the importance of considering global features that provide



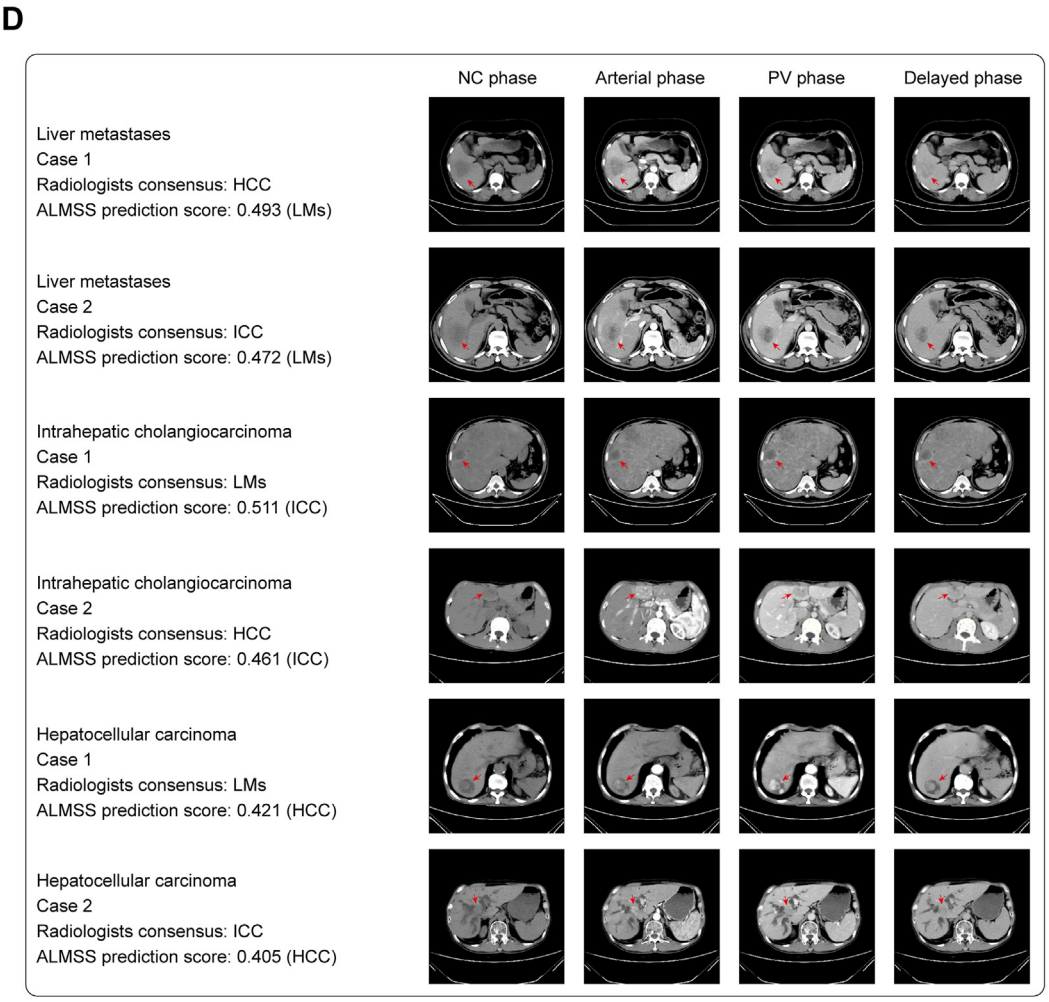


Fig. 6: **The auxiliary performance of artificial intelligence-based liver metastases screening system (ALMSS) in real-world scenario**. **(A–C)** The sensitivity and specificity of the radiologists with and without the assistance of ALMSS in identifying LMs, HCC, and ICC in the internal test set. **(D)** The case illustration of six typical samples misdiagnosed by radiologists. In LMs case 1 and ICC case 2, the enhancement patterns observed were typical for HCC samples, which involved non-rim-like hyper-enhancement in the arterial phase and wash-out in the portal-venous phase. LMs case 2, exhibited central enhancement in the delayed phase. ICC case 1 presented with multi-focal liver lesions with rim hyper-enhancement in the arterial phase. Additionally, the patient had concurrent suspected lung cancer findings reported in a chest CT examination. HCC case 1, who was undergoing pancreatic cancer treatments at the medical center, displayed abnormal nodular enhancement in the portal-venous phase. In HCC case 2, which exhibited a liver lesion with an infiltrative appearance and had a homogeneous and progressive enhancement pattern. Simultaneously, there was a dilation of intrahepatic bile ducts. JR = junior radiologist. SR = senior radiologist.

information about intrahepatic spatial distribution between lesions and the pathological changes of the whole liver. Notably, the age at which LMs are diagnosed can provide some indication of the primary origin. Data-driven analysis has shown that breast cancer is the most common primary origin for LMs in young women aged 20–50 years, while colon cancer is more prevalent in young men.[31] Building upon these findings, ALMSS leverages global information, gender and age at diagnosis as learnable features in addition to local information, resulting in improved performance.

