Hepatology International (2021) 15:868–880
https://doi.org/10.1007/s12072-021-10229-z

REVIEW ARTICLE

The overview of the deep learning integrated into the medical imaging of liver: a review

Kailai Xiang[1,2] · **Baihui Jiang**[3] · **Dong Shang**[1,2]

Received: 9 March 2021 / Accepted: 24 June 2021 / Published online: 15 July 2021

Abstract

Deep learning (DL) is a recently developed artificial intelligent method that can be integrated into numerous fields. For the imaging diagnosis of liver disease, several remarkable outcomes have been achieved with the application of DL currently. This advanced algorithm takes part in various sections of imaging processing such as liver segmentation, lesion delineation, disease classification, process optimization, etc. The DL optimized imaging diagnosis shows a broad prospect instead of the pathological biopsy for the advantages of convenience, safety, and inexpensiveness. In this paper, we reviewed the published representative DL-related hepatic imaging works, described the general situation of this new-rising technology in medical liver imaging and explored the future direction of DL development.

Keywords Deep learning · Artificial intelligence · Convolutional neural network · Liver disease · Ultrasonography · Computed tomography · Magnetic resonance imaging · Imaging diagnosis · Image segmentation · Lesion classification

Introduction

Liver is the largest parenchymatous organ and also one of the crucial metabolic apparatus of human beings [1]. This vital organ is rather vulnerable and always suffers from various diffuse or focal lesions. Although biopsy is still regarded as the reference standard for the majority of hepatic diseases' diagnoses for now, the screening and detection of liver diseases actually depend heavily on imaging examinations for the advantages of convenience, noninvasiveness, and low cost, etc. [2, 3]. The medical images provide independent and accurate evidence for liver diseases that could be comprehensively considered with clinical symptoms and laboratory examination results. The ultrasound (US), computed tomography (CT), and magnetic resonance imaging (MRI) are frequently used modalities as cornerstones of imaging examination [4–8]. However, the evaluation of the images relies on radiologists' professional knowledge, which is of low sensitivity and likely exhibits observer bias [9]. In addition, the visual inspection for huge number of medical scans is labor-intensive and time-consuming [8]. In recent years, the computer-aided diagnosis (CAD) systems combined with computer vision (CV) and machine learning have been developed prosperously for this problem through facilitating the interpretation of imaging, instead of manual segmentation and detection on the images [10, 11]. Differing from the conventional machine learning methods which are basically based on hand-crafted visual descriptors and usually not optimal for specific tasks, the DL algorithm, as the core content of the artificial intelligence, is gradually applied in CAD, obviating the necessity of manual customization (Figs. 1, 2) [12, 13].

In this review, we introduced the DL technology briefly, illustrated the current developments of DL integrated in the liver imaging based on previously published works, and explored the future direction of DL application.

✉ Dong Shang
shangdong@dmu.edu.cn

Kailai Xiang
2273649899@qq.com

Baihui Jiang
2872643400@qq.com

[1] Department of General Surgery, First Affiliated Hospital of Dalian Medical University, Dalian 116011, Liaoning, China

[2] Clinical Laboratory of Integrative Medicine, First Affiliated Hospital of Dalian Medical University, Dalian 116011, Liaoning, China

[3] Department of Ophthalmology, First Affiliated Hospital of Dalian Medical University, Dalian 116011, Liaoning, China

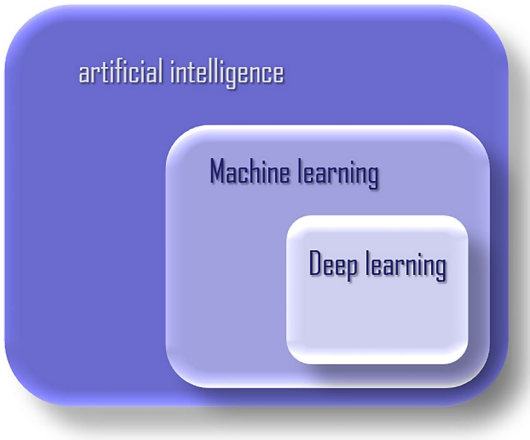

Fig. 1 The relationship of DL-related concepts. The machine learning is the core method to implement artificial intelligence, and the DL is a new-rising algorithm belonging to machine learning that benefits from the improvement of computing power

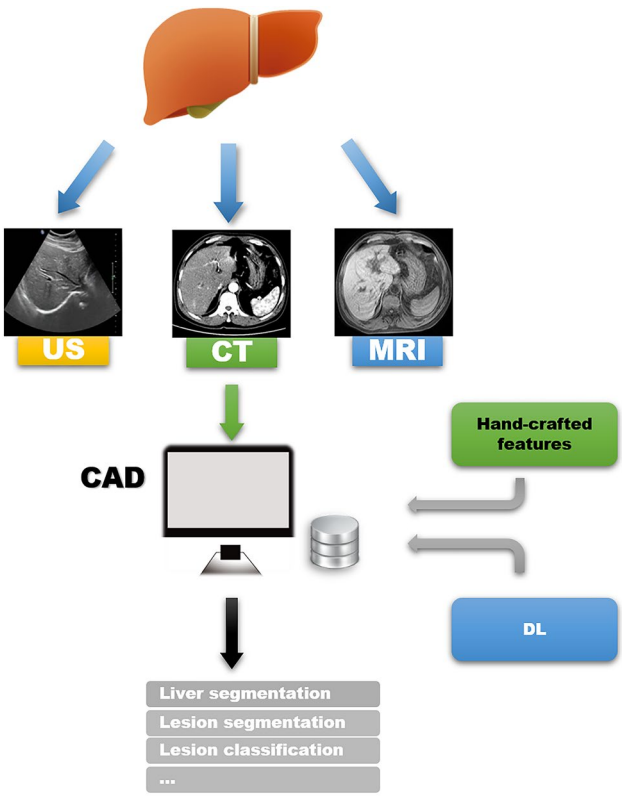

Fig. 2 Flow chart of the general liver imaging processing with current technologies. The diversity imaging modalities of US, CT, and MRI are applied for screening and diagnosis of liver diseases. The scans are delivered to CAD where conventional machine learning and burgeoning DL methods are integrated to investigate the inputs. The overall system deals with various tasks such as liver segmentation, lesion segmentation, lesion classification, etc.

The structures, mechanisms and different types of DL models

DL replaces the hand-crafted feature filters with hierarchical architecture of representation learning which processes the input data through several transformation layers with nonlinear elements. The features of the input data are learned and represented in the framework layer-by-layer, in which the high-level features are derived from the lower ones [14, 15].

There are various architectures of DL, such as convolutional neural network (CNN), auto-encoder (AE), deep belief network (DBN) and recurrent neural network (RNN) [16–19]. These diverse frameworks have become the current topic in most fields like natural language processing (NLP), CV and bioinformatics [20, 21]. CNN is mainly used for image-related tasks. In virtue of convolution layers, rectilinear units and pooling layers sandwiched between input and output layers, CNN models are able to represent the feature maps of input data extracted by convolution kernels or compressed by pooling operation in corresponding layers (Fig. 3) [23]. The convolution is a weighted average operation that the convolution kernel slides over the image to generate the feature map [24]. Subsequently, the activation function is applied to bring in nonlinearity and increase fitting ability for the model, reducing the likelihood of the gradient vanishing concurrently [25]. In addition, the feature maps are pooled in pooling layers to accomplish down sampling and compression of features (Fig. 4). The ordered and repeated arrangement of these primary elements forms the basic structure of a CNN model. And the parameters of the networks are self-optimized by gradient descent and back propagation algorithms as the training goes on [26].

