Federated-Learning-based Hierarchical Diagnosis of Liver Fibrosis

Yueying Zhou*, Xinping Ren†✉, Xiaoying Zheng*, Yongxin Zhu*✉, Kang Xu*, Shijin Song‡, Li Tian*✉
*Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai, China
†Ultrasound Department, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China
‡Shanghai Jiaotong University, Shanghai, China

***Abstract*—Hepatic fibrosis is an important prognostic factor as severe liver fibrosis may lead to liver cancer or even death. To grade liver fibrosis, ultrasound gray-scale images and ultrasound elastic images are commonly used in clinical diagnosis to judge the severity of liver fibrosis. However, these two diagnoses methods are often vulnerable to disturbances, such as personal experience or instrument differences. Moreover, these individual differences usually lead to conflicting stand-alone machine learning diagnosis models at each hospital whose medical data are not allowed to share in public due to data privacy. To handle the conflicts among diagnosis models, we propose a federated learning based hierarchical diagnosis method of liver fibrosis by utilizing shear wave elasticity pictures of multiple users across hospitals without sharing the original data. Our method is validated with authentic shear wave elasticity pictures of hepatic fibrosis patients in Shanghai, China. Experimental results show that our method is able to preprocess these shear wave elasticity pictures, train local diagnosis models at each hospital and securely consolidate into a shared global diagnosis model whose accuracy is over 70% with only a small dataset containing a few hundreds of labeled pictures. Our method is expected to further improve in its accuracy with more training samples. Our method would be the first practice based on federated learning in liver fibrosis diagnosis.**

***Index Terms*—Federated learning, liver fibrosis diagnosis, shear wave elastography, deep learning, Nonalcoholic fatty liver disease**

I. Introduction

In recent years, Artificial Intelligence (AI) [1], [2] has made many attempts in the field of medical images. Deep learning methods [3], [4] have been widely used in model prediction and image recognition. One important application for AI is liver diagnosis. We apply novel techniques, such as Federated-Learning-based techniques to deal with NAFLD (*Nonalcoholic fatty liver disease*).

NAFLD is the most common liver disease in the world, also accounting for about 50% of the causes of chronic liver disease in China [5]. At present, more than one quarter of the world's adult population has NAFLD, of which 15%-30% with *nonalcoholic steatohepatitis* (NASH) [8] [9]. Simple NAFLD (also known as simple fatty liver) is a relatively benign liver disease, which is easy to reverse and progresses slowly. But once Simple NAFLD develops into NASH, it has a much higher possibility to develop to liver cirrhosis or even liver cancer [10]. So, it is of great importance to judge the prognosis of NAFLD patients. Symptoms of NAFLD include simple fatty liver, NASH, and liver fibrosis, among which liver fibrosis is the most important prognostic factor [11].

So far, the diagnosis of NASH can only be made by liver biopsy. However, invasive liver biopsy is difficult to be routinely used in distinguishing simple fatty liver and NASH patients. The severity of liver fibrosis is the only histological indicator that can independently improve the long-term prognosis of NAFLD patients [12]. Detection rate of steatohepatitis in NAFLD patients is positively correlated with the stage of liver fibrosis, confirmed by liver biopsy. It increased from 35% in patients without liver fibrosis (S0) to 94% in patients with cirrhosis (S4). Also, detection rate of advanced liver fibrosis (S3-4) in patients with NASH is significantly higher than that in patients with simple fatty liver (17%-21% vs 2%-11%) [13].

Therefore, accurate assessment of the degree of liver fibrosis during the dynamic development of NAFLD is very important to improve the prognosis and prolong the life of patients. Apart from liver biopsy (or liver puncture), there are many other clinical tools to help diagnose liver diseases. Having the advantages of real-time, non-invasive, low price, good repeatability and high popularity, ultrasonography is widely used in the diagnosis of liver diseases. The two commonly used ultrasonic images are described as follows:

