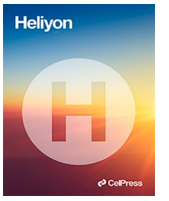
Heliyon 10 (2024) e35115

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Diagnostic performance of EfficientNetV2-S method for staging liver fibrosis based on multiparametric MRI

Haichen Zhao [a], Xiaoya Zhang [c], Yuanxiang Gao [a], Lili Wang [b], Longyang Xiao [a], Shunli Liu [a], Baoxiang Huang [c,1], Zhiming Li [a,*,1]

[a] Department of Radiology, The Affiliated Hospital of Qingdao University, Qingdao, China
[b] Department of Pathology, The Affiliated Hospital of Qingdao University, Qingdao, China
[c] College of Computer Science and Technology of Qingdao University, Qingdao, China

ARTICLE INFO

Keywords:
Deep learning
Chronic liver diseases
Convolutional neural network
EfficientNetV2-S
MRI

ABSTRACT

Problem: Previous studies had confirmed that some deep learning models had high diagnostic performance in staging liver fibrosis. However, training efficiency of models predicting liver fibrosis need to be improved to achieve rapid diagnosis and precision medicine.
Aim: The deep learning framework of EfficientNetV2-S was noted because of its faster training speed and better parameter efficiency compared with other models. Our study sought to develop noninvasive predictive models based on EfficientNetV2-S framework for staging liver fibrosis.
Methods: Patients with chronic liver disease who underwent multi-parametric abdominal MRI were included in the retrospective study. Data augmentation methods including horizontal flip, vertical flip, perspective transformation and edge enhancement were applied to multi-parametric MR images to solve the data imbalance between different liver fibrosis groups. The EfficientNetV2-S models were used for the prediction of liver fibrosis stages F1-2, F1-3, F3, F4 and F3-4. We evaluated the diagnostic performance of our models in training, validation, and test sets by using receiver operating characteristic curve (ROC) analysis.
Results: The total training time of EfficientNetV2-S was about 6 h. For differentiating of F1-2 vs F3, the accuracy, sensitivity and specificity of EfficientNetV2-S model were 96.2 %, 96.4 % and 96.0 % in the test set. The AUC in test set was 0.559. The accuracy, sensitivity and specificity were 82.1 %, 74.5 % and 89.6 % in the test set by using EfficientNetV2-S model to differentiate F1-2 vs F3-4, and the AUC in test set were 0.763. For differentiating F1-3 vs F4, the accuracy, sensitivity and specificity of EfficientNetV2-S model were 71.5 %, 73.4 % and 69.5 % in the test set. The AUC was 0.553 in test set. For differentiating F1-2 vs F4, the accuracy, sensitivity and specificity of our model were 84.3 %, 80.2 % and 88.3 % in the test set, and the AUC was 0.715, respectively. For differentiating F3 vs F4, the accuracy, sensitivity and specificity of EfficientNetV2-S model were 92.5 %, 89.1 % and 95.6 % in the test set, and the AUC was 0.696 in the test set.
Conclusions: The EfficientNetV2-S models based on multi-parametric MRI had the feasibility for staging of liver fibrosis because they showed high training speed and diagnostic performance in our study.

* Corresponding author. Department of Radiology, The Affiliated Hospital of Qingdao University, No.59 Haier Road, Qingdao, 266000, China.
E-mail address: zhiminglee@yeah.net (Z. Li).
[1] Zhiming Li and Baoxiang Huang contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e35115
Received 22 December 2023; Received in revised form 19 July 2024; Accepted 23 July 2024
Available online 27 July 2024

1. Introduction

Liver fibrosis is a common pathological process that can be caused by chronic liver injury of different etiologies [1]. Progressive liver fibrosis can result in cirrhosis, liver dysfunction and hepatocellular carcinoma [2,3]. Therefore, precise diagnosis of liver fibrosis indicates great significance for clinical practice, and patients with mild or moderate liver fibrosis can obtain good prognosis by undergoing timely clinical care. Undoubtedly, liver biopsy is still the gold standard for assessment of liver fibrosis. But it is invasive and has some inevitable limitations including procedure-related complications, sampling error, and interobserver variances [4,5].

Thus, previous studies proposed some noninvasive methods including serum biomarkers [6], ultrasound (US) elastography [7], and magnetic resonance elastography (MRE) [8] for staging liver fibrosis. Some serum biomarkers used for staging liver fibrosis have inadequate accuracy and are difficult to reflect the dynamic process of liver fibrosis. US is usually unsuitable for obese patients. MRE needs costly specialized equipment.

Notably, deep learning algorithms have achieved excellent performance in the diagnosis of diseases such as Parkinson's disease and liver tumor [9,10]. Moreover, a number of studies using convolutional neural network (CNN) models based on MRI showed high diagnostic performance for staging liver fibrosis. For instance, the VGG16 model pretrained from ImageNet was used for staging liver fibrosis with hepatobiliary phase (HBP) images [2]. This study showed that the area under the curve (AUC) of VGG16 model ranged from 0.77 to 0.91 for staging F1-4, F2-4 and F3-4 in the test sets. Huang et al. proposed a cross-contrast neural network (CCNN) which combined the advantages of statistical analysis and convolutional neural network. The CCNN model was performed for staging liver fibrosis and showed high accuracies for binary classification [11]. Nevertheless, some deep learning models used for staging liver fibrosis may have the imbalance between training speed and accuracy [12].

Recently, Tan and his co-worker proposed a deep model framework of EfficientNetV2, and the model framework was demonstrated to have high training speed and parameter efficiency due to proper combination of MBConv and Fused-MBConv blocks and adaptively adjusting regularization along with image size during training [12]. Thus, our study sought to develop noninvasive predictive models based on EfficientNetV2-S framework for staging liver fibrosis. The main contributions of our study are summarized as follows:

- We confirmed the feasibility of using deep learning models based on EfficientNetV2-S framework for liver fibrosis staging.
- Our models had high diagnostic performance for staging liver fibrosis, which may help clinicians make proper therapeutic decisions for patients with different stages of liver fibrosis.
- Training speed of our models was accelerated compared with other models. It can contribute to rapid diagnosis and precision medicine.

The rest of the paper is structured as follows: The related works are shown in Section 2. The materials and methods are presented in Section 3. The results are shown in Section 4, and discussion is carried out in Section 5. The limitations are presented in Section 6. Finally, Section 7 is devoted to the conclusion.

2. Related works

Automatic detection of the disease has been introduced in growing numbers to
facilitate lesion detection and classification [10]. This section presents few chosen earlier works on deep learning methods supported liver diseases detection.