Regarding the potential global information that could explain the source of various LMs, we speculated that the intrahepatic spatial distribution pattern of metastatic tumors may serve as an indicator. Although our results are not sufficient to fully support this assumption, we have gathered some indirect evidence from several aspects. Firstly, when metastatic tumors from different primary sites within the body arise in the liver, they may share similar morphology but possess distinct and identifiable spatial distribution patterns.[11] This distinction arises from the fact that cancer cells metastasize to the liver through specific pathways, including direct invasion, blood spread, lymphatic metastasis, and peritoneal dissemination. For instance, ovarian cancer commonly metastasizes to the liver through peritoneal dissemination, often presenting as perihepatic metastases on CT imaging.[32] Moreover, several studies had explored the possibility that intrahepatic metastatic lesions have heterogeneous spatial distribution patterns influenced by the streamline flow in the portal vein.[33,34]

One of the key limitations of ALMSS is its lack of interpretability. While we tried to visualize the region of interest in model prediction, it was hard to confirm whether the highlighted region was a true clue or an occasional error. Our model still does not provide explicit information about the factors that contribute to its decision-making process. Selection bias is another important limitation in this study. The entire study was conducted on retrospective data and restricted to the population in the southern areas of China. Thus, large-scale prospective studies involving diverse populations should be implemented in the future. Additionally, it should be noted that our model primarily focused on CT-positive liver tumors. This means that its applicability to CT-negative cases, which are confirmed by other modalities such as MRI, may be limited. In this work, we excluded samples with any missing or defective data, which could potentially contribute to selection bias. We aim to address the issue of missing data in enrolled cases in future prospective studies. There is also a lack of subjects with LMs from rare origins that do not preferentially metastasize to the liver but rather to the lung, such as glioblastoma, osteosarcoma, and melanoma. Such cases should be taken into consideration in further investigation to provide a more comprehensive origin network and improve the model's applicability across different metastatic sites. Finally, it is imperative to acknowledge the potential impact of the multiplicity issue on the results obtained in this exploratory study, particularly when dealing with multiple comparisons. More appropriate multiplicity adjustments in statistical analyses should be incorporated into future validation studies to ensure the robustness and reliability of our findings.

In conclusion, our study presents the first proof-of-concept of using a deep-learning framework to differentiate LMs from primary liver cancer and accurately classify their primary origins directly in liver CECT images. The results highlight the potential of AI systems in augmenting conventional diagnostic methods, mitigating the shortage of clinical experts to provide timely referral of patients to the appropriate clinical care.

Contributors

HX, YWZ, WY and YPZ conceived and designed the study. HX, QL, NL, PH, YL, JH, JL and YCZ did the data collection. QL and NL performed image annotations. YL, ZC and JZ reviewed images and performed diagnostic analysis. YWZ and WY contributed to development of methodology. HX, YWZ, and PH performed statistical and computational analysis of data. HX and YWZ wrote the original draft manuscript. WY and YPZ reviewed and performed revision of the manuscript. WY and YPZ had full access to all the data and verified the underlying data. All authors read and approved the final manuscript for publication.

Data sharing statement

To protect patient privacy, CT image datasets and other patient-related data are not publicly accessible, but all data are available upon reasonable request emailed to the corresponding author. To gain access, data requestors will need to sign a data-access agreement. The algorithm source codes used in this study are fully available at https://github.com/whisney/ALMSS.

Declaration of interests

We declare no competing interests.

Acknowledgements

This study was supported by the grants from the National Natural Science Foundation of China (81772923, U21A6005) and the Guangdong Provincial Key Laboratory of Medical Image Processing (2020B1212060039). We sincerely appreciate the great support from www.biorender.com for figures drawing.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2024.102464.

References

1 Azubuike UF, Tanner K. Biophysical determinants of cancer organotropism. *Trends Cancer*. 2023;9:188–197.
2 Tsilimigras DI, Brodt P, Clavien PA, et al. Liver metastases. *Nat Rev Dis Primers*. 2021;7:27.
3 Yu J, Green MD, Li S, et al. Liver metastasis restrains immunotherapy efficacy via macrophage-mediated T cell elimination. *Nat Med*. 2021;27:152–164.
4 Tumeh PC, Hellmann MD, Hamid O, et al. Liver metastasis and treatment outcome with anti-PD-1 monoclonal antibody in patients with melanoma and NSCLC. *Cancer Immunol Res*. 2017;5:417–424.
5 Wang C, Sandhu J, Ouyang C, et al. Clinical response to immunotherapy targeting programmed cell death receptor 1/programmed cell death ligand 1 in patients with treatment-resistant microsatellite stable colorectal cancer with and without liver metastases. *JAMA Netw Open*. 2021;4:e2118416.