- Ultrasonic Gray-scale Image: In gray-scale images, the *echo intensity* (EI) of tissues in different regions is represented by different gray levels. The intensity of ultrasound echoes is divided into 256 levels, so as to be recorded in 8bit gray data. The stronger the echo is, the brighter the color is on the image, and the more likely it is to have hepatic stones and calcifications. A common ultrasonic gray-scale image is shown as in Figure 1.
- Ultrasonic Elastic Image: Ultrasonic elastography is an imaging method that uses the elastic information of biological tissue to reflect its hardness. The elasticity of soft tissue is large, while that of hard tissue is small. Traditional elastography exerts pressure on tissue by manual operation, which may lead to deviation in force, achieving poor repeatability.
- SWE: *Ultrasonic shear wave elasticity* (SWE), as a relatively new non-invasive quantitative assessment of liver fibrosis, has been applied more and more in clinical research in recent years. Calculating elastic modulus by measuring the propagation velocity of shear wave in the tissue, this technique exerts pressure without manual

978-1-6654-5179-6/22/$31.00 ©2022 IEEE
DOI 10.1109/SmartCloud55982.2022.00023

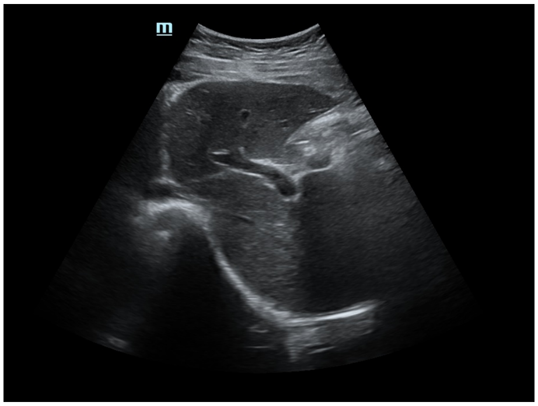

Fig. 1. Ultrasonic gray-scale image sample.

operation, unlike traditional ultrasonic elastography. SWE has good repeatability, high sensitivity and specificity. In this paper, we studied a dataset consists of STE images from patients in Shanghai Ruijin Hospital. STE, the sound touch elastography, is a new SWE technique developed by Mindray company. This technology adopts ultra-wide beam tracking and can receive the shear wave data in the whole frame at one time, as well as to achieve real-time two-dimensional shear wave elastic imaging. A sample STE image is shown as Figure 2.

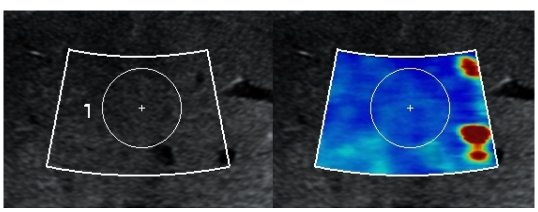

Fig. 2. STE image sample.

However, this method cannot diagnose liver fat with content less than 30% and early liver fibrosis. What's more, it also vulnerable to disturbances, such as operator subjectivity and instrument dependence. This is why people begin to use artificial intelligence to assist the classification of ultrasonic images.

Ananya et al. use the semi-supervised learning method to train liver fibrosis classification model on the MICCAI 2017 LiTS challenge dataset [14]. She also combines the information from CT and H&E stained pathology data to improve fibrosis stage prediction, and proves its effectiveness on the data of 30 patients [18]. K. Prakash et al. use 52 features to classify and predict liver fibrosis, such as *gray level co-occurrence matrix* (GLCM) texture feature and gradient co-occurrence matrix (GLGCM) texture feature, training on MRI images [19]. AI realizes the automatic screening of NAFLD patients, improves consistency of observers and reduces mis-diagnosis rate. However, at the moment, the application of artificial intelligence in medical imaging [6], [7] has the following difficulties:

- First, data privacy and communication security [15]–[17] are among major security concerns in secure applications of deep learning in distributed or web systems [20]. Medical image data have strict privacy protection requirements, so data cannot be shared across centers, even between doctors. AI, especially deep learning, needs a large dataset to train a good result. How to provide sufficient data to AI methods without sharing the original data is a concerned problem.
- Second, in the training process, the model is easy to become overfitted. During the training, the model complexity and errors on training data gradually decrease, but the errors on the validation set increase. How to avoid over fitting when the data set is limited is also a concern.