A study by Rahman [13] used fuzzy parameterized complex fuzzy hypersoft set (FpcFHSS) with five relevant laboratory tests for the assessment of the susceptibility of patients for liver diseases. It may provide a promising tool with which clinicians can conduct therapeutic decisions on fewer invasive procedures. In addition, deep learning method is used for image recognition tasks for detecting liver diseases. Some studies suggested that deep learning model showed high diagnostic performance for staging liver fibrosis using CT [14] and MRI [15,16]. However, there were few studies using deep model framework of EfficientNet for image classification of liver diseases.

Tan and his co-worker designed model framework of EfficientNets which had better accuracy and training speed with much fewer parameters than other networks by compound scaling of model depth, width, and image size [17]. Moreover, they proposed the deep model framework of EfficientNetV2 based on the combination of training-aware neural architecture search and scaling to achieve few model parameters and higher training efficiency [12].

A previous research showed that EfficientNetB0 had high performance in identifying MRI radiological features of liver tumors [18]. In a more recent study, EfficientNet was demonstrated to have higher sensitivity in staging liver fibrosis based on US images [19]. In conclusion, existing researches suggest that deep learning model framework of EfficientNets is used for automatic detection of the liver fibrosis. Thus, the aim of this study was to investigate the performance of the EfficientNetV2 model in staging liver fibrosis by using multiparametric MRI images.

3. Materials and methods

This retrospective study was approved by our institutional review board, which waived the requirement for written informed consent.

3.1. Study population

One radiologist (Xiao LY) retrieved our institutional Picture Archiving and Communication System (PACS). Patients who met the criteria were included in this study. The inclusion criteria were as follows: (1) accepting multiparametric MRI with conventional MR sequences and gadoxetic acid contrast-enhanced MRI; (2) liver fibrosis stage confirmed by histopathologic examination; (3) high quality MR images without obvious artifacts; (4) without any treatment before operation or liver biopsy. Finally, a total of 230 patients were selected in this study. All procedures performed in studies involving human participants were in accordance with the ethical standards of the Ethics Committee of The Affiliated Hospital of Qingdao University (Approved No. of ethic committee: QYFY WZLL 28305).

3.2. Histopathologic examination

Of 230 patients, there were 226 patients accepting liver resection, three patients accepting liver explantations and one patient accepting liver biopsy. The interval between MR examination and histopathologic examination did not exceed 4 weeks. Histopathologic sections of 230 patients were reviewed by a pathologist (Wang LL, with 8 years' experience in liver pathology). Based on the Metavir fibrosis staging system [11,20], the degree of liver fibrosis is divided into five stage: F0, no fibrosis; F1, fibrous portal expansion; F2, few bridges or septa; F3, numerous bridges or septa; and F4, cirrhosis. Because of small number of patients in fibrosis stage F1 or F2 in our study, we integrated patients in stage F1 and stage F2 as a new group of stage F1-2. Finally, the numbers of patients with liver fibrosis F1-2, F3 and F4 were respectively 56, 32 and 142.

3.3. Multiparametric MRI

Multiparametric MRI was performed by using two brands of 3.0-T MR systems (Discovery MR750, GE Healthcare; Magnetom Skyra, Siemens Healthcare). MR sequences included in-phase and out-of-phase T1-weighted imaging (T1WI-IP & T1WI-OP), fat-saturated T2-weighted imaging (T2WI-FS), diffusion-weighted imaging (DWI) and gadoxetic acid-enhanced MR imaging.

The parameters of GE MR system were as follows: T1WI-IP & OP: repetition time (TR)/echo time (TE) 3.7/2.2&1.1 msec, field of view (FOV) 400 × 336 mm, bandwidth 166.67 kHz, slice thickness 5 mm, space of between slices 0 mm, number of excitation (NEX) 1; T2WI-FS: TR/TE 10000/91.4 msec, FOV 400 × 400 mm, bandwidth 50 kHz, slice thickness 5 mm, space of between slices 2 mm, NEX 2; DWI: TR/TE 6667/50.7 msec, FOV 400 × 400 mm, bandwidth 250 kHz, slice thickness 6 mm, space of between slices 2 mm, NEX 6; ADC map was generated by GE AW workstation; gadoxetic acid-enhanced MR imaging was performed using liver acquisition with volume acceleration sequence: TR/TE 3.8/2.2 msec, FOV 400 × 336 mm, bandwidth 166.67 kHz, slice thickness 5 mm, space of

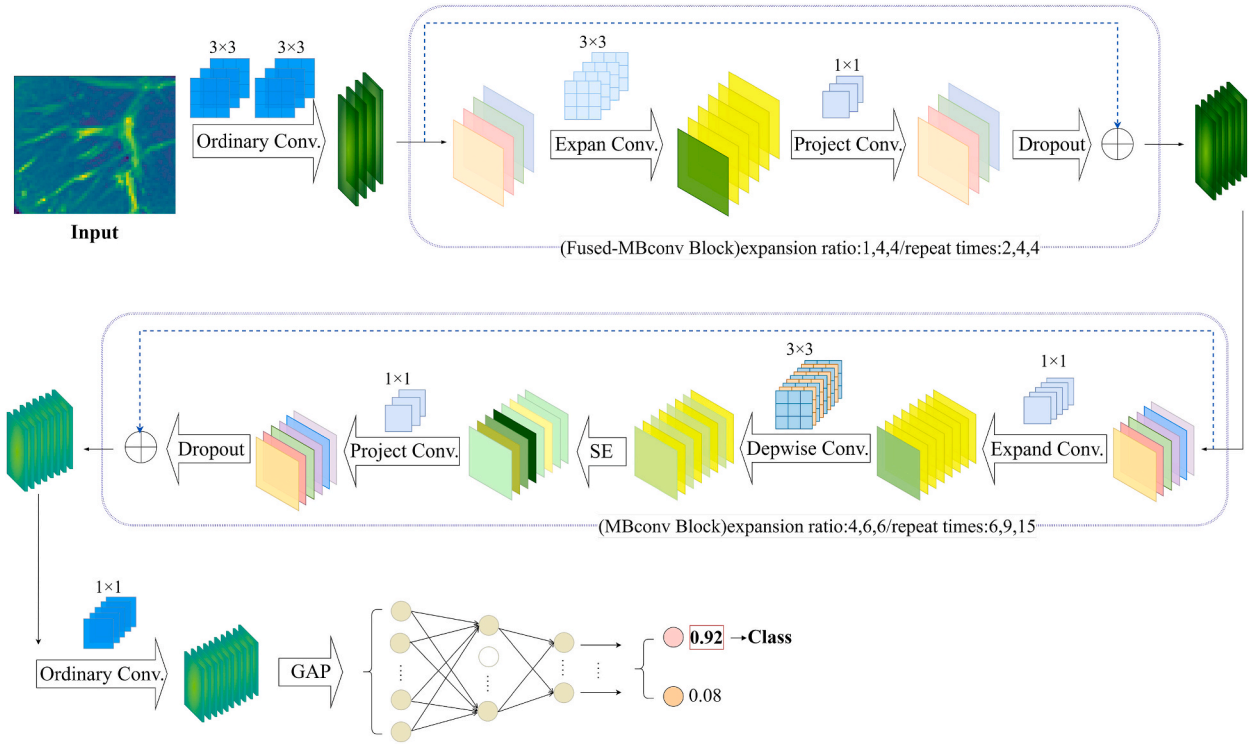


Fig. 1. The network architecture of EfficientNetV2-S used in the study.

