TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 01/22 pp1–12
DOI: 10.26599/TST.2022.9010016
Volume 29, Number 1, February 2024

Deep Learning Blockchain Integration Framework for Ureteropelvic Junction Obstruction Diagnosis Using Ultrasound Images

Yu Guan, Pengceng Wen, Jianqiang Li*, Jinli Zhang, and Xianghui Xie

Abstract: UreteroPelvic Junction Obstruction (UPJO) is a common hydronephrosis disease in children that can result in an even progressive loss of renal function. Ultrasonography is an economical, radiationless, noninvasive, and high noise preliminary diagnostic step for UPJO. Artificial intelligence has been widely applied to medical fields and can greatly assist doctors' diagnostic abilities. The demand for a highly secure network environment in transferring electronic medical data online, therefore, has led to the development of blockchain technology. In this study, we built and tested a framework that integrates a deep learning diagnosis model with blockchain technology. Our diagnosis model is a combination of an attention-based pyramid semantic segmentation network and a discrete wavelet transformation-processed residual classification network. We also compared the performance between benchmark models and our models. Our diagnosis model outperformed benchmarks on the segmentation task and classification task with MIoU = 87.93, MPA = 93.52, and accuracy = 91.77%. For the blockchain system, we applied the InterPlanetary File System protocol to build a secure and private sharing environment. This framework can automatically grade the severity of UPJO using ultrasound images, guarantee secure medical data sharing, assist in doctors' diagnostic ability, relieve patients' burden, and provide technical support for future federated learning and linkage of the Internet of Medical Things (IoMT).

Key words: data mining; image processing and computer vision; machine learning; medical information systems

1 Introduction

As a common kidney disease in children, hydronephrosis is generally congenital[1, 2] and has increasing morbidity above 1%[3, 4]. UreteroPelvic Junction Obstruction (UPJO) is an obstruction at or along the pelvic-ureteral junction and is the main cause of children's hydronephrosis[5]. It can result in symptoms, such as nausea and vomiting, abdominal pain, abdominal mass, urinary tract infection, hematuria, uremia, hypertension, and other progressive loss of renal function, such as uremia or even renal rupture.

Ultrasonography is a commonly adopted examination in the medical field and a preliminary diagnostic step for UPJO, owing to its economical, radiationless, and noninvasive nature[6]. However, because ultrasonic images have high noise, it is laborious for even trained sonographers to obtain sufficient information[7, 8].

With the development of computing power, Artificial Intelligence (AI) technology has been applied to numerous regions, especially in medical image analysis. This technology can greatly improve the diagnostic ability of doctors and alleviate the medical resource imbalance and shortage in less-developed regions[9–14]. As a subset of AI, deep learning can learn from source data and determine the important features to automatically draw conclusions[15]. If we can combine

• Yu Guan, Pengceng Wen, Jianqiang Li, and Jinli Zhang are with the Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China. E-mail: guanyu0010@126.com; wpc18581311918@outlook.com; lijianqiang_bjut@126.com; jlzhangcs@bjut.edu.cn.
• Xianghui Xie is with Beijing Children's Hospital Affiliated to Capital Medical University, Beijing 100020, China. E-mail: xiexianghui_1@126.com.
∗ To whom correspondence should be addressed.
Manuscript received: 2022-03-14; revised: 2022-04-12; accepted: 2022-06-07

AI medical imaging technology and automatically distinguish diseased regions during the ultrasound examination stage, we can not only support urologist surgeons for further diagnosis, but also save considerable medical resources, manpower, and money to help burdened patients.

In the interdisciplinary field of AI and medical image processing, medical data are usually digitized (i.e., electronic medical records or electronic health records) and transmitted on the network[16–18]. In this kind of medical data sharing, a key challenge is the privacy protection of patients' sensitive information and the safe sharing of data by multi-party organizations[19]. As a shared ledger that provides a secure, decentralized, transparent, and trusted manner for data management, blockchain has been extensively applied in medical data with the arrival of the 5G era[20–26].

Previous literature that integrates deep learning and blockchain in the medical field has always focused on general and conceptual construction. We, however, not only propose a deep learning model that can diagnose the ultrasonic images of UPJO, but we also adopt a blockchain protocol to guarantee the secure transfer between different parities. The diagnosis model consists of an attention-based semantic segmentation network, the A-PSPNet, and a classification algorithm wavelet-CNN. After integrating this Deep Learning based Computer-Aided Diagnosis (DL-CAD) model with the InterPlanetary File System (IPFS) and a classic blockchain protocol, we finally built up the deep learning blockchain integrating the UPJO diagnosis framework. The performance of this framework is validated and reported.

