2024 IEEE 11th International Conference on Cyber Security and Cloud Computing (CSCloud) | 979-8-3503-7698-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/CSCLOUD62866.2024.00023

Liver Fibrosis Classification Based on Multimodal Imaging Feature Fusion

Xinyan Jiang[a], Xinping Ren[b*], Yongxin Zhu[a*], Xiaoying Zheng[a], Yueying Zhou[a], Li Tian[a]
[a]Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai, China
[b]Ultrasound Department, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

***Abstract*—This study introduces a liver fibrosis staging method based on the fusion of multi-modal imaging features. By leveraging ultrasound gray-scale images and ultrasound shear wave elastography (SWE), the method employs an attention-based weighted strategy to effectively fuse different feature maps, integrating information from diverse modalities. Additionally, it integrates multi-modal network branches and designs a multi-modal integrated loss function to update network parameters, thereby enhancing the model's generalization and anti-interference capabilities. The experimental results demonstrate that the proposed fusion network achieves a high AUC of 95.01 and an accuracy of 76.21%. Compared to existing liver fibrosis classification methods for the five-class classification task, which integrate multi-modal features from various liver imaging modalities, our approach shows a significant improvement of 6% in accuracy. With its lightweight model architecture and low computational resource consumption, the proposed method effectively performs liver fibrosis staging, holding significant promise for clinical auxiliary diagnosis.**

***Index Terms*—Liver fibrosis diagnosis, Multi-modal fusion, Attention, Model ensemble, Transfer learning**

I. Introduction

Metabolic-associated fatty liver disease (MAFLD) has become the leading cause of chronic liver disease worldwide [1]. The grading of liver fibrosis is of paramount importance for the prognosis of MAFLD [2]. The METAVIR liver fibrosis assessment system categorizes liver fibrosis into five stages, ranging from F0 to F4, with fibrosis severity increasing progressively [3]. Traditional diagnosis of liver fibrosis involves invasive liver biopsy for histopathological examination, which often imposes significant pain and treatment costs on patients [16].

In recent years, with the development of cloud [4]–[6] and big data [7]–[9] technologies, *artificial intelligence* (AI) technology has been widely applied in the field of medical image-assisted diagnosis [10]–[12] and other big data areas [13]–[15]. Research on liver fibrosis diagnosis using different modalities of medical imaging mainly focuses on several aspects. Some studies utilize B-mode ultrasound (US) [17], while others employ CT [18] and MRI [19]. Currently, research on image-based liver fibrosis diagnosis often simplifies the grading problem [20]. For instance, Wang et al. [21] developed a four-layer convolutional neural network to classify liver fibrosis into two stages. Lee et al. [22] grouped the challenging F2 and F3 stages into the same category and conducted a four-class classification study. In five-class classification, three studies [23] utilized ultrasound (US), one study utilized *Shear Wave Elastography* (SWE) [24].

Furthermore, some studies have combined multi-modal images to investigate fibrosis diagnosis. Gao et al. [25] introduced an active learning module employing a mid-fusion strategy with multi-modal ultrasound images. Incorporating four modalities of images, they investigated the classification performance of various modalities combinations. Their study underscored the efficacy of multi-modal fusion in evaluating liver fibrosis severity, yielding an AUC of 89.27% and an accuracy of 70.59%.

However, the current research still faces significant limitations. Firstly, liver fibrosis image datasets are often small in scale, and subjective annotations by physicians can lead to low dataset quality. This hampers the performance improvement of training models. Balancing model complexity with dataset size to conserve computational resources, reduce diagnostic costs, and simultaneously improve classification accuracy remains a primary challenge in the field. Secondly, most current fusion strategies for multi-modal liver images are limited to a few traditional fusion methods [26], without considering the importance of different feature maps, thus restricting the improvement of model performance. Finally, deep learning models are susceptible to adversarial attacks, where attack samples introduce small perturbations to mislead model decisions, resulting in weak model robustness [27].