between slices 0 mm, NEX 1. Five scanning phases of gadoxetic acid-enhanced MR included early-arterial phase (17 s after contrast injection), late-arterial phase (35 s), portal venous phase (60–70 s), transitional phase (3 min), and hepatobiliary phase (20 min). Gadoxetic acid-enhanced MR imaging was with gadoxetic acid contrast material (10 mL, Primovist, Bayer) at a dose of 0.025 mmol per kilogram of body weight.

The parameters of Siemens MR system were as follows: T1WI-IP & OP: TR/TE 6.24/2.49&1.26 msec, FOV 400 × 300 mm, bandwidth 1.04 kHz, slice thickness 3.5 mm, space of between slices 0 mm, number of excitation (NEX) 1; T2WI-FS: TR/TE 3000/79 msec, FOV 380 × 380 mm, bandwidth 0.71 kHz, slice thickness 6 mm, space of between slices 1.2 mm, NEX 4; DWI: TR/TE 5600/58 msec, FOV 400 × 300 mm, bandwidth 2.442 kHz, slice thickness 6 mm, space of between slices 1.2 mm, NEX 6; ADC map was generated by Siemens Syngo.via workstation; gadoxetic acid-enhanced MR imaging was performed using volumetrically interpolated breath-hold sequence: TR/TE 3.81/1.36 msec, FOV 380 × 380 mm, bandwidth 0.45 kHz, slice thickness 3 mm, space of between slices 0 mm, NEX 1. The scanning phases were as described above.

3.4. EfficientNetV2-S architecture

In this study, we applied the EfficientNetV2-S architecture which was described in detail by Tan et al. [12]. The architecture has 8 stages (Stage 0–7), and extensively uses MBConv and fused-MBConv. In Stage 0, it has a convolutional layer with a kernel size of 3 × 3. In Stage 1–3, these blocks apply fused-MBConv, and the number of layers ranges from 2 to 4. In Stage 1, the convolution operator is with an expansion ratio of 1 and a kernel size of 3 × 3. In Stage 2 or 3, it is with an expansion ratio of 4 and a kernel size of 3 × 3. In stage 4–6, these blocks use MBConv, and the number of layers ranges from 6 to 15. In Stage 4, the operator is with an expansion ratio of 4, a kernel size of 3 × 3 and a block of SE0.25. In Stage 5 or 6, it is with an expansion ratio of 6, a kernel size of 3 × 3 and a block of SE0.25. In Stage 7, the convolutional operators include a convolution layer with a kernel size of 1 × 1, a pooling layer and a full connection layer. The framework of EfficientNetV2-S was shown in Fig. 1 and Table 1, and more details were shown in previous study [12].

3.5. Data pre-processing

Multiparametric MR images including T1WI-IP, T1WI-OP, T2WI-FS, DWI, ADC map, dynamic contrast-enhanced images (early-arterial phase, late-arterial phase, portal venous phase, transitional phase) and HBP images, were saved from our institutional PACS system, and then were imported into a free open-source software 3D Slicer (v.5.0.3) (https://www.slicer.org/). By using 3D Slicer, one radiologist with 10-year experience for abdominal MR imaging drew the bounding boxes on multiparametric MR images. Bounding boxes were placed in the right hepatic lobe by excluding major vessels and liver masses.

By using Python 3.8, original MR images of DICOM format were transferred into RGB images. According to bounding boxes, regions of interest were extracted from the RGB images. To resolve the data imbalance between different groups, data augmentation was performed with horizontal flip, vertical flip, perspective transformation and edge enhancement. After data augmentation, the number of images in stage F1-2, F1-3, F3, F4 and F3-4 increased from 534, 834, 466, 1356 and 1656 to 1599, 1668, 1497, 1656 and 1656, respectively. Each binary classification of F1-2 vs. F3, F1-2 vs. F3-4, F1-3 vs. F4, F1-2 vs. F4, and F3 vs. F4, was regarded as a separate group. And then they were randomly divided into training, validation and test sets with a ratio of 6:1:3 by Pytorch. Specifically, 70 % of images in each group of binary classification were randomly used for training and validating the deep learning model by a ratio of 6:1, and the remaining 30 % of images were divided into test set. The workflow was shown in Fig. 2.

3.6. Deep learning analysis

The deep learning model of the EfficientNetV2-S was performed by Pytorch on a computer with a GeForce RTX 2080 Ti GPU with 11 GB memory. For each binary classification, 70 % of images were used for training and validating the deep learning model, and the remaining 30 % were used for testing the accuracy of the model. The network applied the progressive learning which could adaptively adjust regularization along with image size and thus reduce the training time. The maximum input image size was set as 300 × 300 pixels. During the training process, transfer learning loading pretrained weight on ImageNet was performed to accelerate convergence.

Table 1
The architecture of EfficientNetV2-S framework.

Stage	Operator	Stride	Channels	Layers
0	Conv3x3	2	24	1
1	Fused-MBConv1, k3x3	1	24	2
2	Fused-MBConv4, k3x3	2	48	4
3	Fused-MBConv4, k3x3	2	64	4
4	MBConv4, k3x3, SE0.25	2	128	6
5	MBConv6, k3x3, SE0.25	1	160	9
6	MBConv6, k3x3, SE0.25	2	256	15
7	Conv1x1 & Pooling & FC	–	1280	1

Abbreviations: SE0.25, Squeeze-and-Excitation module including a global average pooling and two fully connected layers, 0.25 means that the number of nodes in the first fully connected layer in the SE module is 1/4 of the channels which input MBConv.