To tackle the above two difficulties, we use the method of federated learning to process medical images. In the scenario of federated learning, the data of each user are stored on their local devices without sharing with other users. In each algorithm iteration, users only need to upload the gradient information of the model to the central server. Then, server aggregates the received gradients, and the model converges after multiple iterations. One of the most well-known cases of federal learning is the Google search automatic matching system. This method is also applicable to the processing of medical images, where privacy protection is required. We applied several kinds of small deep learning models in the experiments, such as CNN, ResNet, and DenseNet. Although with limited ultra-sound images available, we still achieve an accuracy of 70% using resnet18 under the framework of federated learning distributed over five medical centers, at the same time providing privacy protection across these centers. Therefore, federated learning demonstrates its ability to improve the accuracy of quantitative diagnosis and exact staging of liver fibrosis in multi-center joint learning with privacy protection.

The rest of the paper is organized as follows. In Section 2, the training method is described. In Section 3, we present the experiment results. The conclusion is drawn in Section 4.

II. Our Methodology

Our hierarchical diagnosis method, based on federated learning, mainly consists of three major steps. The first step is image preprocessing, while the second step is data augmentation, and the third is federated learning. The overall work flow of our methodology is shown in Figure 3.

A. Image Preprocessing

Deep learning network extracts feature information from images through large-scale computation. Image preprocessing can eliminate irrelevant information in the image as much as possible, thus enhancing detectability of feature information, also improving accuracy of deep learning. The methods of enhancing low resolution medical images include domain

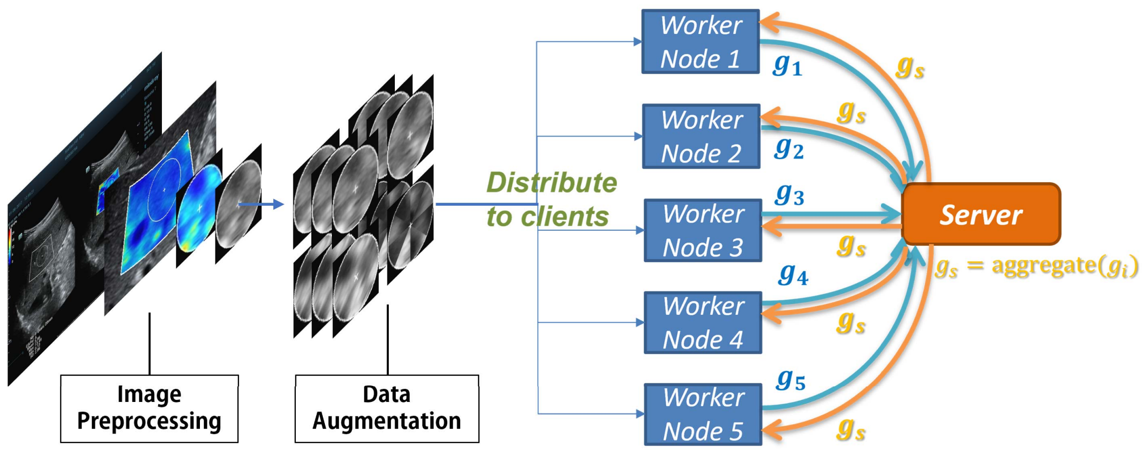

Fig. 3. A schematic diagram illustrating the process of one round training.

transformation, shape-adaptive edge enhancement and adaptive histogram equalization, etc. [25]

We also need to use preprocessing methods to avoid the interference of artifacts. In the original STE image, we can see a half sector colored area, which is the elastic measurement area. But elastic information is vulnerable to disturbances, and artifacts are very easy to be generated in ultrasonic images. Manifested by special addition, reduction, or distortion, the area where the artifact is formed will seriously interfere with liver fibrosis rating. Common artifacts include reverberation, partial volume effect, sound shadow, etc. These interferences may exist at the edge of the colored area or the place where there are blood vessels passing through.

- Greyscale Transformation. The color of STL image is added for readability, each pixel only records the pressure on that point, changing from 0 to 30kPa. So it is of no need to train the colored image with 24bit per pixel. We can transform it to greyscale image using the formula below:

$$
\begin{aligned}
grey = (1.5 - \frac{g}{255})r + 0.49g + 0.11b \\
-20(1 - \frac{g}{255})\frac{e^{10-r}}{1+e^{10-r}}
\end{aligned}
\tag{1}
$$

- Hough Transform. To avoid the interference of artifact, we only use the elastic data within the sampling frame, which is the circular frame in the semi sector. Hough transform is a feature extraction technique in image processing, it finds imperfect instances of objects by a voting procedure in a parameter space. In this paper, we use Hough transform to locate the accurate position of circular STE detection frame. After circle candidates are produced, then the function selects local maxima in an accumulator matrix. We generate a mask to cover the parts that are not in the circular frame, so as to improve the performance of training.