To address the aforementioned issues, this study proposes an attention-based weighted strategy to fuse different feature maps, effectively integrating complementary information from different modalities and improving fibrosis diagnosis performance. Additionally, a transfer learning strategy is employed, where the model is initialized with parameters trained on a large-scale dataset. Subsequently, the high-level parameters of the feature extraction network are fine-tuned using small-scale image datasets of *elastogram modality* (EM) and *gray scale modality* (GM). This approach effectively mitigates overfitting issues caused by training on small-scale medical datasets. Finally, both single-modal and multi-modal network branches are integrated using model ensemble. A model ensemble loss is designed to update the network parameters, enhancing the model's resistance to noisy data and increasing its robustness. We use the single-modal model as a baseline model and discuss the improvement in liver fibrosis diagnosis performance achieved by the fusion and ensemble strategies proposed in

*Corresponding authors: peaceheart80@163.com , zhuyongxin@sari.ac.cn

2693-8928/24/$31.00 ©2024 IEEE
DOI 10.1109/CSCloud62866.2024.00023

this study.

II. Dataset preparation

A. Multi-Modal Dataset

In our study, we utilized clinical medical images from 250 cases provided by Ruijin Hospital, including two modalities: Ultrasonic Gray-scale Image and Ultrasonic *Shear Wave Elasticity* (SWE), as illustrated in Fig. 1.

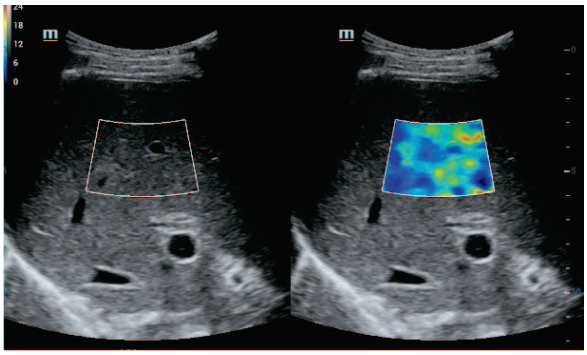

Fig. 1. Dataset illustration.

The descriptions of the images from different modalities are as follows:

- Ultrasonic Gray-scale Image: The Ultrasonic Gray-scale Image is a common type of image used in ultrasound examinations, reflecting the distribution of echo intensity within the liver tissue. By analyzing ultrasonic gray-scale images, the structure and echo characteristics of liver tissue can be observed, allowing for the assessment of the degree of liver fibrosis. Normal liver tissue appears as a uniform distribution of echoes on ultrasonic gray-scale images, while in cases of liver fibrosis, the echo distribution becomes uneven, with localized areas of strong or weak echoes.
- Ultrasonic shear wave elasticity: Ultrasonic shear wave elastography is an advanced ultrasound technique that provides quantitative information about tissue elasticity. This technique evaluates tissue hardness by measuring the propagation of shear waves in the tissue, thereby indirectly reflecting tissue elasticity. Different colors represent different values of the elastic modulus, with fibrotic areas typically showing higher elastic modulus values, indicated by darker colors in the elastography image.

B. Data Preprocessing

From a total of 250 cases, we collected 1414 pairs of effective US and SWE images from different ultrasound perspectives. Among them, there are 162 cases classified as F0, 287 as F1, 284 as F2, 197 as F3, and 94 as F4. There is a significant difference in the number of samples among different fibrosis grades. To achieve higher accuracy and model generalization, it is usually necessary to train the model with datasets of tens of thousands of samples. Therefore, we adopted a data augmentation strategy to achieve overall dataset expansion and balance the number of samples among different fibrosis stages.

Due to the presence of clinically annotated sectorial *regions of interest* (ROI) in the image dataset, we achieved automated segmentation of the sectorial through contour detection in our work. Subsequently, we applied data augmentation techniques such as horizontal flipping and small-distance translation to the ROI, as illustrated in fig.2.

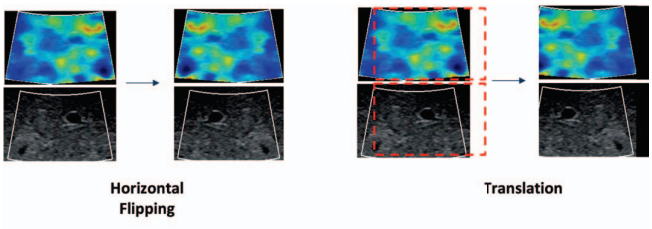

Fig. 2. Dataset augmentation.