Our study has several contributions:

• Our proposed deep learning diagnosis model can identify the UPJO severity according to high-noised, cheap, nonradiative, and noninvasive ultrasound images, and has an accuracy of 91.77%, which is higher than those of other benchmark models.

• We adopt a Discrete Wavelet Transformation (DWT) feature engineering method and replace the pooling layer in our deep learning network with it. This method increases the number of model input features and enhances the model performance.

• The blockchain system we formed ensures the online storage and transmission of sensitive data, and provides technical support for further federated learning and cooperation among multiple parties.

• The blockchain system enables patients to safely store and track their medical history whenever and wherever, even if they changed hospitals. It also helps build trust among institutions owing to its auditability and traceability.

• Moreover, the combination of the deep learning model and blockchain system can play the role of an online AI doctor. Without going to the hospital, and worrying about security, the patient can get a diagnosis with an accuracy rate of more than 91% by just uploading an ultrasound image to our deep learning blockchain framework. This process allows the linkage with the Internet of Medical Things (IoMT)[27], such as portable and ultraportable ultrasound scanners[28].

Our paper is organized as follows for a better understanding: In Section 2, we present the proposed model architecture and introduce every element in detail. In Section 3, we demonstrate the experimental setting and network training. In Section 4, we illustrate our performance and the comparison results with other networks. Finally, in Section 5, we discuss our whole study and present the conclusion in Section 6.

2 Methodology

In this study, we developed a DL-CAD model that combines a semantic segmentation network and classification network to grade the UPJO disease from hydronephrosis ultrasound images. Then, we adopt a blockchain protocol to integrate with our DL-CAD model for secure transmission between separate parties. In this section, we introduce our pathological grading scheme for UPJO ultrasound images, and the architecture is shown in Fig. 1. This scheme combines an attention-based semantic segmentation algorithm, Pyramid Scene Parsing Network (PSPNet), with a classification algorithm named wavelet-CNN

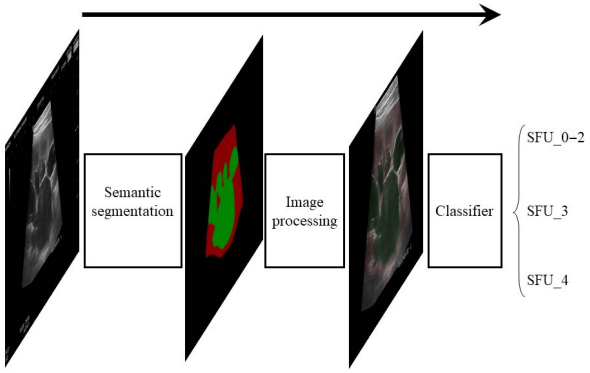


Fig. 1 Overall structure of our UPJO ultrasonic image diagnosis model. The input is the original image, and the output is the grading result.

(hereinafter referred to as the segmentation and the classifier) to increase the available representation of the images and realize the final classification of UPJO. In this study, we follow the classification system of congenital hydronephrosis that designed by the Society of Fetal Urology[29, 30] to class our data into five grades, SFU_0 to SFU_4.

We then separately demonstrate these parts as follows.

2.1 A-PSPNet: A semantic segmentation network

Because ultrasonic images have high and dense noise, the accuracy of a direct classification tends to become unsatisfactory. Hence, it is necessary to supplement feature representations that are different from noise. Our segmentation network adopts the PSPNet[29] and attention mechanism Convolutional Block Attention Module (CBAM)[30]. The overall structure is shown in Fig. 2. The attention layer will help the network ignore unnecessary features and focus on important features to improve the segmentation accuracy. The parsing structure of PSPNet has multilayer perceptrons, which enable the network to obtain more global feature information and then fuse it with the basic features to help distinguish different regions.