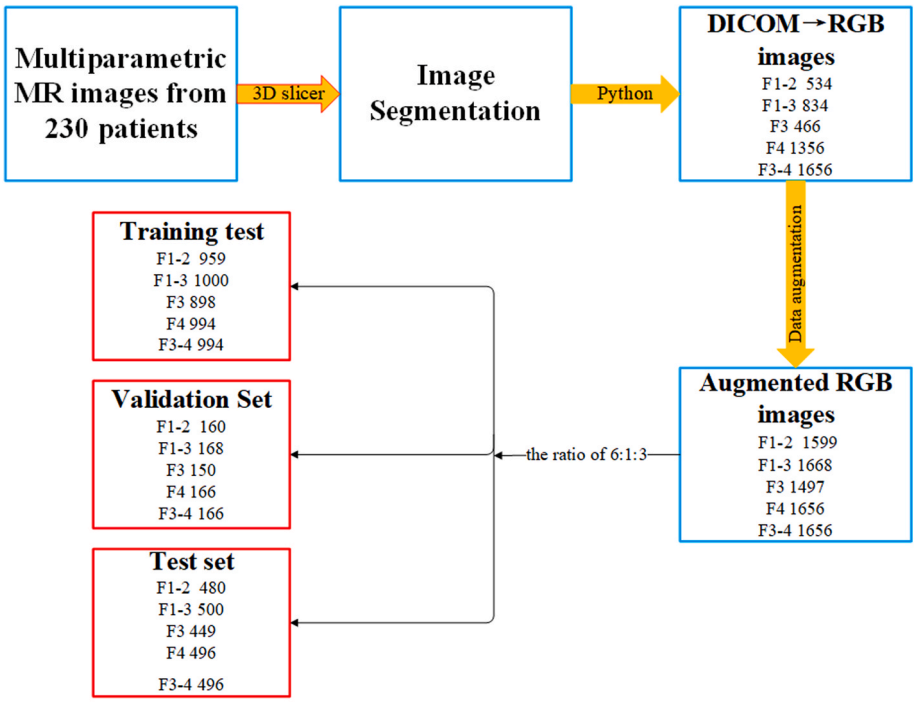


Fig. 2. An illustration of the workflow of data pre-processing and images stratification.

Parameters of all layers in the network were trained with batch size of 8. The parameters were updated by using Stochastic gradient descent (SGD) optimizer. The initial learning rate of Cross-entropy loss function was 0.01, and the learning rate of Cross-entropy loss function decayed to the minimum by using the learning rate annealing algorithm of cosine decay.

3.7. Feature visualization

Gradient-weighted class activation mapping (Grad-CAM) was deployed to generate the Grad-CAM maps with the last convolution layer of the last MBConv to interpret the features contributing to our models predictions [21]. All Grad-CAM maps of the test cohorts were systematically reviewed by a radiologist with 5-year experience for abdominal MR imaging.

3.8. Statistical analysis

All statistical analyses were performed by Python (Version 3.8). The accuracy, sensitivity and specificity were calculated for EfficientNetV2-S model in the training, validation and test sets. The accuracy, sensitivity and specificity were calculated following the equations:

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \times 100\% \quad (1)$$

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \times 100\% \quad (2)$$

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}} \times 100\% \quad (3)$$

Receiver operating characteristic (ROC) analysis and area under the ROC (AUC) were performed for assessing the diagnostic performance of the model. A p value less than 0.05 was considered significant. The AUC was calculated following the equation:

$$\text{AUC} = \frac{\sum_{i}^{n} (pos\_score > neg\_score) + 0.5 * \sum_{i}^{n} (pos\_score = neg\_score)}{\text{M} * \text{N}}$$

Where M denotes the number of positive samples, and N is the number of negative samples, and n denotes the total number of samples.

4. Results

All results of EfficientNetV-S model in training, validation and test sets were shown in the Table 2.

For prediction of F1-2 vs F3, 3096 images were randomly divided into training, validation and test sets with a ratio of 6:1:3. The accuracy, sensitivity and specificity of EfficientNetV2-S model were 98.6 %, 99.1 % and 98.1 % in the training set, 96.7 %, 95.2 % and 98.1 % in the validation set, and 96.2 %, 96.4 % and 96.0 % in the test set. The AUCs were 0.571, 0.500, and 0.559 in the training, validation and test sets. Furthermore, Fig. 3(a–d) showed representative RGB images and the corresponding Grad-CAM heat maps for the prediction of liver fibrosis stages F1-2 vs. F3 in test set.

For prediction of F1-2 vs F3-4, 3255 images were randomly divided into training, validation and tests sets with a ratio of 6:1:3. The accuracy, sensitivity and specificity of EfficientNetV2-S model were 97.7 %, 96.6 % and 98.8 % in the training set, 85.2 %, 80.8 % and 89.5 % in the validation set, and 82.1 %, 74.5 % and 89.6 % in the test set. The AUCs of EfficientNetV2-S model were 0.723, 0.776, and 0.763 in the training, validation and test sets. Fig. 4(a–d) showed representative RGB images and the corresponding Grad-CAM heat maps for the prediction of liver fibrosis stages F1-2 vs. F3-4 in test set.

For prediction of F1-3 vs F4, 3344 images were randomly divided into training, validation and tests sets with a ratio of 6:1:3. The accuracy, sensitivity and specificity of EfficientNetV2-S model were 97.7 %, 97.5 % and 98.3 % in the training set, 69.8 %, 74.8 % and 64.8 % in the validation set, and 71.5 %, 73.4 % and 69.5 % in the test set. The AUCs of EfficientNetV2-S model were 0.495, 0.535, and 0.553 in the training, validation and test sets. Fig. 5(a–d) showed representative RGB images and the corresponding Grad-CAM heat maps for the prediction of liver fibrosis stages F1-3 vs. F4 in test set.

For prediction of F1-2 vs F4, 3255 images were randomly divided into training, validation and tests sets with a ratio of 6:1:3. The accuracy, sensitivity and specificity of EfficientNetV2-S model were 97.6 %, 96.8 % and 98.3 % in the training set, 85.5 %, 81.4 % and 89.5 % in the validation set, and 84.3 %, 80.2 % and 88.3 % in the test set. The AUCs of EfficientNetV2-S model were 0.662, 0.790, and 0.715 in the training, validation and test sets. Fig. 6(a–d) showed representative RGB images and the corresponding Grad-CAM heat maps for the prediction of liver fibrosis stages F1-2 vs. F4 in test set.