- Horizontal Flipping: The image is flipped along the horizontal axis, enhancing the model's invariance to changes in the vertical position of the image and thus improving the model's generalization capability.
- Translation: The image is horizontally translated by a certain distance, typically ranging from a few pixels to tens of pixels. This process simulates minor variations in the position of the liver in different ultrasound angles, thereby enhancing the model's stability to minor translational changes in the image.

All data were subjected to horizontal flipping and translated 5 pixels to the left and 5 pixels to the right. Additionally, to balance the number of samples in each class, data augmentation included horizontal translation ranging from 0 to 10 pixels for classes with fewer samples. Following data augmentation, the number of samples per class was 1103 for F0, 1148 for F1, 1136 for F2, 1121 for F3, and 1087 for F4.

III. Proposed method

A. Overall Framework

It is a significant challenge to extract and correlate information from different modalities of images to improve the performance of liver fibrosis classification. Fig. 3 illustrates an overview of our proposed model. The proposed network takes US and SWE images as input. Subsequently, feature extraction networks are applied to extract features from the two modalities separately. The extracted features are then subjected to dimension reduction and feature enhancement. These processed features are fed into a fusion network to obtain attention maps, and then weighted with the original feature maps. The features from different modalities are then concatenated and fed into a classification network. Finally, the US, SWE, and fusion branches are integrated to form the model ensemble.

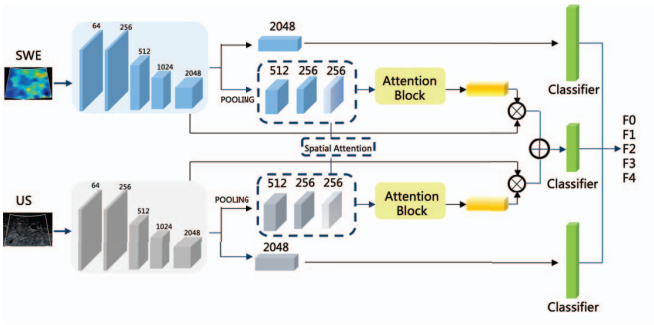

Fig. 3. Network structure.

B. Feature Extraction Networks

Vision Transformers have shown great potential in computer vision research [28]. However, compared to CNNs, they cannot leverage prior knowledge such as scale, translation invariance, and local feature locality inherent in images [29]. This implies that Vision Transformers require large-scale datasets to learn high-quality intermediate representations, posing significant challenges for their application in medical image research where high-quality data is scarce. Therefore, in our work, we still employ CNN architectures.

To better extract deep features, we adopted the ResNet-50 [30] model as the backbone for feature extraction. Considering the multi-modal input, the feature extraction networks for the two modalities are independently designed to effectively extract features from different modalities. High-level deep feature maps are extracted from stage 4 of the ResNet network and used as input for the subsequent fusion network, corresponding to the middle two branch in Fig. 3. The feature maps obtained from stage 4 are flattened and fed into a classifier consisting of three fully connected layers, forming the single-modal branches represented by the top and bottom branches in Fig. 3.

C. Multi-Modal Fusion

In the attention fusion segment, the features extracted from the network possess 2048 channels. To reduce redundancy and facilitate cross-channel information exchange, the features are passed through a pooling layer to reduce dimension to 512, and then a 1x1 convolution is employed to compress the feature map's depth to 256 channels. Subsequently, the dimension-reduced features are processed through a spatial attention module. Within this module, the channel-wise mean $\mathrm{M}_{\mathrm{avg}}$ and maximum $\mathrm{M}_{\mathrm{max}}$ are computed, and the resulting two-channel feature maps are concatenated along the channel axis. This concatenated representation is then convoluted to integrate the spatial characteristics, followed by a sigmoid activation function, generating a spatial attention map that accentuates the spatial attributes of the feature map. The spatially emphasized feature map is then directed into an attention block, as depicted in Fig. 4.