B. Data Augmentation

Sufficient data is significant to model training. In most deep learning scenarios, hundreds of thousands of pictures are trained to get a better accuracy. But in the medical field, samples are not so easy to get. The dataset we use is sampled from only 250 patients, which is a drop in the bucket even for the simplest model. Therefore, it is very important to use data augmentation methods to expand training data.

In addition to supplementing training data, data augmentation can also play a role as a regularizer, reducing overfitting and increasing robustness of the model. Common methods include adding noise to the picture, enlarging certain parts of the picture, clipping and splicing the image, etc. These methods enable the computer to ignore the minutiae and find the features that really need attention.

The techniques we apply are mentioned as follows:

- Random flip and Rotation. This is the most generally used data augmentation method. Once the data is referred, every image of which randomly flips horizontally or vertically. On this basis, our algorithm applies an additional random rotation of certain angle to the picture to further expand the dataset. The transformation results are shown in Figure 4-2,3.
- Slicing and Recombining. In order to generate enough images for training, we cut the images in training data and put the components of different images together to create new synthetic images for training. We can see the synthetic image in Figure 4-4.

C. Federated Learning

Federated learning is a distributed learning technique that trains across multiple edge devices. Federated learning is categorized into 3 types: horizontal federated learning, which is feature aligned; vertical federated learning, which is sample aligned, and federated transfer learning. In this paper, we use horizontal federated learning approach to train STE images.

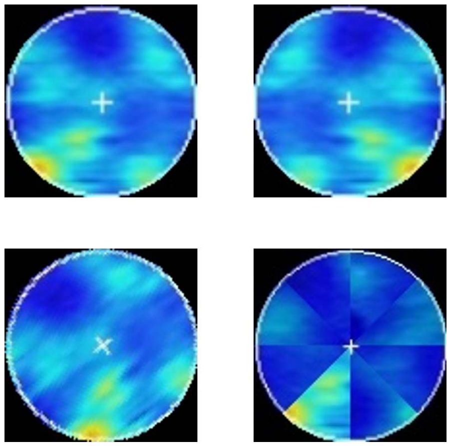

Fig. 4. Data Augmentation $^{1\,2}_{3\,4}$

These images are distributed to five different users. The workflow of federal learning is shown in Figure 5.

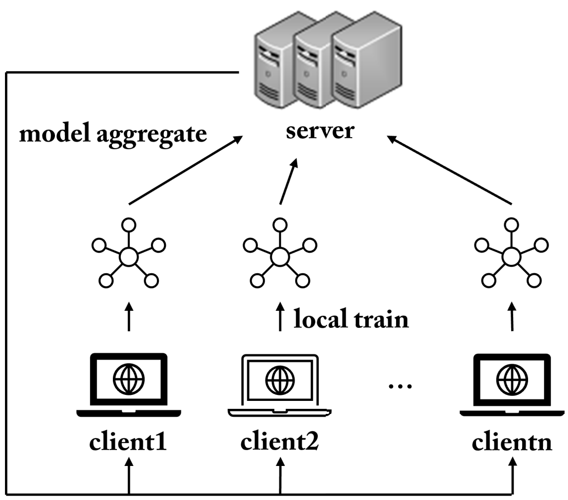

Fig. 5. Federated Learning

During the training process, server and each user (or worker node) holds part of the data respectively. Edge devices cannot communicate with each other, all they can do is to upload gradients to the server. These gradients are aggregated to upgrade model parameters. In general, data volume has a great impact on the upper performance limit of the model, unbalanced data distribution also leads to poor robustness. Therefore, users with better data will be assigned greater weights, in order to achieve better performance.

$$g_s = w_1 g_1 + w_2 g_2 + ... + w_n g_n \tag{2}$$

After the central server distributes the initialized model, each user uses local data to calculate gradient g_i respectively. Users then upload their gradient to the server. After receiving all the gradients, server performs a weighted average on them to calculate the parameters of the global model, and then distributes the new parameters to each user. These steps repeat until the model converges. The details are shown in Algorithm 1.