For prediction of F3 vs F4, 3153 images were randomly divided into training, validation and tests sets with a ratio of 6:1:3. The accuracy, sensitivity and specificity of EfficientNetV2-S model were 97.7 %, 97.0 % and 98.4 % in the training set, 92.5 %, 91.8 % and 93.2 % in the validation set, and 92.5 %, 89.1 % and 95.6 % in the test set. The AUCs of EfficientNetV2-S model were 0.523, 0.685, and 0.696 in the training, validation and test sets. Fig. 7(a–d) showed representative RGB images and the corresponding Grad-CAM heat maps for the prediction of liver fibrosis stages F3 vs. F4 in test set.

For the test set, the AUC for differentiating F1-2 vs F3-4 was higher than those for differentiating F1-2 vs F3, F1-3 vs F4, F1-2 vs F4, and F3 vs F4, which was described in Fig. 8.

5. Discussion

In our study, EfficientNetV2-S model was used for differentiating liver fibrosis stage, and some satisfying results were obtained. The result showed high accuracies of EfficientNetV2-S model ranged from 71.5 % to 96.2 % in the test set.

Compared with other deep learning model, the major advantages of EfficientNetV2-S model are of faster training speed and better parameter efficiency. The total training time of EfficientNetV2-S was about 6 h with high accuracies of 97.6 %–98.6 %.

Table 2
Diagnostic performance of EfficientNetV-S model in training, validation and test sets.

Training set					
Parameter	F1-2 vs F3	F1-2 vs F3-4	F1-3 vs F4	F1-2 vs F4	F3 vs F4
Accuracy	98.6(97.9–99.1)	97.7(96.9–98.3)	97,7(96.9–98.3)	97.6(96.7–98.2)	97.7(96.9–98.3)
Sensitivity	99.1(98.2–99.6)	96.6(95.2–97.6)	97.5(96.3–98.4)	96.8(95.4–97.8)	97.0(95.6–98.0)
Specificity	98.1(97.0–98.9)	98.8(97.8–99.3)	98.3(97.2–99.0)	98.3(97.2–99.0)	98.4(97.4–99.0)
AUC	0.571(0.570–0.572)	0.723(0.722–0.724)	0.495(0.494–0.496)	0.662(0.661–0.663)	0.523(0.522–0.524)
Validation set					
Parameter	F1-2 vs F3	F1-2 vs F3-4	F1-3 vs F4	F1-2 vs F4	F3 vs F4
Accuracy	96.7(93.8–98.3)	85.2(80.7–88.8)	69.8(64.5–74.7)	85.5(81.1–89.1)	92.5(88.9–95.1)
Sensitivity	95.2(90.0–97.9)	80.8(73.5–86.5)	74.8(67.3–81.2)	81.4(74.2–87.0)	91.8(85.8–95.5)
Specificity	98.1(94.0–99.5)	89.5(83.5–93.6)	64.8(56.9–72.0)	89.5(83.5–93.6)	93.2(87.9–96.4)
AUC	0.500(0.493–0.506)	0.776(0.771–0.782)	0.535(0.529–0.541)	0.790(0.785–0.795)	0.685(0.679–0.691)
Test set					
Parameter	F1-2 vs F3	F1-2 vs F3-4	F1-3 vs F4	F1-2 vs F4	F3 vs F4
Accuracy	96.2(94.7–97.3)	82.1(79.5–84.5)	71.5(68.5–74.2)	84.3(81.8–86.5)	92.5(90.6–94.0)
Sensitivity	96.4(94.2–97.9)	74.5(70.3–78.3)	73.4(69.3–77.2)	80.2(76.3–83.6)	89.1(85.7–91.7)
Specificity	96.0(93.8–97.5)	89.6(86.4–92.1)	69.5(65.2–73.5)	88.3(85.1–91.0)	95.6(93.2–97.1)
AUC	0.559(0.557–0.562)	0.763(0.761–0.764)	0.553(0.551–0.555)	0.715(0.713–0.717)	0.696(0.694–0.698)

AUC: Areas Under the ROC Curve.

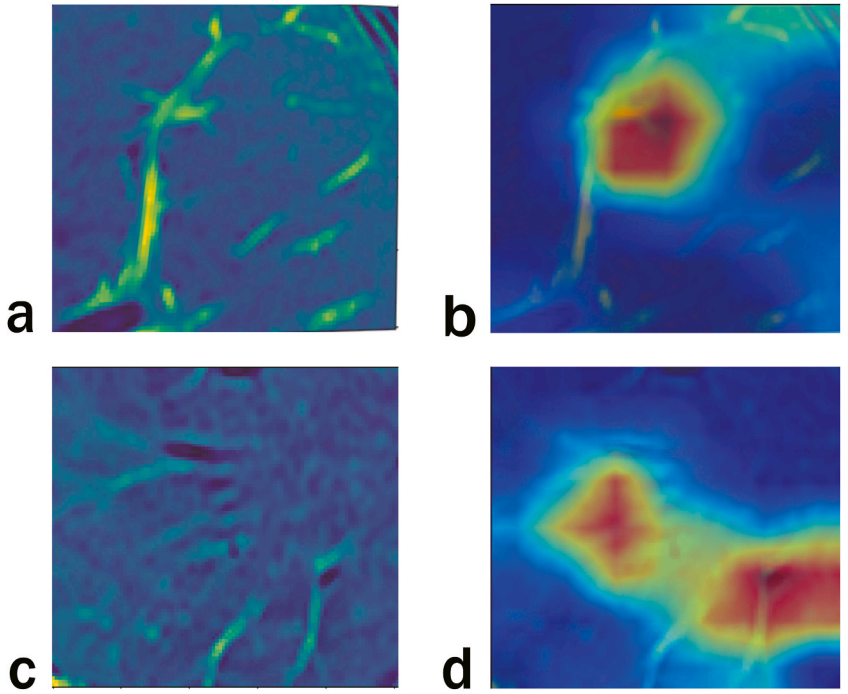


Fig. 3. Representative RGB images and the corresponding Grad-CAM heat maps for differentiating F1-2 vs F3 in test set. Fig. 3a and b showed RGB image and corresponding Grad-CAM heat map in F1-2 group. Fig. 3c and 3d showed RGB image and corresponding Grad-CAM heat map in F3 group.