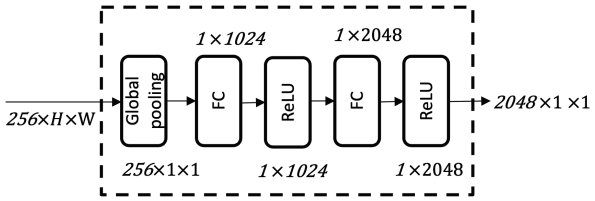

Fig. 4. Attetion block.

To capture the global receptive field, the features are compressed via global average pooling. And then the compressed features are passed through two fully connected layers to augment the number of channels, resulting in a channel-weighted map. This map is then element-wise multiplied with the feature map derived from Stage 4. Following processing through the attention block, a fully connected layer is applied to reduce the channels to 256. The features from both modalities are concatenated, yielding a fused feature vector of 512 dimensions.

D. Model Ensemble Strategy

Considering that different modalities of images focus on different aspects of fibrosis classification, we integrate the single-modal and fusion-modal branches. Features obtained from each branch are fed into fully connected (FC) layers, and cross-entropy loss functions are employed, denoted as $\mathcal{L}_{\mathrm{swe}}$, $\mathcal{L}_{\mathrm{us}}$ and $\mathcal{L}_{\mathrm{joint}}$, respectively. The final loss $\mathcal{L}_{\mathrm{w}}$ is defined as the sum of the individual losses:

$$\mathcal{L}_{\mathrm{w}} = \mathcal{L}_{\mathrm{swe}} + \mathcal{L}_{\mathrm{us}} + \mathcal{L}_{\mathrm{joint}} \tag{1}$$

Although there are significant differences between images of different modalities from the same patient, there is an intuition that they share some consistencies at the feature level. It is imperative that all modalities provide coherent recommendations for the diagnosis. Consequently, we have developed a consistency loss function to ensure harmony in the contributions from each modality. The consistency loss is defined as:

$$L_C = \sum_{i,j} \|f_i - f_j\|_1 \tag{2}$$

The features from different modalities are denoted as f_i. To impose a consistency constraint on these features, we utilize the discrepancies between feature maps of distinct modalities as a loss component, which enhances the model's robustness to noise interference and guides the learning of effective features. Drawing from the findings in [31] that MAE exhibits greater resistance to noise, we employ mean absolute error (MAE) as the loss function for L_C.

The total loss function of the proposed framework is as follows:

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_w + \alpha \mathcal{L}_c \tag{3}$$

In this study, the coefficient α is set to 0.2, serving as the weight for the consistency loss component within the total loss function.

E. Training Strategy

In the field of deep learning for medical image analysis, transfer learning has demonstrated substantial advancements [32], particularly in the task of liver fibrosis staging. The privacy concerns and scarcity of medical data present significant challenges for conventional training approaches that start from scratch. Transfer learning leverages deep learning models pre-trained on large-scale publicly available datasets, significantly reducing the need for large-scale datasets, thereby lowering annotation costs and time, and improving model efficiency.

This study utilizes the ResNet-50 architecture, pre-trained on ImageNet, as the main feature extractor. Subsequently, the network undergoes fine-tuning using a multi-modal dataset. To preserve the model's ability to capture subtle features, we froze the low-level parameters of ResNet-50 and only fine-tune the high-level parameters corresponding to stage 4. Furthermore, the remaining parts of the network are updated to improve their adaptability to the specific task.

IV. Experiments and Results

A. Experimental Settings

To assess the efficacy of the feature fusion and model integration strategies proposed in this paper for liver fibrosis grading, we have designed two experiments:

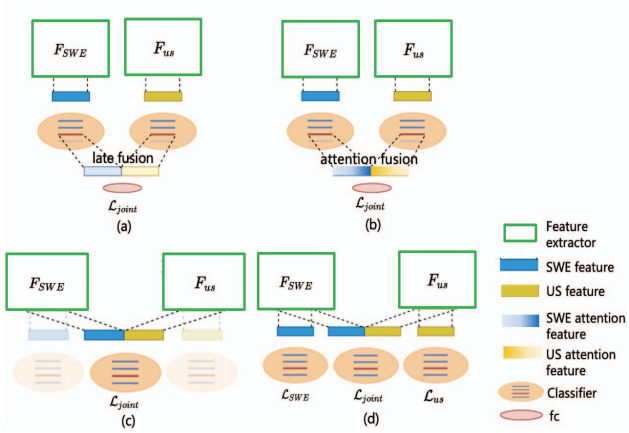

Fig. 5. Experimental settings.((**a**) Late fusion; (**b**) Attention fusion; (**c**) Single loss; (**d**) Multi loss)