In particular, the Fully convolutional network (FCN), which can take arbitrary size of input image and produce corresponding size output since all the fully connected layers in the framework are replaced by convolution layers, is prevailing in various medical imaging tasks as a variant of CNN [27]. With FCN, images could be analyzed globally instead of using patch-based methods and the details of original images are restored. At present, the U-Net architecture is the most popular form of FCN which consists of two symmetric parts that compress and unfold the location information of pixels together with the image features [28].

The training process of the DL model relies on numerous annotated data. If merely trained on limited medical data, DL models tend to suffer from both under-fitting and over-fitting problems. Generally, transfer learning is applied to medical image training of small databases. The transfer learning refers to pre-training the model on

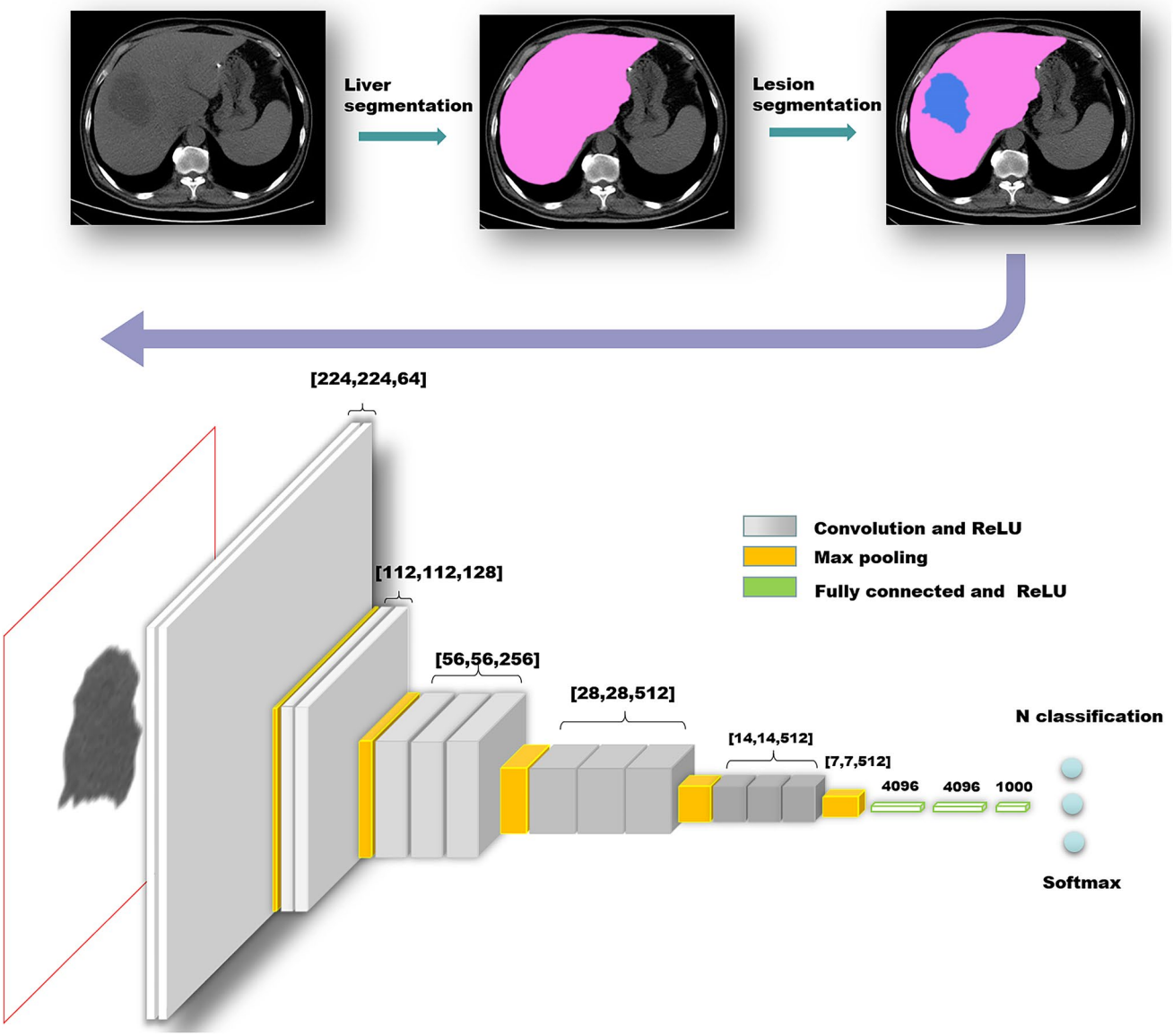

Fig. 3 A liver lesion classification task with a popular CNN framework named VGG-16. In this illustration, the segmented liver lesion from an abdominal CT scan passes through 13 convolution layers with activation function of ReLU and 5 max pooling layers in VGG-16 framework, and the results are further connected to 3 fully connected layers and the final adjusted softmax layer to solve the corresponding n-classification task [22]

existing large databases of natural or task-related images, and the obtained network is further trained on personalized database since the features of the former layers are similar in diverse image tasks empirically. Besides, transfer learning is always combined with fine-tuning which only adjusts the last few layers to acquire a rather quick but accurate result [29, 30].

DL methods for liver segmentation

Generally, the abdominal scans involve other organs besides the liver, while most of the medical liver imaging tasks need to be processed on separate liver technologically. In these years, the model-based segmentation