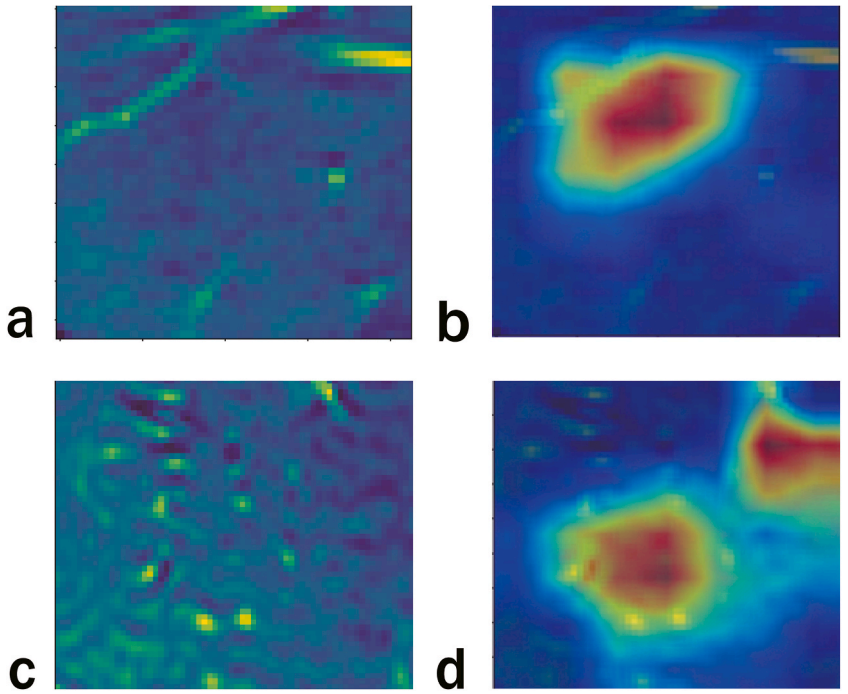


Fig. 4. Representative RGB images and the corresponding Grad-CAM heat maps for differentiating F1-2 vs F3-4 in test set. Fig. 4a and b showed RGB image and corresponding Grad-CAM heat map in F1-2 group. Fig. 4c and d showed RGB image and corresponding Grad-CAM heat map in F3-4 group.

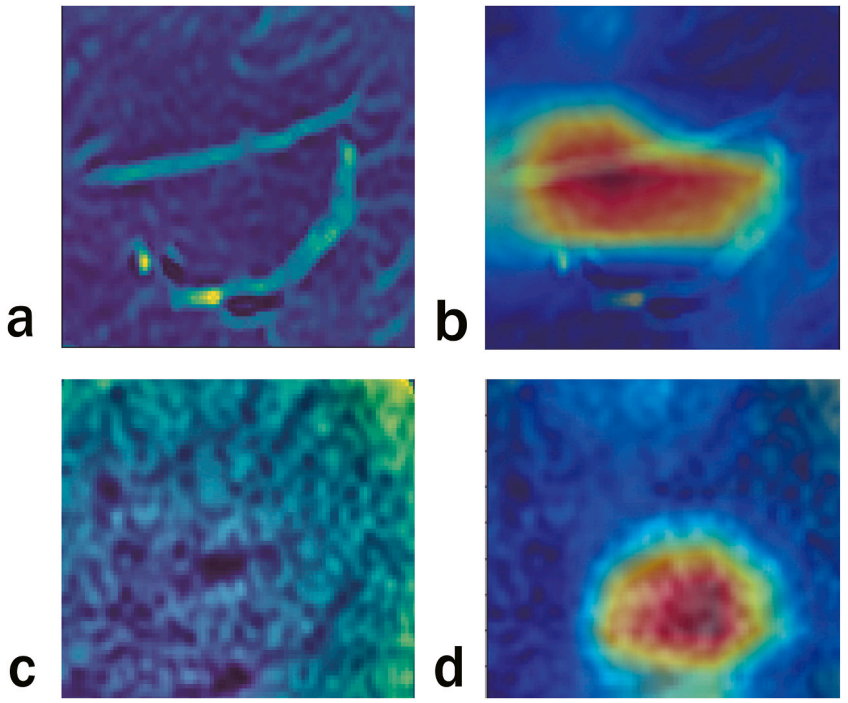


Fig. 5. Representative RGB images and the corresponding Grad-CAM heat maps for differentiating F1-3 vs F4 in test set. Fig. 5a and b showed RGB image and corresponding Grad-CAM heat map in F1-3 group. Fig. 5c and d showed RGB image and corresponding Grad-CAM heat map in F4 group.

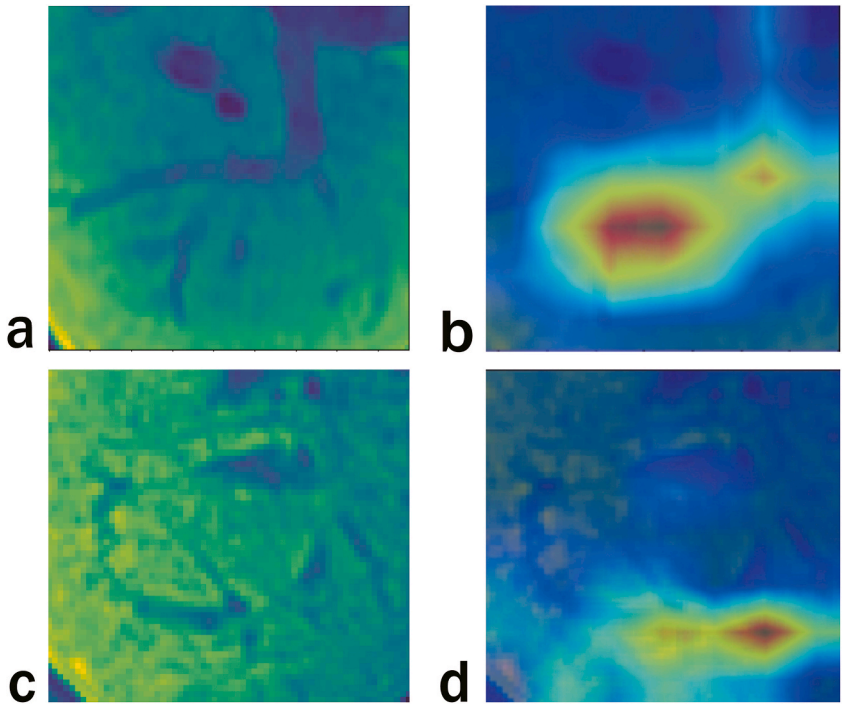


Fig. 6. Representative RGB images and the corresponding Grad-CAM heat maps for differentiating F1-2 vs F4 in test set. Fig. 6a and b showed RGB image and corresponding Grad-CAM heat map in F1-2 group. Fig. 6c and d showed RGB image and corresponding Grad-CAM heat map in F4 group.