1) Late fusion vs. Attention fusion: Our proposed attention-based fusion strategy (Fig. 5(b)) involves directly feeding the fused features into a classifier composed of a single fully connected layer. This is consistent with the late fusion approach (Fig. 5(a)), which also employs a single fully connected layer after concatenating the features. With an identical classifier structure, comparing the classification performance between these two approaches allows us to quantify the contribution of the attention mechanism in enhancing liver fibrosis diagnosis performance.

2) Single loss vs. Multi loss: To investigate the impact of model integration on the performance of liver fibrosis grading, we compare the classification performance of the model before and after integration [33]. For the non-integrated model, we evaluate the individual branches of fusion, corresponding to the losses $\mathcal{L}_{\text{joint}}$ (single loss) as depicted in Fig. 5(c). For the integrated model, we employ a combined decision-making approach using all three branches, corresponding to the total loss $\mathcal{L}_{\text{total}}$ (multi loss), as illustrated in Fig. 5(d).

B. Implementation Details

The training and inference processes were implemented using the PyTorch deep learning framework. The experiments were conducted on a server equipped with an Intel(R) Xeon(R) Gold 6430 CPU, featuring 16 vCPUs and a Nvidia GeForce RTX 4090 GPU with 24GB of GPU memory. And the operating system was Ubuntu 18.04.

The models are trained with the Adam optimizer for 40 epochs. The learning rate, batch size and weight decay are 7e-5, 4 and 0.001, respectively. For our model training and evaluation, we selected 80% of the images from the dataset at random to serve as the training set, while the remaining 20% were designated as the test set.

C. Evaluation Metrics

To evaluate the performance of our model, we employ a suite of evaluation metrics, which includes the Area Under the Receiver Operating Characteristic Curve (AUC) [34], as well as accuracy, sensitivity [35], specificity [36], and the confusion matrix. These metrics collectively furnish a thorough assessment of the model's performance. Sensitivity is defined as:

$$sensitivity = \frac{TP}{TP + FN} \tag{4}$$

Sensitivity measures the proportion of true positives that are correctly identified by the classifier, thereby quantifying the model's ability to detect positive cases accurately. Specificity is defined as:

$$specificity = \frac{TN}{TN + FP} \tag{5}$$

Specificity, which assesses the proportion of true negatives that are correctly identified by the classifier, serves as a measure of the model's ability to discern negative cases accurately. For the AUC metric, in the context of fibrosis grading, a five-class problem, the AUC values for each class are calculated by treating the other two classes as a single class. Subsequently, ROC curves are plotted for each class, and the average AUC across all classes is computed to obtain the overall AUC value.

D. Results and Discussion

Table 1 presents the results of comparative experiments.The first column of the table represents different experimental settings, where "late" denotes the late fusion strategy, "att" represents the attention fusion method proposed in this work, "single" indicates the single-model approach using a Single loss ($\mathcal{L}_{\text{joint}}$), and "multi" indicates the ensemble model using Multi-loss ($\mathcal{L}_{\text{total}}$).

It demonstrates that models employing a single imaging modality exhibit significantly poorer performance compared to those that integrate multiple imaging modalities. This corroborates the existence of complementary features across different imaging modalities, which can enhance classification performance. The multimodal strategy offers a wealth of diagnostic information, indicating significant potential for multimodal fusion-based imaging diagnostic techniques.