Algorithm 1 Federated Learning Algorithm.

for $i = 0, i < client\_number, i = i + 1$ **do**
  $model[i] \leftarrow initial\_model$
end for
$rounds \leftarrow 200, client\_number \leftarrow 5$
for $round = 0, round < rounds, round = round + 1$ **do**
  for $i = 0, i < client\_number, i = i + 1$ **do**
    $model[i] \leftarrow model[i] + params$
    $new\_model[i] \leftarrow train(model[i], local\_data[i])$
    $g[i] \leftarrow new\_model[i] - model[i]$
    $upload(g[i])$
  end for
  $g_s \leftarrow weighted\_avg(g)$
  $global\_model \leftarrow update(g_s)$
  $params \leftarrow params - g_s$
  $broadcast(params)$
end for

Following these procedures, we have realized the training of deep learning network from a small set of data sets under the premise of privacy protection.

III. Performance Evalution

A. Training results

We obtained STE images of around 250 patients with liver diseases from Shanghai Ruijin Hospital. The degree of liver fibrosis within these patients is divided into three grades. 0 stands for no liver fibrosis, while 2 for cirrhosis, degree gradually deepened with the increase of number. We distributed these images to five clients and used several models to carry forward, including ResNet18, ResNet34, DenseNet121 and VGG16. Although data augmentation methods have been used to increase the training data, the amount of original data is still too small after all. In order to prevent the model from failing to converge, I specially introduced three-layer CNN as a control experiment.

We trained more than 150 rounds on each model, the accuracy of the two ResNet networks stabilized at around 70%. VGG16 also trained with moderate results. But the performance of 3-layer-CNN and DenseNet was mediocre. These two models cannot converge during the training. The reasons why these two models can't converge are completely opposite. The 3-layer CNN cannot converge because of too few layers, while DenseNet for too many layers.

Figure 6 is a line chart showing the change of model accuracy with the number of training rounds, and Figure 7 is a line chart showing the change of cross entry loss of each client during the first 150 rounds of training. We can see from the two figures, at about 150 rounds, loss gradually approaches 0, and the recognition rate tends to be stable.

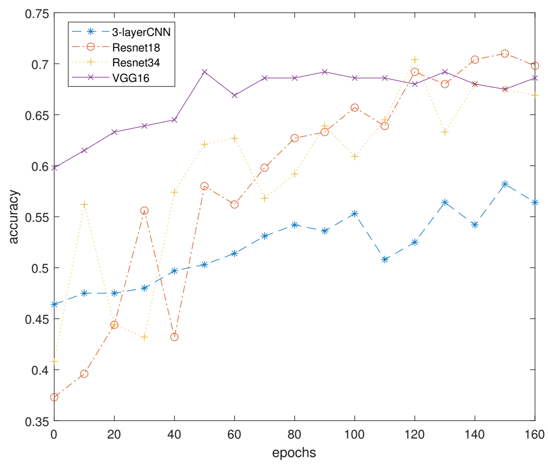

Fig. 6. model accuracy increasing with the training rounds

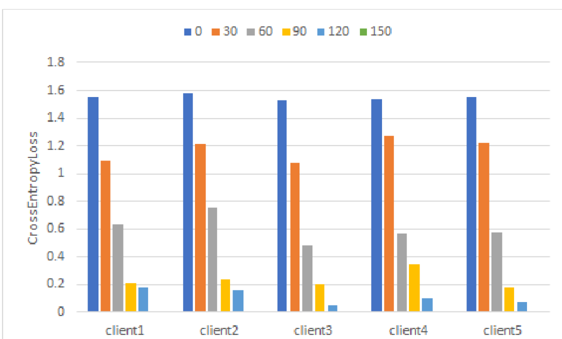

Fig. 7. clients' cross entropy loss decaying with the training rounds

B. Communication cost

Communication cost [21], [24] is positively related to the size of the model. When using ResNet, each communication takes about 1s, and the data exchanged is about 224MB. But if we change ResNet to VGG16, the model of which is much larger than ResNet, each communication time change to 3.5s, and the data exchanged increase to nearly 3GB. So, if the model we use become more and more complex, data transmission will inevitably reach the physical upper limit, resulting in communication congestion [23], [24]. In the follow-up study, a better scheme should be adopted to reduce the size of transmission data.