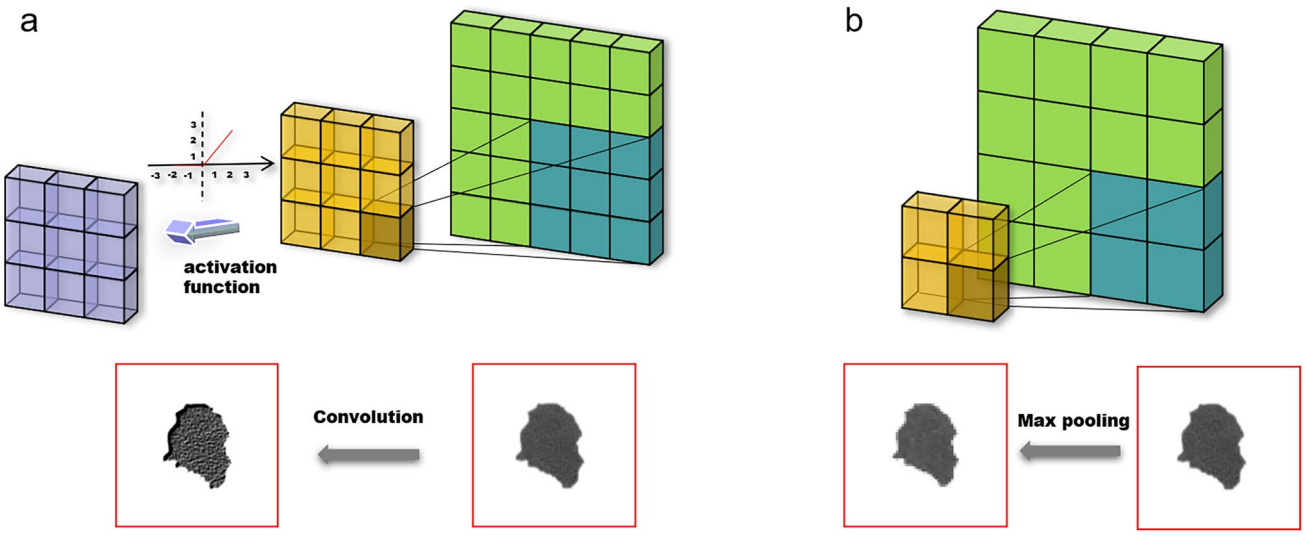

Fig. 4 The convolution with activation function and pooling operation. **a** The convolution kernel slides over the image by certain stride to extract the feature map. The products of image elements and corresponding kernel elements are summed up as the convolution results, which are further processed by activation function to introduce nonlinearity. **b** The max and average poolings are common pooling methods, which represent the 2 × 2 non-overlapping pixels with their maximum or mean values

approaches are gradually developed and taking the main position with extraordinary accuracy and fast processing speed in virtue of DL.

FCN is the preferred model for liver segmentation task. The work of Ben-Cohen et al. [31] proposed VGG-16 based FCN architectures for liver segmentation and further metastatic lesion detection on CT examinations. Input axial CT slices were merged with two adjacent slices to introduce the spatial information in pretreatment. The output layers of the models are fused with the former layers to unite the different scales and restore the visual details close to origin images. As a result, the so-called FCN-8 s architecture with adjacent slices achieved a Dice index of 0.89, an average sensitivity of 0.86 and an average positive predictive value of 0.95 for liver segmentation. Zhang et al. [32] presented an elaborate coarse-to-fine method for liver segmentation in another form, in which the region of interest (ROI) of the liver is primarily delineated with a 2D U-net, and further refined by a 3D FCN, providing rough tumor contours simultaneously. The method was evaluated on an International Symposium on Image Computing and Digital Medicine (ISICDM) dataset and the MICCAI 2017 Liver Tumor Segmentation (LiTS17) Challenge database with dice similarity coefficients (DSCs) of 96.31% and 96.7%, respectively, and was proven to be efficient in reducing computation time and superior to other state-of-the-art methods.

Organ segmentation is also extensively used in image-guided operations. For instance, CT-guided biopsy and ablation rely on interventional scans to observe the spatial relationship of equipment and organs. It would be safe to fuse the interventional image with preprocedural diagnostic image when lesions are not directly visible in interventional image. Thus, Fang et al. [33] constructed a multi-scale input and multi-scale output feature abstraction network (MIMO-FAN) architecture for liver segmentation and applied the segmented surfaces to register the interventional CT with preprocedural image. Meanwhile, the same system was proven to be well-worked on registration of positron emission tomography-CT (PET-CT), contrast-enhanced CT (CE-CT) and MRI.

In selective internal radiation therapy (SIRT), an accurate liver segmentation is the guarantee of dose calculation. For example, an automatic liver segmentation method modified by 3D DeepMedic was developed by Tang et al. [34] on both CE-CT and plain CT scans for SIRT patients. The CNN model adopted is a modified dual-pathway framework, which hybridizes the common patch-wise training and dense training on full images. The median DSC between the model and manual segmentation was 0.94 in test cohort, and the accuracy could be further improved by manual adjustment.

And the liver delineation is a significant preoperative preparation in 3D imaging assisted liver surgery planning. A 3D segmentation model combined with globally optimized surface evolution algorithm was constructed by Hu et al. [35] for this task. In the two steps of the framework, the 3D CNN predicts the probability map of liver primarily, and the result is further refined by an energy-functional model. Considering the differences of liver appearances, the global and local information of the initial segmentation are adaptively integrated into the energy function, which is formulated in a

hybrid model that includes region statistics, shape prior constraint as well as gradient-edge map. With the minimization of the energy function, the predicted surface is converged to its optimal position. This method was evaluated on public Segmentation of the Liver Competition 2007 (SLIVER07) and local databases. On the SLIVER07 testing set, the framework achieved a mean DSC of 97.25% and an average symmetric surface distance of 0.84 mm which were 97.28% and 0.87 mm, respectively for the local database.

Besides the overall outline of liver, the liver volume is also a vital value in clinic of which the analysis is extensively applied in acute liver failure, treatment planning and response evaluation, and hepatic surgery [36]. Combined with other clinical information, Marinelli et al. [37] adopted a weakly supervised active transfer learning method to construct the DL model for liver segmentation. The liver volume extracted from clinic report by NLP substitutes traditional Dice score to evaluate segmentation result in this method. A 3D U-Net was trained on the LiTS17 and further applied on the Mount Sinai Health System database. The mispredicted cases were combined with the original LiTS17 database to form new training sets, among which the LiTS17 with over- and under-estimated cases produced the best model with absolute volume difference of 176 ml. To a certain extent, adopting the liver volume value for liver segmentation is beneficial to the clinical application of DL.

A significant question for DL architectures is that the model is determined by several hyper-parameters so that searching through parameter space is crucial for training. Ghoniem [38] combined the artificial bee colony (ABC) algorithm, a bio-inspired optimization algorithm, to the established DL models for liver segmentation and lesion classification on CT images. The model adopts SegNet for liver segmentation, and applies U-Net to extract lesions from the liver, with ABC algorithm integrated to fine-tune the hyper-parameters. This approach was tested on LiTS17 and another Radiopaedia database. Of note, the SegNet encompasses an encoder–decoder architectural engine ended at a pixel-wise classification layer. And the hyper-parameters such as initial learning rate, maximum epochs and minibatch size are optimized by ABC, in which the optimal parameters are regarded as "foods of the best quality" and the bee agents represent a group of computational agents. The proposed model achieved 0.96, 0.968 and 0.962 in terms of the Jaccard index, Dice index and correlation coefficient on the Radiopaedia dataset, and 0.964, 0.97, 0.958 for the three metrics on LiTS17 database.

The various segmentation tasks of this section are listed in Table 1. Most of the cases were trained on public databases and tested on private ones. Several conventional machine learning methods were combined to support the segmentations. In general, the majority of the constructed models exhibited extraordinary performances in this upstream task.

DL methods applied in detection, segmentation and classification of liver focal lesions (FLLs)

Current CAD systems are mostly based on traditional feature engineering, which has several drawbacks such as redundant features and large computational cost [39]. It is difficult to design hand-crafted feature filters optimal for specific tasks. For the classification of diverse liver tumors, Yamakawa et al. [40] constructed a VGG-based model on US images. Accuracy of the model by lesion type was 98.1% for cyst, 86.8% for hemangioma, 86.3% for hepatocellular carcinoma (HCC) and 29.2% for metastatic liver cancer. Meanwhile, the group parallelly established a binary classification model for malignant lesions identification, of which the accuracy, sensitivity and specificity were 94.8%, 93.8% and 95.2%, respectively. Schmauch et al. [41] created a DL algorithm that simultaneously detects and identifies FLLs from liver US images. The attention mechanism is integrated into the framework and a 50-layer ResNet, pretrained on the ImageNet, was adopted as the backbone of the framework. In tradition CNN, each position of the image shares the same weight in feature extraction. The attention mechanism of the model helps to locate the focus on the decisive pixels. Finally, the model reached a weighted mean area under the receiver operating characteristic curve (ROC-AUC) of 0.891 for seven different tasks on the testing cohort.