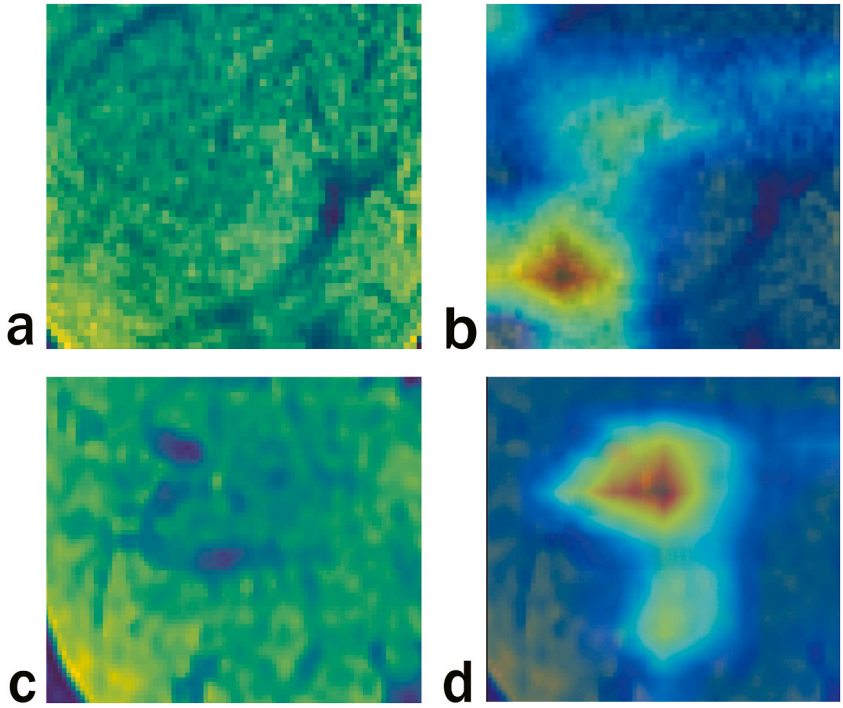


Fig. 7. Representative RGB images and the corresponding Grad-CAM heat maps for differentiating F3 vs F4 in test set. Fig. 7a and b showed RGB image and corresponding Grad-CAM heat map in F3 group. Fig. 7c and d showed RGB image and corresponding Grad-CAM heat map in F4 group.

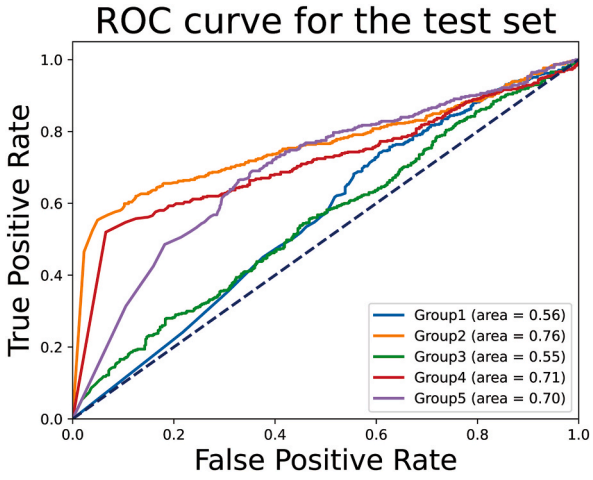


Fig. 8. ROC curve for the test set. Group1 represents the prediction of F1-2 vs F3. Group2 represents the prediction of F1-2 vs F3–F4. Group3 represents the prediction of F1-3 vs F4. Group4 represents the prediction of F1-2 vs F4. Group5 represents the prediction of F3 vs F4. The area under the ROC for differentiating F1-2 vs F3-4 was higher than those for differentiating F1-2 vs F3, F1-3 vs F4, F1-2 vs F4, and F3 vs F4.

In our study, deep learning showed a great help in staging liver fibrosis with high accuracy. We used the same data to performed a radiomics analysis for staging liver fibrosis [22]. The accuracies of radiomics analysis ranged from 68.8 % to 77.5 % in training cohort, and from 26.4 % to 70.4 % in testing cohort. The accuracy of radiomics analysis in differentiating F3 vs F4 were lowest, and we speculated that stage F3 could have incomplete cirrhosis nodule which mimic the nodules of cirrhosis. In this study, the accuracies for staging liver fibrosis were from 97.6 % to 98.6 % in the training set, and from 71.5 % to 96.2 % in the test set. Therefore, the EfficientNetV2-S model outperforms radiomics model by using the same dataset. We believed that discriminating F3 vs F4 has an important role in clinical practice because compared with liver fibrosis stage F3, individuals with liver fibrosis stage F4 tend to have a higher risk for development of liver disease complications and worse survival [23].

Moreover, our results showed higher AUCs in differentiating between F1-2 and F3-4 in the training, validation and test sets. The EfficientNetV2-S model may be a promising method for diagnosing advanced liver fibrosis with high accuracy of 82.1 % in test set. Huang et al. used a cross-contrast neural network for staging liver fibrosis, and the accuracy for diagnosing advanced fibrosis was 71.11 % in test set [2]. The EfficientNetV2-S model could be superior to the deep model based on cross-contrast neural network. However, when compared with another deep learning study, deep transfer learning model showed higher accuracy than EfficientNetV2-S model in diagnosing liver cirrhosis [3]. We inferred that only one MR parameter could increase the robustness of deep learning model. But, compared with application of multiparameter MR, using only one MR parameter may inevitably loss too much image information.

The Grad-CAM produced location maps that could indicate the focus of the liver fibrosis staging network. Thus, it was utilized to provide a visual-based explanation for features contributing to our models predictions. We thought that the method might be an auxiliary tool to evaluate the prognosis of liver fibrosis after treatment and improve the accuracy of liver biopsy.

6. Limitations

There were some limitations in our study. First, although the accuracy, sensitivity and specificity of our models were high in training, validation and test sets, our results showed relatively low AUCs for staging liver fibrosis. Some reasons attributed that (1) intermediate stages of liver fibrosis could exit subtle and heterogeneous liver changes because liver fibrosis is chronic progressive process; (2) there was a challenge for a pathologist to differentiate stage F3 in clinical practice; (3) prediction probability of our models for positive samples might be low; (4) multi-parametric MRI data was used for staging liver fibrosis, which might increase data complexity and reduce model generalization in our study. In addition, there are lack of external validation datasets in our study due to the difficulty of data sampling and lack of public dataset for liver fibrosis. In the future work, we will focus on these limitations.

7. Conclusion

In conclusion, the deep learning framework of EfficientNetV2-S is valuable for staging liver fibrosis with multi-parametric MRI.

Our models had higher accuracy, sensitivity and specificity for staging liver fibrosis compared with other models reported by other studies. Additionally, the EfficientNetV2-S model has higher diagnostic performance for differentiating between F1-2 and F3-4, which can impact clinical decision-making because some researchers suggested that nanoparticulates treatment can reverse the progression of liver fibrosis for patients with mild or moderate liver fibrosis [24].