Using x denotes the input. The original input image is decomposed into a basic feature representation $\mathcal{R}(x)$ and advanced feature representation $\mathcal{P}_i(x)$, where $i \in (1, 2, 3, 6)$, i represents the pooling scales in pyramid pooling and can further adjust the feature size parameters for concatenation. The two kinds of feature maps are concatenated to form the structure of Fig. 2c as $\mathcal{H}$, where $C(\cdot, \cdot)$ stands for the concatenate processing,

$$\mathcal{H} = C\left\{\mathcal{R}(x), C\left(\mathcal{P}_i(x) \mid i \in (1, 2, 3, 6)\right)\right\} \quad (1)$$

Then, the concatenated feature layers are processed, predicted, and colored to output the segmentation results. As shown in Fig. 3, the concatenated features are sent through a bottleneck layer to adjust the number of channels, and then each pixel is predicted through a softmax function and colored[31]. The final segmentation results are shown as the output in Fig. 2.

2.2 ResNet-W: A DWT-utilized classifier

As previously mentioned, we add a classifier to acquire the UPJO pathological grading. In traditional classification networks, the pooling layer employs downsampling to decompose images and reduce the total amount of calculation, which, however, results in the loss of information. Because wavelet transformation and its inverse operation are reversible, thus guaranteeing no information loss when reducing the complexity. DWT[32], which is deemed as an effective operator to decompose images and obtain the desired frequency-

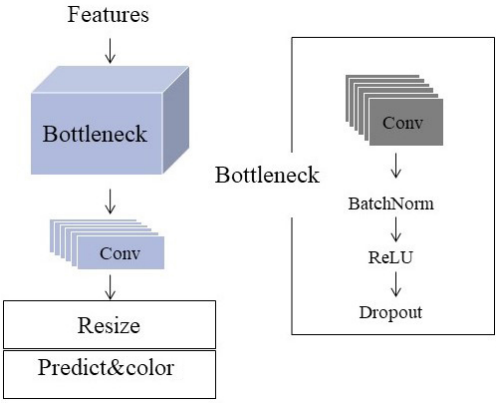


Fig. 3 Dimensionality reduction and prediction steps after feature concatenation.

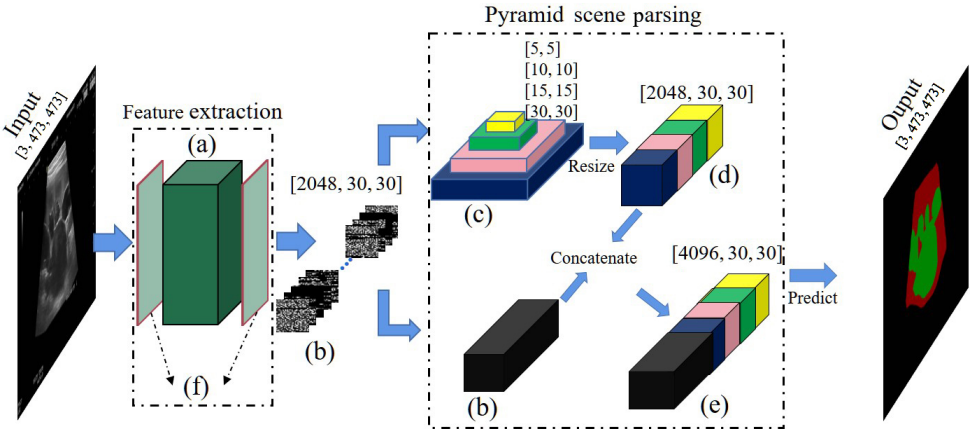


Fig. 2 Architecture of the A-PSPNet. The input is the original image, and the output is the segmentation result distinguished by color. (a) ResNet structure embedded in two CBAMs. (b) Basic feature of the output after (a). (c) Advanced feature of (b) after pyramid pooling. (d) Result of the unified dimension operation on (c). (e) Integrated structure of the features of modules of (b) and (d).