For the numerous unmarked medical images, an unsupervised algorithm is favorable to make full use of the data instead of training only on the labeled ones. Deep auto-encoder is a specific unsupervised DL framework to extract latent patterns of unlabeled data, in which the input data are compressed through the hidden layers and restored to itself in the output layer. The latent comprehensive features of the input are saved in the form of dimensional reduction. After the lesion segmentation by the level set and fuzzy c-means clustering methods, Hassan et al. [42] applied a stacked sparse auto-encoder (SSAE) model to pre-train and learn the feature extraction method of the various unlabeled FLLs US images. Iteratively, a sequence of auto-encoders were trained and connected to the classifier network for identification of the different focal liver diseases. The SSAE system surpassed other three comparative models with overall classification accuracy of 97.2%.

The contrast-enhanced ultrasound (CEUS) is developed as an enhanced imaging examination, which is supported by the contrast agent injected from peripheral veins [43, 44]. The CEUS visualizes the enhancement status of liver especially for FLLs in real-time and provides more information than color doppler US which improves classification accuracy for FLLs and increases the detection rate of

Table 1 The summary of representative liver segmentation tasks based on DL methods

Author	Model (backbone)	Other mechanism or algorithm	Main performance	Database	Modality	References
Ben-Cohen et al.	FCN-8 s (VGG-16)		Test set: Dice index: 0.89 Sensitivity: 0.86 Positive predictive value: 0.95	SLIVER07, private database	Plain CT	[31]
Zhang et al.	2D slice-based U-net, 3D patch-based FCN		ISICDM: DSC: 96.31% LiTS: DSC: 96.70%	ISICDM, LiTS	Plain CT	[32]
Fang et al.	MIMO-FAN	Iterative closest point algorithm, spatial pyramid pooling, deep pyramid supervision	LiTS: Dice score: 96.1% Speed: 0.04 s/slice	LiTS, private database	Plain CT, PET-CT, CE-CT, MRI	[33]
Tang et al.	Modified DeepMedic (CNN)		Test set: DSC: 0.94 (0.98 with manual correction)	SLIVER07, LiTS, Medical Segmentation Decathlon challenge, private database	Plain CT, CE-CT	[34]
Hu et al.	3D CNN	Globally optimized surface evolution	Sliver07: DSC: 97.2% Average symmetric surface distance: 0.84 mm	Sliver07, private database	Plain CT	[35]
Marinelli et al.	3D U-Net	Active transfer learning, NLP	MSHS: Absolute volume difference: 176 ml	LiTS, private database	Plain CT	[37]
Ghoniem	SegNet-UNet-ABC	ABC algorithm	Radiopaedia: Jaccard index: 0.96 Dice index: 0.968 Correlation coefficient: 0.962 LiTS: Jaccard index: 0.964 Dice index: 0.97 Correlation coefficient: 0.958	LiTS, Radiopaedia	Plain CT	[38]

cancers. For the extensive application of CEUS on liver cancer diagnosis, Guo et al. [45] established a CEUS-based CAD system with multi-view learning (MKL) to classify liver lesions. The input of the system consisted of arterial, portal venous and late-phase images of CEUS. Statistical features of the input images are summarized in advance in this system and the deep canonical correlation analysis (DCCA) is subsequently performed on the pairwise images. The generated results are then sent to a MKL classifier for further diagnosis. The DCCA is different from the traditional canonical correlation analysis in which a deep neural network (DNN) is applied to learn the complex nonlinear transformations of two CEUS views to seek the common information. The diagnosis result derived from the MKL classifier achieved a pretty good performance along with the low computational cost and great flexibility. Wu et al. [46] proposed a DL diagnostic model for benign and malignant FLLs classification with time intensity curves (TICs) of CEUS imaging. TICs of arterial and portal vein phases are obtained from the dynamic CEUS videos with the correction of respiratory motion. The final evaluation showed that all of the quantitative metrics, e.g., accuracy, sensitivity and specificity of the DL architecture performed better than that of other comparative methods.

In the ABC integrated method for CT images proposed by Ghoniem et al. [38] mentioned above, the lesions are extracted by U-Net, and the LeNet-5 framework with ABC algorithm (LeNet-5/ABC) was established to extract the features of lesions and accomplish the categorization. The ABC helps to seek for the optimal topology of the network. As results indicated, LeNet-5/ABC exhibited quite outstanding

performances regarding specificity, *F*1-score, accuracy and computational time. DL could also be integrated with the legacy systems on CT. Das et al. [47] developed a watershed Gaussian based DL method to isolate cancer lesions in liver CT scans. The liver is segmented using marker controlled watershed segmentation and the cancer lesions are subsequently delineated with the Gaussian mixture model. After tumor segmentation, various statistical, geometrical and texture features of the ROI are concluded and sent to DNN for identification of hemangioma, HCC, and metastatic carcinoma. This model achieved a classification accuracy of 99.38% and Jaccard index of 98.18% on training set and approximate values on the validation set.

As for the CE-CT, of which each phase plays a significant role, the features distributed in diverse phases constitute the image signature of different lesions [48]. Wang et al. [49] used the fine-tuned AlexNet and 50-layer ResNet to classify the FLLs using multi-phase CT scans including non-contrast, arterial and portal venous phases. The images of different phases were registered and formed a 3-channel image as the input for the CNN model. Researchers found that the accuracies of AlexNet and ResNet trained with limited medical images were only 78.23% and 83.67%. While combined with fine-tuning, their accuracies increased to 82.94% and 91.22%. For the 4-phase CT scanning with the additional delayed phase, the additional phase means more radiation exposure, which likely gives rise to radiation-related malignancies. Shi et al. [50] constructed a 3-phase convolutional dense network to classify HCC FLLs and non-HCC FLLs, and compared the lightweight model to the conventional 4-phase one. The result of the model without precontrast phase was close to model of 4-phase images with the accuracies of 85.6% and 83.3% but was different from the model excluding portal venous phase in discriminating HCC from other FLLs, suggesting the multi-phase CT protocol for HCC diagnosis could be optimized by removing the precontrast phase to reduce radiation dose with DL.