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT EXPERIMENTAL CONFIGURATIONS

Experimental Settings	Performance Metrics			
	Accuracy	*AUC*	*Sensitivity*	*Specificity*
US	55.10%	89.86	0.56	0.69
SWE	52.42%	91.37	0.97	0.68
US+SWE+late+single	61.32%	91.77	0.76	0.84
US+SWE+late+multi	62.45%	91.93	0.82	0.85
US+SWE+att+single	73.86%	93.85	0.83	0.87
***US+SWE+att+multi**	**76.21%**	**95.01**	**0.97**	**0.86**

In the comparative experiments of fusion methods, our proposed attention-weighted feature fusion strategy shows significant improvements in classification performance metrics such as accuracy and AUC compared to the late fusion strategy [37]. This demonstrates that fusion after attention weighting helps the model observe feature regions that are effective for improving classification performance.

Our model achieves an accuracy of 76.21%, representing a 6% improvement over the state-of-the-art (SOTA) method [25], which employs US and SWE images for five-class fibrosis staging.

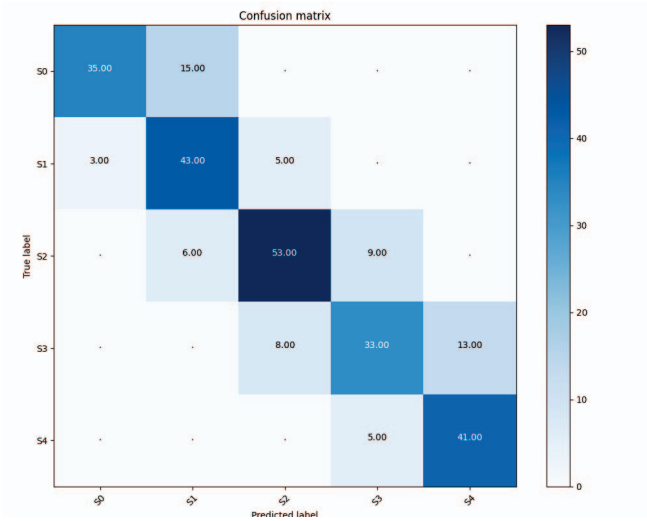

Fig. 6. Confusion matrix of our model.

Moreover, the attention mechanism proposed in this work is relatively simple, which significantly improves the model's performance while adding little complexity. For the medical imaging research field with small-scale datasets, our model converges after training for 20 epochs, with a model size of approximately 516M. Compared to large visual models based on transformer structures [38], our model is easy to train and more lightweight. In terms of models with similar parameter sizes, we achieve better performance in liver fibrosis diagnosis.

The comparison results from the model ensemble experiments demonstrate that the performance of the ensemble model is improved compared to individual models, indicating the effectiveness of our proposed model ensemble strategy in liver fibrosis staging tasks. Combining predictions from multiple models reduces the bias of individual models, enhances the robustness of the model, and improves its generalization ability on the test dataset.

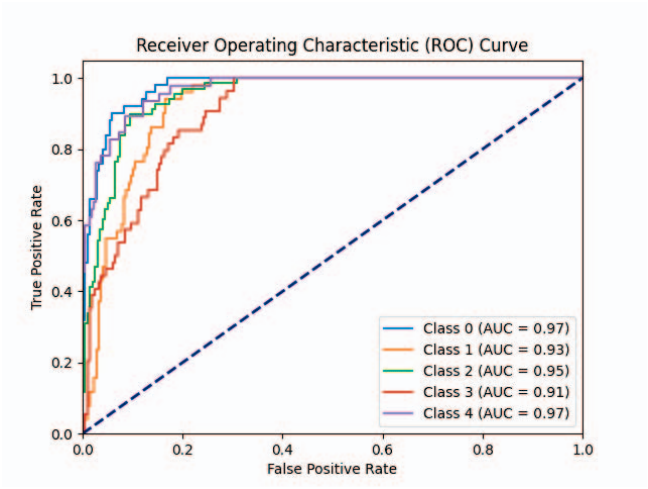

Fig. 7. ROC of our model.

In Fig. 6, the confusion matrix intuitively demonstrates our model's classification ability in different categories.The values in the cells along the diagonal represent the number of correctly classified instances for each liver fibrosis level, while the color intensity indicates the corresponding percentage accuracy. The model performs well in staging F2 and F4, but tends to misclassify F3 as adjacent stages F2 or F4. As shown in Fig. 7, the AUC values for each category are all above 0.9, indicating the model's overall high performance in liver fibrosis staging. Notably, in certain categories, the AUC surpassed 0.95, highlighting the model's excellent classification performance in these specific categories.