domain information, is therefore applied to the classifier to replace the traditional pooling layer. For a better understanding of the process in the two-dimensional (2D) wavelet transformation, we start with the one-dimensional (1D) ones. First, we define two basic functions, a wavelet function $\Psi(\cdot)$ and scale function $\varphi(\cdot)$. Take a 1D image of size $M \times 1$, for 1D discrete sequence $f(n)$, its discrete wavelet series expansion coefficient is

$$W_{\varphi}(j_0,k) = \frac{1}{\sqrt{M}} \sum_n f(n)\varphi_{j_0,k}(n) \tag{2}$$

$$W_{\Psi}(j,k) = \frac{1}{\sqrt{M}} \sum_n f(n)\Psi_{j,k}(n),\ j \geqslant j_0 \tag{3}$$

where $j = 0, 1, \ldots, k-1$, $k = 0, 1, \ldots, 2^j - 1$, and $n = 0, 1, \ldots, M-1$. k denotes the position in a given direction, j denotes the width of the sample, and j_0 is any determined starting scale. $\varphi_{j_0,k}(n)$ and $\Psi_{j,k}(n)$ are samples that use M equal intervals on the support region of the basis function on the J scale. Usually, if $j_0 = 0$ and the transform itself consists of M coefficients, then the minimum scale is 0, and the maximum scale is J,

$$\begin{aligned} \varphi_{j,k}(x) &= 2^{j/2}\varphi\left(2^j x - k\right), \\ \Psi_{j,k}(x) &= 2^{j/2}\Psi\left(2^j x - k\right) \end{aligned} \tag{4}$$

The obtained expanded set $\{\varphi_{j_0,k}(n)\}$ is a subset of $\{\varphi_{j,k}(n)\}$, and the final transformation itself is composed of M coefficients.

Similar to the 1D DWT, the 2D DWT uses the 2D scale function and wavelet function. We first take the 1D transformation of the row of the 2D array, and then take the 1D transformation of the column of the last step result. For a 2D image of size $M \times N$, its DWT is

$$W_{\varphi}(j_0,m,n) = \frac{1}{\sqrt{M \times N}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y)\varphi_{j_0,m,n}(x,y) \tag{5}$$

$$W_{\Psi}^{i}(j,m,n) = \frac{1}{\sqrt{M \times N}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y)\Psi_{j_0,m,n}^{i}(x,y), \quad i = \{H, V, D\} \tag{6}$$

Finally, we get four half-sized output images W_{φ}, W_{Ψ}^{H}, W_{Ψ}^{V}, and W_{Ψ}^{D}, which represent the average, horizontal, vertical, and diagonal information from the input source image, respectively.

We replace the first pooling layer in our residual classification network with the above 2D-DWT processing to construct our classifier, the ResNet-W, whose overall structure is shown in Fig. 4. Then, the four half-sized images are separately passed through

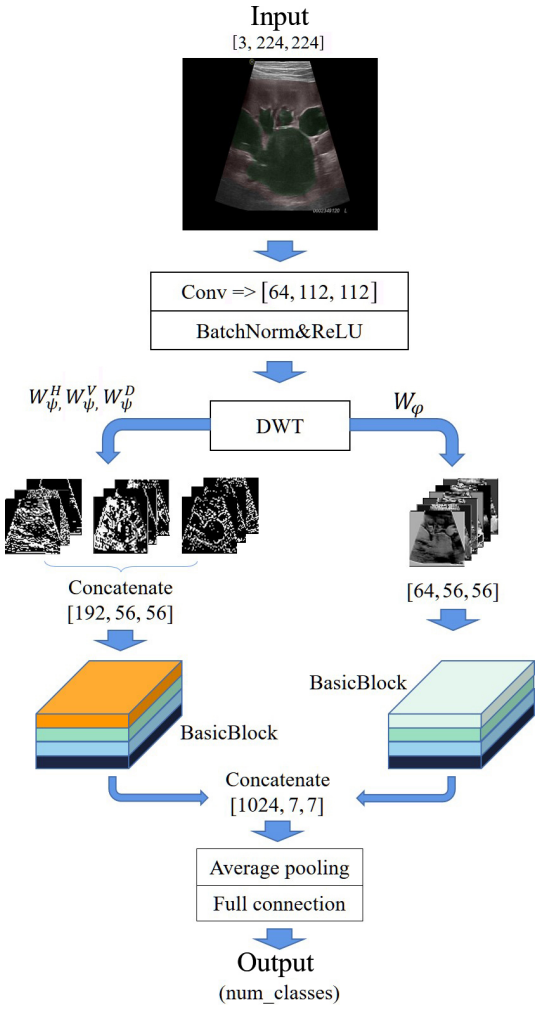


Fig. 4 Structure of the classifier. Through DWT, the image is generated into one approximate component and three detail components, and then the output results are concatenated through a neural network.

two BasicBlocks[28], which are only different from the dimensions of its first layer. Then, the two components are concatenated, and the final classification results are outputted through the average pooling dimension reduction and full connection layer. The frameworks of the BasicBlock in Fig. 4 are demonstrated in Fig. 5. Its core idea is to add a residual term called the identity shortcut connection at the output end to directly skip one or more layers and then add it to the backbone. This method consequently suppresses the gradient disappearance phenomenon caused by the very deep network.