Another imaging modality, MRI, is widely used in combination with CT in liver lesion classification. Hamm et al. [51] developed and validated a CNN system for classification of hepatic lesions on contrast-enhanced MRI (CE-MRI). The Monte Carlo cross-validation was performed and the accuracy of DL model was compared with two qualified radiologists' inspections on another test set. Of note, the model performed overwhelming sensitivity of 90% in HCC classification over the 60% and 70% for two radiologists. The true positive and false positive rates were 93.5% and 1.6% for all lesions of the model, with a ROC-AUC of 0.992 and a computation time of 5.6 ms per lesion. And then the group attempted to interpret this black box by mapping the radiological features to diverse lesions [52]. The post hoc algorithm recorded and compared the activating patterns of the 100-neuron fully connected layer of different lesions and highlighted the feature-related parts of the input images. The possibility of a certain feature's appearance was quantified as relevance score for the input image. The model with 14 selected representative radiological features achieved 76.5% positive predictive value and 82.9% sensitivity in identifying the right radiological features presented in each test lesion. It is worth noting that the integrated model visualizes the features of raw images and is user-friendly to communicate the feature information for medical workers. 3D tomographic images of MRI could also be possessed in a framework proposed by Trivizakis et al. [53]- , to classify primary and metastatic liver carcinoma on diffusion weighted MRI (DW-MRI) as a new input form. The accuracy of the 3D model was 83% that surpassed 69.6% and 65.2% of the parallelly trained 2D models, suggesting that introducing full spatial information and constructing 3D models on DW-MRI are improved methods compared with classical models. Moreover, this 3D method also provides feasible reference for similar tasks of medical imaging.

The representative FLL-specific works are displayed in Table 2, in which the emphasis was placed on HCC identification. Multiple imaging modalities were involved in this topic, representing the general applicability of DL in liver imaging tasks.

DL methods related to diffuse liver diseases

There is an urgent need for an accurate noninvasive and highly cost-effective imaging method to replace the high-risk biopsy for fatty liver disease (FLD). The US is always applied as the primary modality to detect FLD [54]. Reddy et al. [55] proposed a transfer learning and fine-tuning based VGG-16 framework trained on US images. The testing accuracy of the model reached 90.6% for FLD, which exceeded the 84.3% accuracy for basic CNN. Specifically, for the nonalcoholic fatty liver disease (NAFLD), Byra et al. [56] proposed a DL framework to assess the disease on US. An Inception-ResNet-v2 framework was trained to evaluate the steatosis level confirmed by biopsy. The proposed approach was proven to perform better than the indicator of hepatorenal index and gray-level co-occurrence matrix algorithm with a ROC-AUC of 0.977. The work of Cao et al. [57] verified the value of DL in diagnosing NAFLD on US images. 2-dimensional liver images were assessed by several indicators, resulting in DL index presenting the best diagnostic ability in distinguishing patients of moderate and severe stages.

In the matter of liver fibrosis, Lee et al. [58] trained a DL framework to discriminate the diverse stages of liver fibrosis according to METAVIR scoring system (F0, no fibrosis; F1, portal fibrosis; F2, periportal fibrosis; F3, septal fibrosis; F4, cirrhosis.) on US images of healthy persons and chronic liver

Table 2 The FLLs detection, segmentation and classification tasks based on DL methods

Author	Model (backbone)	Other mechanism or algorithm	Main performance	Database	Modality	References
Yamakawa et al.	4-classification CNN(VGG-Net)		Accuracy: cyst: 98.1% Hemangioma: 86.8% HCC: 86.3% Metastatic liver cancer: 29.2% Malignant lesions: 94.8%	Private database	US	[40]
Schmauch et al.	ResNet50	Attention mechanism	Weighted mean ROC-AUC score for seven different tasks of FLLs diagnosis: 0.891	Data challenge organized at the 2018 Journées Francophones de Radiologie	US	[41]
Hassan et al.	stacked sparse auto-encoder	Level set method, fuzzy c-means clustering algorithm	Accuracy of FLLs diagnosis: 97.2%	Private database	US	[42]
Guo et al.	DCCA-MKL	MKL	Classification of liver tumors: Accuracy: 90.41% Sensitivity: 93.56% Specificity: 86.89%	Private database	CEUS	[45]
Wu et al.	DBN	Sparse non-negative matrix factorizations	Classification of FLLs: Accuracy: 86.36% Sensitivity: 83.33% Specificity: 87.50%	Private database	CEUS	[46]
Ghoniem	LeNet-5/ABC	ABC algorithm	Liver cancer diagnosis on Radiopaedia: Specificity: 0.986 *F*1-score: 0.98 Accuracy: 0.99 on LiTS: Specificity: 0.982 *F*1-score: 0.976 Accuracy: 0.985	LiTS, Radiopaedia	Plain CT	[38]
Das et al.	Watershed Gaussian based deep learning(DNN)	Marker controlled watershed segmentation, Gaussian mixture model algorithm	Liver cancer detection: Accuracy: 98.38% Sensitivity: 100% Specificity: 97.72%	Private database	Plain CT	[47]
Wang et al.	AlexNet, ResNet50		Classification of FLLs: Accuracy of AlexNet: 82.94% Accuracy of ResNet: 91.22%	Private database	CE-CT	[49]
Shi et al.	Automatic multiphase convolutional dense network		3-phase images without precontrast phase for HCC classification: Accuracy: 85.6% ROC-AUC: 0.920	Private database	CE-CT	[50]
Hamm et al.	CNN		HCC classification: True positive rate: 93.5% False positive rate: 1.6%	Private database	CE-MRI	[51, 52]
Trivizakis et al.	3D CNN		Liver Tumor classification: Accuracy: 83%	Private database	DW-MRI	[53]

disease (CLD) patients. The ImageNet pre-trained VGG-Net was selected as the basic classification framework. And the output layer of the model was replaced with a 4-node layer corresponding to F0-F4 stages, where F2 and F3 were combined. The established model reached an accuracy of 83.5% on the validation set and 76.4% on the test set.

The ultrasound elastography (USE) as a new form of US modality has been widely used to diagnose and evaluate the CLD [59]. By means of USE, clinicians are able to acquire the quantitative stiffness of liver tissue. The commercial variants of USE include transient elastography (TE), acoustic radiation force impulse elastography, real-time tissue elastography (RTE), shear wave elastography (SWE), etc. [60]. For the application of SWE, Gatos et al. [61] designed an algorithm to select out the low- and high-stiffness temporal stability regions of image sequences. 200 SWE image sequences of F1–F2 fibrosis constituted the dataset. After the inverse colormap-to-stiffness conversion, wavelet transform and fuzzy c-means clustering are applied to create a temporal stability mask. The stable and reliable images constrained in the mask are then applied to train a CNN for fibrosis stages confirmation. This optimization undoubtedly decreased observer bias and improved the accuracy of DL model. For assessing fibrosis stage in HBV-infected patients, Wang et al. [62] evaluated a recently developed deep learning Radiomics of elastography (DLRE) and compared the result with 2D-SWE values and serological indicators. The accuracies of the model were 0.98 for stages over F2, and 0.85 for stages over F1, which were significantly better than other methods except for the 2D-SWE values on the stages over F1. The traditional liver stiffness measurement (LSM) of TE generally relies on the liver stiffness value, which fails to reflect the comprehensive information of image patterns. It is believed that the M-mode strain image (MSI) estimated from the M-mode radiofrequency (RF) data contains much more information than traditional LSM. Li et al. [63] proposed an 8-layer DL method based on MSI to improve the prediction performance of fibrosis staging. The staging results were recorded by METAVIR scoring system from F0 to F4. The ROC-AUC of the DL method were 0.85 for stages over F2, 0.95 for stages over F3 and 0.93 for stages over F4, respectively, better than those of traditional LSM method. To conclude, the DL method with MSI performed better for advanced fibrosis and cirrhosis cases.