V. CONCLUSION

In this paper, we proposed a multi-modal framework for liver fibrosis diagnosis, capable of handling two modalities of liver fibrosis images and achieving automated fibrosis classification. The model takes into account the influence of different modal images on fibrosis staging, guiding the model to focus

on effective feature areas through an attention mechanism and encouraging interaction between different modal branches. By integrating models from different modal branches and employing a consistency loss function to constrain the modal features extracted by different branches, we enhanced the model's generalization ability and robustness to noisy data.

Experimental results demonstrated that our method outperforms other studies focusing on the five-class fibrosis problem using SWE and B-mode modalities, showing relatively better liver fibrosis staging performance. Moreover, our model has significant potential in lightweight deployment and saving computational resources. In the future, fusion strategies considering the complementary relationship between different modalities need to be designed, and the recognition ability for adjacent fibrosis stages should be improved as required.

VI. Acknowledgment

This work is supported in part by the National Key Research and Development Program of China under grant No.2023YFB2807001 the National Natural Science Foundation of China under grant No.12373113 and grant No.62004201, the Shanghai Talent Development Fund under Grant No.E1322E1, and the Research on Key Technologies of Energy Spectrum Detection Electronics Equipment Based on AI under grant No.E2520F1.

References

[1] Lin H, Zhang X, Li G, Wong GL, Wong VW. Epidemiology and Clinical Outcomes of Metabolic (Dysfunction)-associated Fatty Liver Disease. J Clin Transl Hepatol. 2021 Dec 28;9(6):972-982.

[2] Ng, Cheng Han, et al. "Mortality outcomes by fibrosis stage in nonalcoholic fatty liver disease: a systematic review and meta-analysis." Clinical Gastroenterology and Hepatology 21.4 (2023): 931-939.

[3] Mohamadnejad M, Tavangar S M, et al. Histopathological study of chronic hepatitis B: a comparative study of Ishak and METAVIR scoring systems[J]. Intl. J. of organ transplantation medicine, 2010, 1(4): 171.

[4] Qiu M, Guo M, et al., "Loop scheduling and bank type assignment for heterogeneous multi-bank memory", JPDC, 69 (6), 546-558, 2009

[5] Qiu M, Zhang K, Huang M, Usability in mobile interface browsing, Web Intell. and Agent Systems: An Intl. J. 4 (1), 43-59, 2006

[6] Qiu M, Ming Z, et al., Phase-change memory optimization for green cloud with genetic algorithm, IEEE Transactions on Computers 64 (12), 3528-3540, 2015

[7] Qiu M, Dai W, Vasilakos A, Loop parallelism maximization for multimedia data processing in mobile vehicular clouds, IEEE Transactions on Cloud Computing 7 (1), 250-258, 2016

[8] Song Y, Li Y, et al. Retraining strategy-based domain adaption network for intelligent fault diagnosis, IEEE Trans. on Industrial Informatics 16 (9), 6163-6171, 2019

[9] Qiu M, Chen Z, et al., Data allocation for hybrid memory with genetic algorithm, IEEE Transactions on Emerging Topics in Computing 3 (4), 544-555, 2015

[10] Zhang J, Li H, et al., Decouple and Decorrelate: A Disentanglement Security Framework Combing Sample Weighting for Cross-Institution Biased Disease Diagnosis," IEEE IoTJ, 2024

[11] Zhang Y, Qiu M, et al., Health-CPS: Healthcare cyber-physical system assisted by cloud and big data, IEEE Systems Journal 11 (1), 88-95, 2015

[12] Qiu M, Qiu H, Review on image processing based adversarial example defenses in computer vision, IEEE 6th BigDataSecurity, 2020

[13] Zhang Y, Qiu M, and Gao H, Communication-Efficient Stochastic Gradient Descent Ascent with Momentum Algorithms, IJCAI 2023.