2.3 Deep learning blockchain framework

In the era of smart medicine, traditional medical institutions have started to get assistance with the DL-CAD system. In this background, sensitive data sharing

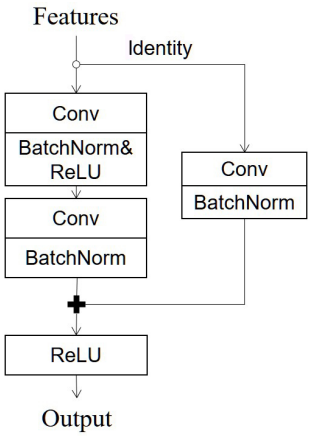


Fig. 5 Framework of the BasicBlock. The identity branch is the residual term and is added to the trunk.

and storing face the most challenges. Therefore, we use the Self-certifying File System (SFS)[33] to solve the problem of sensitive data transfer.

As a classical protocol of blockchain and a very successful application in the SFS, the IPFS[34] provides a secure, economical, tamper-resistant, and decentralized storage solution. Actually, there are eight layers of sub-protocol stacks in the IPFS framework, which ensure its effectiveness and security. From the bottom up, the layers are identity, network, routing, exchange, object, file, naming, and application. Each protocol stack has its own jobs and matches one another, as shown in Fig. 6.

We have also added an encryption algorithm to avoid the leakage of sensitive information in the external structure of the file through the RSA encryption algorithm. First, this algorithm generates a public key and private key, which are used to encrypt information to obtain ciphertext. Then, hexadecimal encoding will replace sensitive information in the file. The decryption

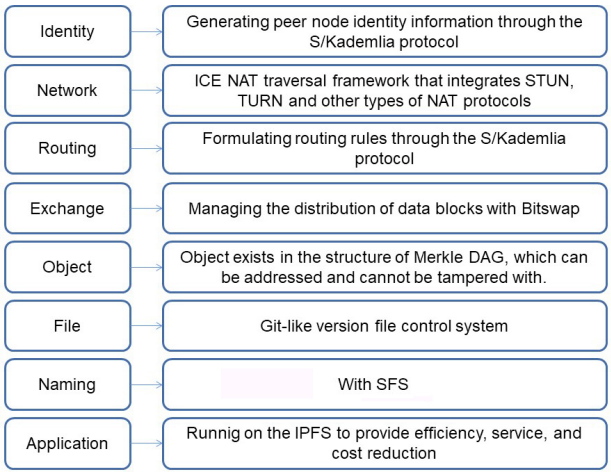


Fig. 6 Overall structure of the IPFS.

process decodes and decrypts the password with a private key to obtain plaintext and cover sensitive information in the original file.

When a file is uploaded to the IPFS, it is divided into several blocks and stored on different network nodes. Each block gets a unique ID through the hash operation to facilitate network nodes to identify and de-duplicate. This specific ID is also required when people want to access a file on the IPFS.

In our blockchain framework, as shown in Fig. 7, the three ends, namely, hospitals, patients, and our proposed DL-CAD model, together constitute a private IPFS. For example, in Step 3, hospital H_i uploads a pathological image dataset to the IPFS, gets a unique CID_H_i and sends it to the DL-CAD model. In Step 4, the DL-CAD model uses CID_H_i to download the desensitized dataset and returns the diagnosis results to the hospital through the IPFS and a new unique CID_M_i. This framework is also helpful for individuals, such as when uploading a single pathological image and safely obtaining a diagnosis suggestion if they have an IoMT, e.g., a portable or ultraportable ultrasound scanner[35, 36], or checking their relevant medical history if they had ever changed a hospital. A detailed demonstration of this framework and examples of how it works can be found in the supplementary video. Notice that our proposed blockchain system can be extended to contain more sites as long as this site has its own node.