6 Ciner AT, Jones K, Muschel RJ, et al. The unique immune microenvironment of liver metastases: challenges and opportunities. *Semin Cancer Biol*. 2021;71:143–156.
7 Hayashi H, Kurata T, Takiguchi Y, et al. Randomized phase II trial comparing site-specific treatment based on gene expression profiling with carboplatin and paclitaxel for patients with cancer of unknown primary site. *J Clin Oncol*. 2019;37:570–579.
8 Culine S, Lortholary A, Voigt JJ, et al. Cisplatin in combination with either gemcitabine or irinotecan in carcinomas of unknown primary site: results of a randomized phase II study–trial for the French Study Group on Carcinomas of Unknown Primary (GEFCAPI 01). *J Clin Oncol*. 2003;21:3479–3482.
9 Briasoulis E, Kalofonos H, Bafaloukos D, et al. Carboplatin plus paclitaxel in unknown primary carcinoma: a Phase II Hellenic Cooperative Oncology Group Study. *J Clin Oncol*. 2000;18:3101–3107.
10 Kato S, Alsafar A, Walavalkar V, et al. Cancer of unknown primary in the molecular era. *Trends Cancer*. 2021;7:465–477.
11 Lincke T, Zech CJ. Liver metastases: detection and staging. *Eur J Radiol*. 2017;97:76–82.
12 Cho MJ, An C, Aljoqiman KS, et al. Diagnostic performance of liver imaging reporting and data system in patients at risk of both hepatocellular carcinoma and metastasis. *Abdom Radiol (NY)*. 2020;45:3789–3799.
13 Yasaka K, Akai H, Abe O, et al. Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. *Radiology*. 2018;286:887–896.
14 Chen X, Wang X, Zhang K, et al. Recent advances and clinical applications of deep learning in medical image analysis. *Med Image Anal*. 2022;79:102444.
15 Shang H, Li J, Jiao T, et al. Differentiation of lung metastases originated from different primary tumors using radiomics features based on CT imaging. *Acad Radiol*. 2023;30:40–46.
16 Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527.
17 Isensee F, Jaeger PF, Kohl SAA, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18:203–211.
18 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9:1735–1780.
19 Zhou J, Wang W, Lei B, et al. Automatic detection and classification of focal liver lesions based on deep convolutional neural networks: a preliminary study. *Front Oncol*. 2020;10:581210.
20 Gao R, Zhao S, Aishanjiang K, et al. Deep learning for differential diagnosis of malignant hepatic tumors based on multi-phase contrast-enhanced CT and clinical data. *J Hematol Oncol*. 2021;14:154.
21 Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: *2017 IEEE international conference on computer vision (ICCV), 22-29 Oct. 2017*. 2017.
22 Nguyen L, Van Hoeck A, Cuppen E. Machine learning-based tissue of origin classification for cancer of unknown primary diagnostics using genome-wide mutation features. *Nat Commun*. 2022;13:4013.
23 Moon I, LoPiccolo J, Baca SC, et al. Machine learning for genetics-based classification and treatment response prediction in cancer of unknown primary. *Nat Med*. 2023;29:2057–2067.
24 Zhao Y, Pan Z, Namburi S, et al. CUP-AI-Dx: a tool for inferring cancer tissue of origin and molecular subtype using RNA gene-expression data and artificial intelligence. *eBioMedicine*. 2020;61:103030.
25 Deng JY, Gou Q, Yang L, et al. Immune suppressive microenvironment in liver metastases contributes to organ-specific response of immunotherapy in advanced non-small cell lung cancer. *J Immunother Cancer*. 2023;11:e007218.
26 Rassy E, Pavlidis N. Progress in refining the clinical management of cancer of unknown primary in the molecular era. *Nat Rev Clin Oncol*. 2020;17:541–554.
27 Petrowsky H, Fritsch R, Guckenberger M, et al. Modern therapeutic approaches for the treatment of malignant liver tumours. *Nat Rev Gastroenterol Hepatol*. 2020;17:755–772.
28 Chernyak V, Fowler KJ, Kamaya A, et al. Liver imaging reporting and data system (LI-RADs) version 2018: imaging of hepatocellular carcinoma in at-risk patients. *Radiology*. 2018;289:816–830.
29 Vogel A, Meyer T, Sapisochin G, et al. Hepatocellular carcinoma. *Lancet*. 2022;400:1345–1362.
30 Brindley PJ, Bachini M, Ilyas SI, et al. Cholangiocarcinoma. *Nat Rev Dis Primers*. 2021;7:65.
31 Horn SR, Stoltzfus KC, Lehrer EJ, et al. Epidemiology of liver metastases. *Cancer Epidemiol*. 2020;67:101760.
32 Akin O, Sala E, Moskowitz CS, et al. Perihepatic metastases from ovarian cancer: sensitivity and specificity of CT for the detection of metastases with and those without liver parenchymal invasion. *Radiology*. 2008;248:511–517.
33 Hata Y, Kondoh M, Nakajima Y, et al. Streamline phenomena in liver metastasis of gastrointestinal tumors - a clinical-study based upon 172 patients. *Oncol Rep*. 1995;2:593–595.
34 Ambrosetti MC, Zamboni GA, Mucelli RP. Distribution of liver metastases based on the site of primary pancreatic carcinoma. *Eur Radiol*. 2016;26:306–310.