[14] Ling C, Jiang J, et al., Deep Graph Representation Learning and Optimization for Influence Maximization, ICML 2023

[15] Li C and Qiu M, Reinforcement learning for cyber-physical systems: with cybersecurity case studies, CRC Press, 2019

[16] Noureddin M, Loomba R. Nonalcoholic fatty liver disease: Indications for liver biopsy and noninvasive biomarkers[J]. Clinical liver disease, 2012, 1(4): 104-107.

[17] Khvostikov A, Krylov A, Kamalov J, et al. Ultrasound despeckling by anisotropic diffusion and total variation methods for liver fibrosis diagnostics[J]. Signal Processing: Image Communication, 2017, 59: 3-11.

[18] Choi K J, Jang J K, Lee S S, et al. Development and validation of a deep learning system for staging liver fibrosis by using contrast agent–enhanced CT images in the liver[J]. Radiology, 2018, 289(3): 688-697.

[19] Yasaka K, Akai H, Kunimatsu A, et al. Liver fibrosis: deep convolutional neural network for staging by using gadoxetic acid–enhanced hepatobiliary phase MR images[J]. Radiology, 2018, 287(1): 146-155.

[20] Zhou Y, Ren X, Zheng X, et al. Federated-Learning-based Hierarchical Diagnosis of Liver Fibrosis, IEEE 7th SmartCloud, 2022: 109-114.

[21] Wang K, Lu X, Zhou H et al (2019) Deep learning radiomics of shear wave elastography significantly improved diagnostic perfor- mance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. Gut 68:729–741

[22] Lee J H, Joo I, Kang T W, et al. Deep learning with ultrasonography: automated classification of liver fibrosis using a deep convolutional neural network[J]. European radiology, 2020, 30: 1264-1273.

[23] Feng X, Chen X, Dong C, et al. Multi-scale information with attention integration for classification of liver fibrosis in B-mode US image[J]. Computer Methods and Programs in Biomedicine, 2022, 215: 106598.

[24] Kagadis G C, Drazinos P, Gatos I, et al., Deep learning networks on chronic liver disease assessment with fine-tuning of shear wave elastography image sequences[J]. Physics in Medicine & Biology, 2020, 65(21): 215027.

[25] Gao L, Zhou R, Dong C, et al. Multi-modal active learning for automatic liver fibrosis diagnosis based on ultrasound shear wave elastography, IEEE 18th Intl. Symposium on Biomedical Imaging (ISBI), 2021: 410-414.

[26] Lipkova J, Chen R J, Chen B, et al. Artificial intelligence for multimodal data integration in oncology[J]. Cancer Cell, 2022, 40(10): 1095-1110.

[27] Eltaeib T, Islam N. Taxonomy of challenges in cloud security, 8th IEEE CSCloud/EdgeCom, 2021: 42-46.

[28] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.

[29] Wu J, Huang X, Liu J, et al., NLP Research Based on Transformer Model, IEEE 10th CSCloud/IEEE 9th EdgeCom, 2023: 343-348.

[30] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[31] Karimi D, Dou H, Warfield S K, et al. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis[J]. Medical image analysis, 2020, 65: 101759.

[32] Weiss K, Khoshgoftaar T M, Wang D D. A survey of transfer learning[J]. Journal of Big data, 2016, 3: 1-40.

[33] Jana, Ananya, et al. "Deep learning based nas score and fibrosis stage prediction from ct and pathology data." 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE). IEEE, 2020.

[34] Korte J C, Cardenas C, Hardcastle N, et al. Radiomics feature stability of open-source software evaluated on apparent diffusion coefficient maps in head and neck cancer[J]. Scientific reports, 2021, 11(1): 17633.

[35] Hossin M, Sulaiman M N. A review on evaluation metrics for data classification evaluations[J]. International journal of data mining & knowledge management process, 2015, 5(2): 1.

[36] Avanzo M, Wei L, Stancanello J, et al. Machine and deep learning methods for radiomics[J]. Medical physics, 2020, 47(5): e185-e202.

[37] Zhang Y, Yang J, Tian J, et al. Modality-aware mutual learning for multi-modal medical image segmentation[C]//Medical Image Computing and Computer Assisted Intervention– MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. Springer International Publishing, 2021: 589-599.

[38] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.