C. Computation cost

Computation cost [26], [27] is positively related to the size of dataset. During the training, the memory consumption of the GPU is 1700MB, while the processed STE pictures only occupy less than 10MB of memory in total. Therefore, it is necessary to crop the images and use masks to filter out measurement frames, otherwise the computational cost will be greater [28], [29].

IV. Conclusion

Early detection and accurate assessment of the degree of liver fibrosis are crucial to improve the prognosis of patients. It is of high value to develop deep learning technology to assist ultrasound shear-wave elasticity technology to determine the prognosis of liver fibrosis grading based on NAFLD, but there are still the following difficulties: First, the model is easy to overfit due to the limited size of the data set; Second, medical image data has the need of privacy protection, and it is difficult to share directly among multi-centers. In response to the above challenges, we use the federated learning method to process STE image data. We make the images scattered and stored on different client-computers and train the model without exchanging the original data between clients. Considering that there may be artifacts and other interferences, we designed a mask to filter out the circle that applies STE. We compared and tested how different deep learning models perform in our system. Although the total amount of data is small, we achieve the accuracy of 70% using resnet18.

How to better apply federated learning in processing STE image is still under study. For the current data set, the utilization of our algorithm is not enough. Subsequently, we plan to add ultrasonic gray-scale images to the training and improve the aggregation method of central server to enhance the accuracy of STE image classification and achieve better performance.

V. Acknowledgment

This work was supported in part by National Key R&D Program of China (grant no. 2019YFB2204204), National Natural Science Foundation of China (grant no. U2032125), Shanghai Talent Development Fund (grant no. E1322E1). Sincere gratitude to all the people who helped me during this period.

References

[1] F. Hu, S. Lakdawala, Q. Hao, M. Qiu, Low-power, intelligent sensor hardware interface for medical data preprocessing, IEEE Trans. on Information Technology in Biomedicine 13 (4), 656-663, 2009

[2] H. Qiu, M. Qiu, R. Lu, "Secure V2X communication network based on intelligent PKI and edge computing", IEEE Network, 34 (2), 172-178, 2019

[3] H. Qiu, Q. Zheng, et al., "Topological graph convolutional network-based urban traffic flow and density prediction", IEEE Trans. on ITS, 2020

[4] Y. Li, K. Gai, Z. Ming, H. Zhao, M. Qiu, "Intercrossed access controls for secure financial services on multimedia big data in cloud systems", ACM Trans. on Multimedia Comp., Comm., and App., 2016