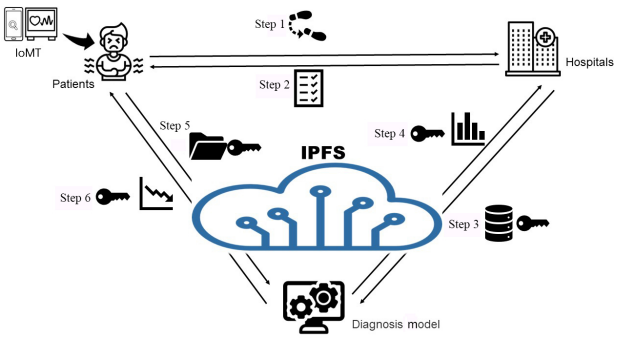


Fig. 7 Sketch map of the deep learning blockchain framework. Step 1: Patients visit hospitals; Step 2: Patients get results report after the diagnosis; Step 3: Hospital *i* on-chains the collected image dataset to the IPFS, gets a unique CID_H_i ($i \geqslant 1$), and gives it to the deep learning based diagnosis model; Step 4: The deep learning model uses CID_H_i to off-chain the desensitized dataset, trains them locally, then on-chains the diagnosis results, gets a CID_M_i, and gives this CID_M_i to hospital *i*. Afterward, the original hospital can use this ID to off-chain its results back. Steps 5 and 6: For an individual patient, the procedure is similar to that for an individual hospital.

3 Experiment

3.1 Dataset

All the experiments were conducted on the outpatient dataset of Beijing Children's Hospital, Capital Medical University, from September 2019 to February 2021. This dataset includes 229 patients, containing 3289 images taken by an experienced sonographer in standard positions (transverse and coronal sections of the kidney). Among them, 1850 images of 17 patients were annotated into the hydronephrosis area, kidney area, and background area by professionals using LabelMe. All the 3289 images were labeled into five grades, SFU_0–4, according to the Classification System of Congenital Hydronephrosis designed by the Society of Fetal Urology[37, 38]. All the experiments were performed using Intel CPU and NVIDIA GPU with 8 GB memory.

In clinical practice, SFU_0 is healthy, SFU_1&2 is mild and does not need surgery, SFU_3 is medium and needs further observation and diagnosis, and SFU_4 is severe and needs surgery. To better fit the clinical requirement, we further divide the 3289 images into three grades, i.e., SFU_0–2, SFU_3, and SFU_4. The stages of the experiments are demonstrated separately below.

3.2 Segmentation process

1850 annotated images were divided into the training set and test set with a ratio of 9:1. Batch training was adopted, and the batch size was set to 8. The training generation was 200. Adam optimizer was added to the training process. The initial learning rate was set to 0.005, and the learning rate was updated by the scheme with a ratio of 0.9 and a frequency of 1 generation. The combination of the Dice Loss and Cross Entropy Loss[39] was used as the loss function. The true value is defined as Y_{True}, the predicted value is Y_{Predict}, and the combined loss value is L, whose calculation method is

$$L = 1 - \sum Y_{\text{True}} \log Y_{\text{Predict}} - \frac{2\sum Y_{\text{True}} Y_{\text{Predict}}}{\sum Y_{\text{True}}^2 + Y_{\text{Predict}}^2} \quad (7)$$

Taking the Mean Intersection over Union (MIoU) and Mean Pixel Accuracy (MPA)[40] as the evaluation standards of the model results, the actual inputs of the whole training network are images reduced and filled to a size of 474 pixel × 473 pixel, and the output segmentation image size is 1024 pixel × 768 pixel.

3.3 Classification process

After using the 1850 annotated images to obtain a segmentation network with a good performance, all the remaining 1439 un-annotated images were sent through the segmentation network again. Then, we merged the segmentation result Seg with their original image Org of all the 3289 images to get the merged result Meg. The merging algorithm is as follows:

$$\text{Meg} = \text{Org} \times (1 - \alpha) + \text{Seg} \times \alpha \quad (8)$$

In the actual experiment, we set α to 0.1. We cropped the irrelevant peripheral area of the merged dataset Meg, keeping the organ regions as much as possible, and Meg was derived into a size of 810 pixel × 608 pixel. We divided the results into the training set and test set with a ratio of 9 : 1. A comparison of the results before and after the 2D-DWT processing and comparative experiments between our model and classic benchmark models were all conducted to prove the superiority of our classification network in this stage. Moreover, accuracy is taken as the evaluation standard. For the classification model f_{cls} and test set D_{test} with size N_{test}, the accuracy is as follows:

$$\text{Acc}(f_{cls}; D_{test}) = \frac{1}{N_{test}} \sum (f\ (x_i) = \text{label}_i) \quad (9)$$

where label_i is the label of the i-th sample.