DL methods of this part are summarized in Table 3, of which the US was the main modality. And for the deficiency

Table 3 DL methods integrated in the tasks for diffuse liver disease

Author	Model (backbone)	Other mechanism or algorithm	Main performance	Database	Modality	References
Santhosh Reddy et al.	VGG-16		FLD diagnosis: Accuracy: 90.6%	Private database	US	[55]
Michał Byra et al.	Inception-ResNet-v2 CNN		Liver steatosis assessment: ROC-AUC: 0.977 Accuracy: 0.963	Private database	US	[56]
Cao et al	CNN		Distinguishing the moderate and severe NAFLD: ROC-AUC: 0.958	Private database	US	[57]
Lee et al.	CNN		Liver fibrosis staging: Accuracy: 76.4% AUC for cirrhosis: 0.857	Private database	US	[58]
Gatos et al.	CNN	wavelet transform, fuzzy c-means clustering	Interclass correlation coefficient: 0.92 Accuracy for various CLD stages: 82.5%—95.5%	Private database	SWE	[61]
Wang et al.	CNN		Liver fibrosis staging: AUC for F4: 0.97 AUC for ≥ F3: 0.98 AUC for ≥ F2: 0.85	Private database	2D-SWE	[62]
Li et al.	CNN		Liver fibrosis staging: AUC for ≥ F3: 0.948 AUC for F4: 0.934	Private database	TE	[63]

of existing databases, researchers have to establish their private database for model training.

Conclusions and future prospects

Past several years have witnessed the prosperity of artificial intelligence. The application of the DL algorithm brought about extraordinary transitions in imageology. For medical imaging of liver, the DL methods typified by CNN have gained considerable attention for segmentation, classification and other tough problems.

Before the advent of DL, the image processing techniques were mainly based on feature engineering and conventional image-based methods such as region growing algorithm, thresholding method, level-set method, graph cut method, etc. [64]. Among the diverse tasks of liver imaging, an accurate liver delineation is the basic step in most instances of clinical applications. Although the liver boundaries can be contoured by radiologists manually, automated segmentation algorithms are still regarded as the best choice for saving time and removing subjectivity. In virtue of DL, the downstream tasks of lesion detection and classification get rid of the complicated assembly of hand-crafted feature filters and inefficient models. Moreover, the DL methods are also adopted for model optimization, explanation, and so on (Table 4). And the conventional methods could be appropriately integrated into the DL framework in some cases to provide interfaces for the existing processes [32].

Besides extracorporeal imaging, DL technologies are also extensively applied to other examinations such as endoscopy, which generally provides accurate diagnoses based on endoscopic findings automatically. Endoscopic examinations and operations lay emphasis on the real-time capability of corresponding systems to reduce the risk of omission and assist the endoscopists to make fast decisions. For the gastrointestinal endoscopy, advanced DL algorithms have already been established to quickly classify gastric neoplasms, identify the colorectal polyps, determine the invasion depth of neoplasms and so on [69–71]. However, in the field of liver disease, the mode of DL combined with endoscope is still in its infancy, while exhibits great potential. For instance, the monitoring of endoscopic ultrasound (EUS) plays a crucial part in diagnosing, tumor ablation, medicine injection and vascular intervention for hepatopathy [72–77]. This modality unifies the US with gastrointestinal endoscope, with ultrasound probe scanning the adjacent liver inside digestive tract through gastrointestinal wall, which is much fitter for the observation of deep-located liver lesions and structures. Multidirectional information of lesion is collected by ultrasonic probe, and with rapid selection and classification of ROI, the integrated DL algorithm could be developed to identify malignant lesions, assess the degree of liver fibrosis, evaluate regional lymph nodes and detect vascular invasion

Table 4 Other auxiliary functions of DL methods for liver imaging

Author	Model (backbone)	Other mechanism or algorithm	Main performance	Database	Modality	purpose	References
Liu et al.	Dense-cycle-generative adversarial network (GAN)		Mean absolute error: 72.87 HU Peaksignal-to-noise Ratio: 22.65 dB Normalized cross-correlation: 0.92	Private dataset	MRI	CT synthesis from MRI for MR-only treatment	[65]
Ibragimov et al.	CNN	Markov random fields, centerline detection algorithm	DSC for segmentation: 0.83 Manual and DL-based prediction results were in agreement for 94% cases	Private database	CT	Automated prediction of hepatobiliary toxicity based on portal vein segmentation for liver stereotactic body radiation therapy	[66, 67]
Esses et al.	CNN		With respect to Reader 1: Sensitivity: 67% Specificity: 81% With respect to Reader 2: Sensitivity: 47% Specificity: 80%	Private database	MRI	Automatically filters out the nondiagnostic T2-weighted (T2WI) scans for detecting lesions and assessing liver morphology	[68]

for EUS automatically and accurately in real time. Furthermore, the real-time property could be fully utilized by DL algorithm for EUS-guided operations, like assisting the EUS-guided portal vein intervention to avoid misoperation, refining the tumor ablation treatment, etc. [75–77]. Meanwhile, it is worth noting that for the intraoperative imaging of laparoscopy and laparoscopic ultrasound, the DL systems are being constructed to indicate anatomical landmarks, identify tumor margins and programme excision extension in surgery [78, 79]. Annotations of DL systems would play a guiding role for surgeons, which greatly reduce the operating difficulty, effectively avoid intraoperative omissions and also have teaching effects for viewers. With the advance of DL research, it is believed that the concerned technologies are going to appear in the near future.

The increasing imaging data encourage us to integrate multiple imaging modalities for certain tasks. And for now, different types of DL are trending to collaborate on solving specific medical problems. RNN is mainly applied to NLP and sequence prediction, which easily extracts information recorded on pathographies and forecasts clinical indexes changing with time [80]. Graph neural network plays its role in graph-based networks such as molecular structures and prevails among the molecular pharmacology [81]. Besides, the prosperity of omics studies such as genomics, proteomics and metabonomics need powerful models to make full use of the raw massive data, concluding expression patterns of certain traits. The development of DNN proposes solutions to the questions precisely. The linkage of diverse forms of DL models and multi-omics realizes the information complementarity and provides multi-view learning in practical application. Thus, various liver imaging tasks could be combined with DL of other aspects to conduct studies of comprehensive diagnosis, mechanism explanation, drug discovery, etc.

However, it is worth noting that there are several restrictions in the application of DL. First of all, this data-driven training is overly dependent on the quantity of high-quality raw data, which are usually difficult to collect in a small-scale study. The construction of the subject-specific databases for DL training relies on the cooperation of medical institutions at all levels for now, together with a well-regulated and privacy-protected data sharing system. Second, most of the DL methods are well encapsulated for the end-to-end design philosophy, which means that the internals of the frameworks are invisible to users. Researchers can hardly figure out the task-related features learned by such black box models; therefore, appropriate visualization methods for the intermediate processes are urgently needed to overcome this intrinsic defect. In addition, although the DL is a powerful solution to most complex tasks, the selection of model should be considered carefully, especially for the relatively simple tasks, for which the conventional approaches may be more suitable to avoid over-fitting. In clinical practice, the model construction needs to meet clinical demand, and in terms of technological development, the developed DL models for now tend to add in cutting-edge mechanisms like attention mechanism, probabilistic programming, etc.; therefore, medical workers need to deepen partnership with computer scientists to link the theory with practice, which is a key point of interdiscipline [82, 83].