[5] Z. M. Younossi, M. Otgonsuren, L. Henry, C. Venkatesan, A. Mishra, Erario M, Hunt S. Association of nonalcoholic fatty liver disease (NAFLD) with hepatocellular carcinoma (HCC) in the United States from 2004 to 2009[J]. Hepatology 2015,62(6): 1723-1730.
[6] K. Gai, M. Qiu, L. Chen, M. Liu, "Electronic health record error prevention approach using ontology in big data", IEEE 17th HPCC, 2015
[7] M. Qiu, E. Khisamutdinov, et al., "RNA nanotechnology for computer design and in vivo computation," Philosophical Transactions of the Royal Society A, 2013
[8] Z. M. Younossi, A. B. Koenig, D. Abdelatif, Y. Fazel, L. Henry, M. Wymer. Global epidemiology of nonalcoholic fatty liver disease—Meta-analytic assessment of prevalence, incidence, and outcomes[J]. Hepatology, 2016, 64(1).
[9] Z. Younossi et al., Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention[J]. Nature Reviews Gastroenterology & Hepatology, 2018.
[10] G. A. Michelotti, M. V. Machado, and A. M. Diehl. NAFLD, NASH and liver cancer. Nat Rev Gastroenterol Hepatol (2013).
[11] N. Li, Y. Jiang, G. Gong, G. Han and J. Ma, Non-Invasive Assessment Model of Liver Disease Severity by Serum Markers Using Cloud Computing and Internet of Things, in IEEE Access, vol. 6, pp. 33969-33976, 2018, doi: 10.1109/ACCESS.2018.2849016.
[12] V. Nobili, A. Alisi, A. Vania, C. Tiribelli, A. Pietrobattista, G. Bedogni. The pediatric nafld fibrosis index: a predictor of liver fibrosis in children with non-alcoholic fatty liver disease[J]. BMC Medicine,7,1(2009-05-01), 2009, 7(1):1-7.
[13] Liang Xu. Pay attention to the evaluation, prevention, and treatment of liver fibrosis in nonalcoholic fatty liver disease [J]. Journal of Doctors Online (Chinese), 2019,9(27): 27-28
[14] A. Jana, H. Qu, C. D. Minacapelli, C. Catalano, V. Rustgi and D. Metaxas, Liver Fibrosis And NAS Scoring From CT Images Using Self-Supervised Learning And Texture Encoding, 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), 2021, pp. 1553-1557, doi: 10.1109/ISBI48211.2021.9433920.
[15] H. Qiu, T. Dong, T. Zhang, et al., "Adversarial attacks against network intrusion detection in IoT systems," IEEE Internet of Things Journal 8(13), 10327-10335, 2020
[16] K. Gai, M. Qiu, S. Elnagdy, "A novel secure big data cyber incident analytics framework for cloud-based cybersecurity insurance," IEEE BigDataSecurity 2016
[17] M. Qiu, H. Su, M. Chen, Z. Ming, L. Yang, "Balance of security strength and energy for a PMU monitoring system in smart grid", IEEE Comm. Mag., 50 (5), 142-149,2012
[18] A. Jana, H. Qu, P. Rattan, C. D. Minacapelli, V. Rustgi and D. Metaxas, Deep Learning based NAS Score and Fibrosis Stage Prediction from CT and Pathology Data, 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), 2020, pp. 981-986, doi: 10.1109/BIBE50027.2020.00166.
[19] K. Prakash and S. Saradha, A Deep Learning Approach for Classification and Prediction of Cirrhosis Liver: Non Alcoholic Fatty Liver Disease (NAFLD), 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), 2022, pp. 1277-1284, doi: 10.1109/ICOEI53556.2022.9777239.
[20] W. Chen, Y. Zeng and M. Qiu, Using Adversarial Examples to Bypass Deep Learning Based URL Detection System, 2019 IEEE International Conference on Smart Cloud (SmartCloud), 2019, pp. 128-130, doi: 10.1109/SmartCloud.2019.00031.
[21] M. Qiu, Z. Chen, Z. Ming, X. Qin, J. Niu, "Energy-aware data allocation with hybrid memory for mobile cloud systems", IEEE Systems J., 11 (2), 813-822, 2014
[22] J. Li, Z. Ming, et al., "Resource allocation robustness in multi-core embedded systems with inaccurate information", Journal of Systems Architecture 57 (9), 840-849, 2011
[23] M. Qiu, C Xue, Z Shao, et al., "Efficient algorithm of energy minimization for heterogeneous wireless sensor network", IEEE EUC, 25-34, 2006
[24] J. Niu, Y. Gao, et al., "Selecting proper wireless network interfaces for user experience enhancement with guaranteed probability", JPDC, 72(12), 1565-1575, 2012
[25] S. Agaian, P. Rad, R. Rajendran and K. Panetta, A Novel Technique to Enhance Low Resolution CT and Magnetic Resonance Images in Cloud, 2016 IEEE International Conference on Smart Cloud (SmartCloud), 2016, pp. 73-78, doi: 10.1109/SmartCloud.2016.34.
[26] M. Qiu, Z. Jia, C. Xue, Z. Shao, E. Sha, "Voltage assignment with guaranteed probability satisfying timing constraint for real-time multiproceesor DSP", Journal of VLSI Signal Processing Systems, 2007
[27] M. Qiu, L. Yang, Z. Shao, E. Sha, "Dynamic and leakage energy minimization with soft real-time loop scheduling and voltage assignment", IEEE Transactions on Very Large Scale Integration (VLSI) Systems 18 (3), 501-504, 2009
[28] M. Qiu, C. Xue, Z. Shao, E. Sha, "Energy minimization with soft real-time and DVS for uniprocessor and multiprocessor embedded systems," IEEE DATE Conf., 1-6, 2007
[29] M. Qiu, H. Li, E. Sha, "Heterogeneous real-time embedded software optimization considering hardware platform", ACM sym. on Applied Comp., 1637-1641, 2009