Nevertheless, several limitations have been identified in our study. First, the AUCs were lower than other deep learning models of previous studies. In addition, there are lack of external validation datasets in our study due to the difficulty of data sampling and lack of public dataset for liver fibrosis. Moreover, hepatitis B is main causes of liver fibrosis in our area, which may lead to different performance of the EfficientNetV2-S model in staging liver fibrosis for patients with other etiologies (e.g., hepatitis C, autoimmune liver disease and nonalcoholic fatty liver disease).

Thus, exploring the diagnostic performance of our models in a multicenter study with subgroups of different etiologies should be investigated in subsequent research. In addition, multi-parametric MRI data was used for staging liver fibrosis, which might increase data complexity and reduce model generalization in our study. Thus, we will utilize few MRI sequences to construct deep learning models and compare their diagnostic performance in future study.

Data availability statement

The dataset and code of this study are available from Supplementary Material.

CRediT authorship contribution statement

Haichen Zhao: Writing – original draft, Data curation, Conceptualization. **Xiaoya Zhang:** Software, Methodology. **Yuanxiang Gao:** Writing – original draft, Project administration, Methodology, Investigation. **Lili Wang:** Software, Resources, Methodology. **Longyang Xiao:** Writing – original draft, Methodology, Investigation, Formal analysis. **Shunli Liu:** Supervision, Software, Resources. **Baoxiang Huang:** Validation, Supervision, Software, Data curation. **Zhiming Li:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Zhiming LI reports administrative support, article publishing charges, and equipment, drugs, or supplies were provided by Diagnostic performance of EfficientNetV2-S method for staging liver fibrosis based on multiparametric MRI. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was funded by Clinical Medicine-X Project of The Affiliated Hospital of Qingdao University (QDFY+X202101017).

Abbreviation

ROC Receiver Operating Characteristic Curve
CNN convolutional neural network
ECM extracellular matrix
DCCN deep convolutional neural network
CCNN cross-contrast neural network
HBP hepatobiliary phase
PACS Picture Archiving and Communication System
T1WI-IP & T1WI-OP in-phase and out-of-phase T1-weighted imaging
T2WI-FS fat-saturated T2-weighted imaging
DWI diffusion-weighted imaging
TR repetition time
TE echo time
FOV field of view
NEX number of excitation
US ultrasound
CT computed tomography
AUC area under the ROC
CLD chronic liver diseases
SGD Stochastic gradient descent

References

[1] A. Caligiuri, et al., Cellular and molecular mechanisms underlying liver fibrosis regression, Cells 10 (10) (2021).
[2] Y. Huang, et al., A liver fibrosis staging method using cross-contrast network, Expert Syst. Appl. 130 (2019) 124–131.
[3] S. Nowak, et al., Detection of liver cirrhosis in standard T2-weighted MRI using deep transfer learning, Eur. Radiol. 31 (11) (2021) 8807–8815.
[4] Y. Yin, et al., Liver fibrosis staging by deep learning: a visual-based explanation of diagnostic decisions of the model, Eur. Radiol. 31 (12) (2021) 9620–9627.
[5] K. Wang, 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 68 (4) (2019) 729–741.
[6] N. Kaur, et al., Potential role of noninvasive biomarkers during liver fibrosis, 世界肝病学杂志:英文版(电子版) 13 (12) (2021) 1919–1935.
[7] G. Ferraioli, et al., Shear wave elastography for evaluation of liver fibrosis, Journal of Ultrasound in Medicine Official Journal of the American Institute of Ultrasound in Medicine 33 (2) (2014) 197–203.
[8] A. Ozturk, et al., Liver fibrosis assessment: MR and US elastography, Abdominal Radiology (2022) 1–14.
[9] C.A. Hamm, et al., Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI, Eur. Radiol. 29 (2019) 3338–3347.
[10] V. Rajinikanth, S. Yassine, S.A. Bukhari, Hand-sketchs based Parkinson's disease screening using lightweight deep-learning with two-fold training and fused optimal features, International Journal of Mathematics, Statistics, and Computer Science 2 (2024) 9–18.
[11] Z.D. Goodman, Grading and staging systems for inflammation and fibrosis in chronic liver diseases, J. Hepatol. 47 (4) (2007) 598–607.
[12] M.L.Q.V. Tan, EfficientNetV2: Smaller Models and Faster Training, 2021 arXiv. 2021. arXiv.
[13] A.U. Rahman, et al., An innovative mathematical approach to the evaluation of susceptibility in liver disorder based on fuzzy parameterized complex fuzzy hypersoft set, Biomed. Signal Process Control 86 (2023) 105204.
[14] K. Yasaka, H. Akai, A. Kunimatsu, O. Abe, S. Kiryu, Deep learning for staging liver fibrosis on CT: a pilot study, European radiology 28 (2018) 4578–4585.
[15] S.J. Hectors, et al., Fully automated prediction of liver fibrosis using deep learning analysis of gadoxetic acid-enhanced MRI, Eur. Radiol. 31 (6) (2021) 3805–3814.
[16] K. Yasaka, et al., Liver fibrosis: deep convolutional neural network for staging by using gadoxetic acid-enhanced hepatobiliary phase MR images, Radiology 287 (1) (2018) 146–155.
[17] M. Tan, Q. Le Efficientnet, Rethinking model scaling for convolutional neural networks, in: International Conference on Machine Learning, PMLR, 2019.
[18] Róbert Stollmayer, et al., Focal liver lesion MRI feature identification using efficientnet and : a feasibility study, Cells 11 (9) (2022).
[19] H.-C. Park, et al., Automated classification of liver fibrosis stages using ultrasound imaging, BMC Med. Imag. 24 (1) (2024) 36.
[20] Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. The French METAVIR Cooperative Study Group, Hepatology 20 (1) (1994) 15–20.
[21] R.R. Selvaraju, et al., Grad-cam: visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017.
[22] L. Xiao, H. Zhao, S Liu, et al., Staging liver fibrosis: comparison of radiomics model and fusion model based on multiparametric MRI in patients with chronic liver disease, Abdominal Radiology 49 (4) (2024) 1165–1174.
[23] P. Axley, et al., Patients with stage 3 compared to stage 4 liver fibrosis have lower frequency of and longer time to liver disease complications, PLoS One 13 (5) (2018) e0197117.
[24] S.S. Poilil, R. George Thomas, M.J. Moon, Y.Y. Jeong, Nanoparticles for the treatment of liver fibrosis, Int. J. Nanomed. (2017) 6997–7006.