To evaluate the model accurately, we weighed the impact of missing diagnosis and misdiagnosis, and set the threshold as a diagnostic reference value. False (F) indicates the misdiagnosis and True (T) indicates the correct diagnosis, as shown in Table 1.

We also analyzed the Receiver Operating Characteristic (ROC) curve of the sensitivity and specificity, including the True Positive Rate (TPR) and False Positive Rate (FPR), which are calculated below. Moreover, we added the Precision-Recall (PR) curve to show the quality of the model in an intuitive way. Their formulas are presented as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},\ \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$
$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}},\ \text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad (10)$$

In addition, we added the macro-average (Macro_F1) and micro-average (Micro_F1) evaluation indexes to the ROC curve,

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

Table 1 Confusion matrix of evaluation indices.

Prediction	Reality	
	Positive (P)	Negative (N)
Positive (P)	True Positive (TP)	False Negative (FN)
Negative (N)	False Positive (FP)	True Negative (TN)

$$\text{Macro\_F1} = \frac{2 \times \left(\frac{1}{n}\sum \text{Precision}_i\right) \times \left(\frac{1}{n}\sum \text{Recall}_i\right)}{\left(\frac{1}{n}\sum \text{Precision}_i\right) + \left(\frac{1}{n}\sum \text{Recall}_i\right)} \tag{12}$$

$$\text{Micro\_F1} = \frac{2 \times \frac{\sum \text{TP}_i}{\sum \text{TP}_i + \sum \text{FP}_i} \times \frac{\sum \text{TP}_i}{\sum \text{TP}_i + \sum \text{FN}_i}}{\frac{\sum \text{TP}_i}{\sum \text{TP}_i + \sum \text{FP}_i} + \frac{\sum \text{TP}_i}{\sum \text{TP}_i + \sum \text{FN}_i}} \tag{13}$$

When comparing our model with other classic models, we set the training epoch to 50, and all the parameters were set the same except the initial learning rate. Batch training with a batch size of 32 was adopted, training generation was set to 50, and Adam optimizer was added to update the learning rate with a 0.9 ratio and 1 generation frequency. To solve the imbalanced dataset problem, a ratio of $0.3:0.6:0.1$ was added to the cross-entropy loss as the weight, which was also used as the loss function of the training process. The input of the whole training network was 224 pixel $\times$ 224 pixel size images.

For the wavelet transform in our target network, the Haar wavelet transform and standard decomposition method were adopted. First, a 1D wavelet was used to transform the pixel values of each row of the image, and then each column of the image, and the approximate component and detail component were generated, in which the approximate component well replaced the output of the pooling layer. To verify the effect of DWT, we conducted comparative experiments, where we used the approximate component of the wavelet transform to replace the output of the pooling layer on other networks and train them. The networks we adopted for this function are AlexNet, GoogLeNet, and ResNet. The training epoch was 30, and other parameters, such as learning rate and input size, were the same as defined in the previous experiment.

3.4 Simulation of the blockchain system

We used VMware to generate two virtual machines of the Windows system, simulate medical institutions, and patients, and build the IPFS and related environment. In our simulation, the name of a file is its private information. The diagnostic results of the image set generated by the CAD model include the classification results. A validation operation was also added. That is, medical institutions can choose to upload marked datasets to test the performance of the CAD model, including the accuracy, precision, and recall of each image class.

4 Result

4.1 Segmentation results

For the segmentation task, we compared several sets of segmentation networks that also considered the pyramid pooling method. Then, we compared it with U-Net, an authoritative network in the field of medical image segmentation. Params means the number of parameters, and MACS means the number of multiplication and addition calculations per second. The results in Table 2 prove the superiority of our A-PSPNet.

Here, we will once again simply introduce the critical processes before classification. After the images passed through the semantic segmentation, the output was combined with the original image and then clipped as the input of our classifier. The specific results and processes are shown in Fig. 8.

4.2 Classification results

In the classifier, we added the wavelet transform without changing the feature dimension and obtained the

Table 2 Comparative experiment results of the semantic segmentation.