Acknowledgements This work was supported by the National Key Research and Development Program of China (No. 2018YFE0195200), National Natural Science Foundation of China (No. 81873156), The Leading Talent of Hundred, Thousand and Ten Thousand Project of Liaoning Province (XLYC1905013).

Author Contributions All authors contributed to this paper. Material preparation and article collection were performed by Kailai Xiang and Baihui Jiang. Reviewing and editing was performed by Dong Shang. The first draft of the manuscript was written by Kailai Xiang and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by the National Key Research and Development Program of China (No. 2018YFE0195200), National Natural Science Foundation of China (No. 81873156), The Leading Talent of Hundred, Thousand and Ten Thousand Project of Liaoning Province (XLYC1905013).

Availability of data and material Not applicable.

Code availability Not applicable.

Declarations

Conflicts of interest/Competing interests Kailai Xiang, Baihui Jiang, and Dong Shang have declared that they have no competing interest.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication All authors agree to publish.

References

1. Trefts E, Gannon M, Wasserman DH. The liver. Curr Biol 2017;27(21):R1147–R1151
2. Di Tommaso L, et al. Role of liver biopsy in hepatocellular carcinoma. World J Gastroenterol 2019;25(40):6041–6052
3. Ayuso C, et al. Diagnosis and staging of hepatocellular carcinoma (HCC): current guidelines. Eur J Radiol 2018;101:72–81
4. Sertorio F, et al. Ultrasonography as first line imaging for the diagnosis of positional plagiocephaly: our experience and literature review. Minerva Pediatr 2019. https://doi.org/10.23736/S0026-4946.19.05424-0
5. Nishida N, et al. Current status and perspectives for computer-aided ultrasonic diagnosis of liver lesions using deep learning technology. Hepatol Int 2019;13(4):416–421

6. Choi KJ, et al. Development and validation of a deep learning system for staging liver fibrosis by using contrast agent-enhanced CT images in the liver. Radiology 2018;289(3):688–697
7. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. Med Phys 2017;44(4):1408–1419
8. Ben-Cohen A, Greenspan H. Liver lesion detection in CT using deep learning techniques. In: Handbook of Medical Image Computing and Computer Assisted Intervention, 2020; p. 65–90. https://doi.org/10.1016/B978-0-12-816176-0.00008-9
9. Waite S, et al. Interpretive error in radiology. AJR Am J Roentgenol 2017;208(4):739–749
10. Chan HP, Samala RK, Hadjiiski LM. CAD and AI for breast cancer-recent development and challenges. Br J Radiol 2020;93(1108):20190580
11. Suzuki K. Overview of deep learning in medical imaging. Radiol Phys Technol 2017;10(3):257–273
12. Sapkota M, Shi X, Xing F, Yang L. Deep convolutional hashing for low-dimensional binary embedding of histopathological images. IEEE J Biomed Health Inf 2019;23(2):805–816
13. Le EPV, et al. Artificial intelligence in breast imaging. Clin Radiol 2019;74(5):357–366
14. Sahiner B, et al. Deep learning in medical imaging and radiation therapy. Med Phys 2019;46(1):e1–36
15. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. Z Med Phys 2019;29(2):102–127
16. Sarıgül M, Ozyildirim BM, Avci M. Differential convolutional neural network. Neural Netw 2019;116:279–287
17. Karim AM, et al. A novel framework using deep auto-encoders based linear model for data classification. Sensors (Basel) 2020;20(21):6378
18. Movahedi F, Coyle JL, Sejdic E. deep belief networks for electroencephalography: a review of recent contributions and future outlooks. IEEE J Biomed Health Inf. 2018;22(3):642–652
19. Barak O. Recurrent neural networks as versatile tools of neuroscience research. Curr Opin Neurobiol 2017;46:1–6
20. Pons E, Braun LM, Hunink MG, Kors JA. Natural language processing in radiology: a systematic review. Radiology 2016;279(2):329–343
21. Min S, Lee B, Yoon S. Deep learning in bioinformatics. Brief Bioinform 2017;18(5):851–869
22. Huan EY, Wen GH. Multilevel and multiscale feature aggregation in deep networks for facial constitution classification. Comput Math Methods Med 2019;2019:1258782
23. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging 2018;9(4):611–629
24. Biswas M, et al. Symtosis: A liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm. Comput Methods Programs Biomed 2018;155:165–177
25. Ohn I, Kim Y. Smooth function approximation by deep neural networks with general activation functions. Entropy (Basel) 2019;21(7):627
26. Soffer S, et al. Convolutional neural networks for radiologic images: a radiologist's guide. Radiology 2019;290(3):590–606
27. Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39(4):640–651
28. Falk T, et al. U-Net: deep learning for cell counting, detection, and morphometry. Nat Methods 2019;16(1):67–70
29. Clancy K, et al. Deep learning pre-training strategy for mammogram image classification: an evaluation study. J Digit Imaging 2020;33(5):1257–1265
30. Jansen MJA, et al. Patient-specific fine-tuning of convolutional neural networks for follow-up lesion quantification. J Med Imaging (Bellingham) 2020;7(6):064003
31. Ben-Cohen A, et al. Fully convolutional network for liver segmentation and lesions detection. Deep Learn Data Label Med Appl 2016. https://doi.org/10.1007/978-3-319-46976-8_9
32. Zhang Y, et al. Deep learning initialized and gradient enhanced Level-Set based segmentation for liver tumor from CT images. IEEE Access 2020;8:76056–76068
33. Fang X, Xu S, Wood BJ, Yan P. Deep learning-based liver segmentation for fusion-guided intervention. Int J Comput Assist Radiol Surg 2020;15(6):963–972
34. Tang X, et al. Whole liver segmentation based on deep learning and manual adjustment for clinical use in SIRT. Eur J Nucl Med Mol Imaging 2020;47(12):2742–2752
35. Hu P, et al. Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution. Phys Med Biol 2016;61(24):8676–8698
36. Winkel DJ, et al. Validation of a fully automated liver segmentation algorithm using multi-scale deep reinforcement learning and comparison versus manual segmentation. Eur J Radiol 2020;126:108918
37. Marinelli B, et al. Combination of active transfer learning and natural language processing to improve liver volumetry using surrogate metrics with deep learning. Radiol Artif Intell 2019. https://doi.org/10.1148/ryai.2019180019
38. Ghoniem RM. A novel bio-inspired deep learning approach for liver cancer diagnosis. Information 2020;11(2):80
39. Halder A, Dey D, Sadhu AK. lung nodule detection from feature engineering to deep learning in thoracic CT images: a comprehensive review. J Digit Imaging 2020;33(3):655–677
40. Yamakawa M, Shiina T, Nishida N, Kudo M. Computer aided diagnosis system developed for ultrasound diagnosis of liver lesions using deep learning. In: 2019 IEEE International Ultrasonics Symposium (IUS), 2019; p. 2330–2333. https://doi.org/10.1109/ULTSYM.2019.8925698
41. Schmauch B, et al. Diagnosis of focal liver lesions from ultrasound using deep learning. Diagn Interv Imaging 2019;100(4):227–233
42. Hassan TM, Elmogy M, Sallam ES. Diagnosis of focal liver diseases based on deep learning technique for ultrasound images. Arab J Sci Eng 2017;42:3127–3140
43. Durot I, Wilson SR, Willmann JK. Contrast-enhanced ultrasound of malignant liver lesions. Abdom Radiol (NY) 2018;43(4):819–847
44. Jaspers N, Pfister R, Kinkel H, Michels G. Kontrastmittelsonographie [Contrast-enhanced ultrasound]. Dtsch Med Wochenschr 2012;137(45):2336–2339
45. Guo L, et al. CEUS-based classification of liver tumors with deep canonical correlation analysis and multi-kernel learning. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017; p. 1748–1751. https://doi.org/10.1109/EMBC.2017.8037181
46. Wu K, Chen X, Ding MY. Deep learning based classification of focal liver lesions with contrast-enhanced ultrasound. Optik 2014;125(15):4057–4063
47. Das A, et al. Deep learning based liver cancer detection using watershed transform and Gaussian mixture model techniques. Cogn Syst Res 2019;54:165–175
48. Xu X, et al. Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma. J Hepatol 2019;70(6):1133–1144
49. Wang W, et al. Classification of focal liver lesions using deep learning with fine-tuning. In: Proceedings of the 2018 International Conference on digital medicine and image processing 2018; p. 56–60. https://doi.org/10.1145/3299852.3299860
50. Shi W, et al. Deep learning assisted differentiation of hepatocellular carcinoma from focal liver lesions: choice of four-phase and three-phase CT imaging protocol. Abdom Radiol (NY) 2020;45(9):2688–2697

51. Hamm CA, et al. Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI. Eur Radiol 2019;29(7):3338–3347
52. Wang C, et al. Deep learning for liver tumor diagnosis part II: convolutional neural network interpretation using radiologic imaging features. Eur Radiol 2019;29(7):3348–3357
53. Trivizakis E, et al. Extending 2-D convolutional neural networks to 3-D for advancing deep learning cancer classification with application to mri liver tumor differentiation. IEEE J Biomed Health Inf 2019;23(3):923–930
54. Xiao G, et al. Comparison of laboratory tests, ultrasound, or magnetic resonance elastography to detect fibrosis in patients with nonalcoholic fatty liver disease: a meta-analysis. Hepatology 2017;66(5):1486–1501
55. Reddy DS, Bharath R, Rajalakshmi P. A novel computer-aided diagnosis framework using deep learning for classification of fatty liver disease in ultrasound imaging. In: IEEE 20th International Conference on e-health networking, applications and services 2018; p. 1–5. https://doi.org/10.1109/HealthCom.2018.8531118
56. Byra M, et al. Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images. Int J Comput Assist Radiol Surg 2018;13(12):1895–1903
57. Cao W, et al. Application of deep learning in quantitative analysis of 2-dimensional ultrasound imaging of nonalcoholic fatty liver disease. J Ultrasound Med 2020;39(1):51–59
58. Lee JH, et al. Deep learning with ultrasonography: automated classification of liver fibrosis using a deep convolutional neural network. Eur Radiol 2020;30(2):1264–1273
59. Sigrist RMS, et al. Ultrasound elastography: review of techniques and clinical applications. Theranostics 2017;7(5):1303–1329
60. Colombo S, et al. Head-to-head comparison of transient elastography (TE), real-time tissue elastography (RTE), and acoustic radiation force impulse (ARFI) imaging in the diagnosis of liver fibrosis. J Gastroenterol 2012;47(4):461–469
61. Gatos I, et al. Temporal stability assessment in shear wave elasticity images validated by deep learning neural network for chronic liver disease fibrosis stage assessment. Med Phys 2019;46(5):2298–2309
62. Wang K, et al. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. Gut 2019;68(4):729–741
63. Li Y, et al. A deep learning trial on transient elastography for assessment of liver fibrosis. In: IEEE International Ultrasonics Symposium. 2018; p. 1–4. https://doi.org/10.1109/ULTSYM.2018.8579992
64. Yip SS, Aerts HJ. Applications and limitations of radiomics. Phys Med Biol 2016;61(13):R150–R166
65. Liu Y, et al. MRI-based treatment planning for proton radiotherapy: dosimetric validation of a deep learning-based liver synthetic CT generation method. Phys Med Bio. 2019;64(14):145015
66. Ibragimov B, et al. Combining deep learning with anatomical analysis for segmentation of the portal vein for liver SBRT planning. Phys Med Biol 2017;62(23):8943–8958
67. Ibragimov B, et al. Automated hepatobiliary toxicity prediction after liver stereotactic body radiation therapy with deep learning-based portal vein segmentation. Neurocomputing 2020;392:181–188
68. Esses SJ, et al. Automated image quality evaluation of T2 -weighted liver MRI utilizing deep learning architecture. J Magn Reson Imaging 2018;47(3):723–728
69. Cho BJ, et al. Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network. Endoscopy 2019;51(12):1121–1129
70. Le Berre C, et al. Application of artificial intelligence to gastroenterology and hepatology. Gastroenterology 2020;158(1):76-94.e2
71. Cho BJ, et al. Prediction of submucosal invasion for gastric neoplasms in endoscopic images using deep-learning. J Clin Med 2020;9(6):1858
72. Sbeit W, et al. A comprehensive narrative review on the evolving role of endoscopic ultrasound in focal solid liver lesions diagnosis and management. Diagnostics (Basel) 2020;10(9):688
73. Lisotti A, et al. EUS liver assessment using contrast agents and elastography. Endosc Ultrasound 2018;7(4):252–256
74. Zhang WY, et al. Endoscopic ultrasound-guided ethanol ablation therapy for tumors. World J Gastroenterol 2013;19(22):3397–3403
75. Chua T, et al. Endoscopic ultrasound-guided ablation of liver tumors. Gastrointest Endosc Clin N Am 2019;29(2):369–379
76. ASGE Technology Committee, Trikudanathan G, et al. EUS-guided portal vein interventions. Gastrointest Endosc 2017;85(5):883–888
77. Bhatia V, et al. Endoscopic ultrasound description of liver segmentation and anatomy. Dig Endosc 2014;26(3):482–490
78. Li YD, et al. Intelligent detection endoscopic assistant: an artificial intelligence-based system for monitoring blind spots during esophagogastroduodenoscopy in real-time. Dig Liver Dis 2021;53(2):216–223
79. Ramalhinho J, et al. A pre-operative planning framework for global registration of laparoscopic ultrasound to CT images. Int J Comput Assist Radiol Surg 2018;13(8):1177–1186
80. Short RG, Bralich J, Bogaty D, Befera NT. Comprehensive word-level classification of screening mammography reports using a neural network sequence labeling approach. J Digit Imaging 2019;32(5):685–692
81. Xiong J, Xiong Z, Chen K, Jiang H, Zheng M. Graph neural networks for automated de novo drug design. Drug Discov Today 2021;26(6):1382–1393
82. Li W, Liu K, Zhang L, Cheng F. Object detection based on an adaptive attention mechanism. Sci Rep 2020;10(1):11307
83. Fourment M, Darling AE. Evaluating probabilistic programming and fast variational Bayesian inference in phylogenetics. PeerJ 2019;7:e8272