Model	MIoU	MPA	Params ($\times 10^6$)	MACS
PSPNet	86.61	92.01	46.71	51.68
Deeplabv3	85.47	91.58	59.23	77.50
DANet	80.54	83.14	214.41	203.38
U-Net	85.12	91.43	26.36	42.77
A-PSPNet (ours)	**87.93**	**93.52**	47.23	51.69

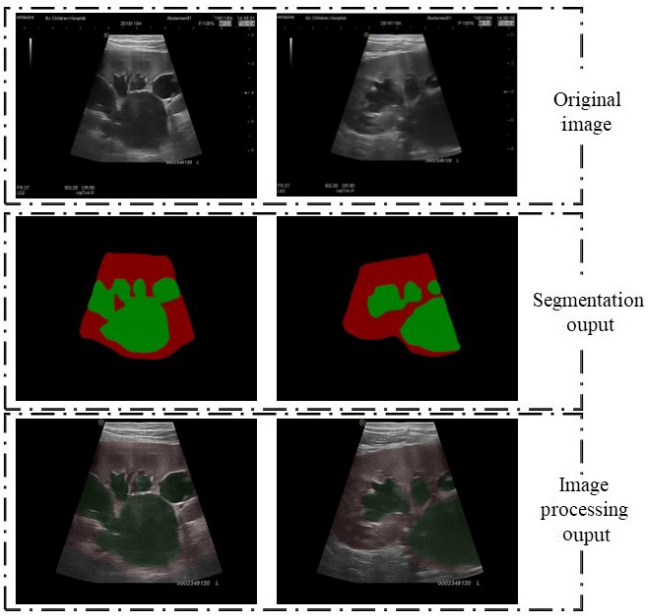
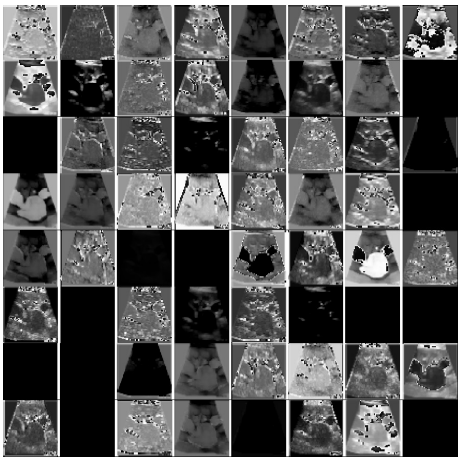


Fig. 8 Two examples of the specific results of the segmentation processing.

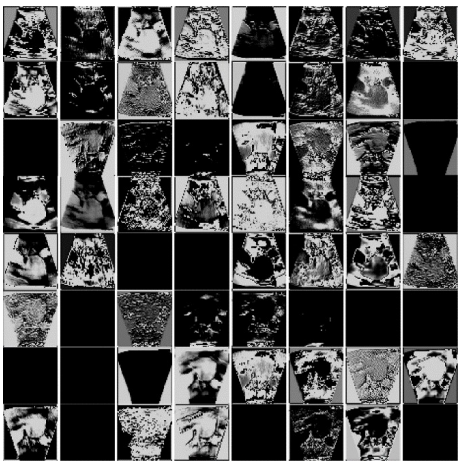
approximate component, which is consistent with the effect of double downsampling. The detail component with high-frequency information hence increases the available feature information of the whole network from the initial size of 64 pixel × 56 pixel × 56 pixel to 256 pixel × 56 pixel × 56 pixel. The feature maps of the approximate components and those after the original pooling process are shown in Fig. 9.

After changing to DWT, the path of the original pooling layer was replaced by the approximate components, and the path of the detailed component was also added. We took 64 of the 192 feature maps to facilitate the display, as shown in Fig. 10.

Comparing Fig. 9a with Fig. 9b, it can be found that the approximate component of the DWT is an effective substitute for pooling operation, and the detail components of Fig. 10 were added in the aspect of feature engineering. In Fig. 10, the new feature map contains dense high-frequency noise information and contour information.



(a) After pooling



(b) After DWT

Fig. 9 Feature maps of approximate components and those after the original pooling process. (a) Feature maps of a randomly selected ultrasonic image in a dataset after the pooling layer of the original residual network, with a dimension of 64 pixel × 56 pixel × 56 pixel. (b) Feature maps of approximate components after DWT with a dimension of 64 pixel × 56 pixel × 56 pixel.

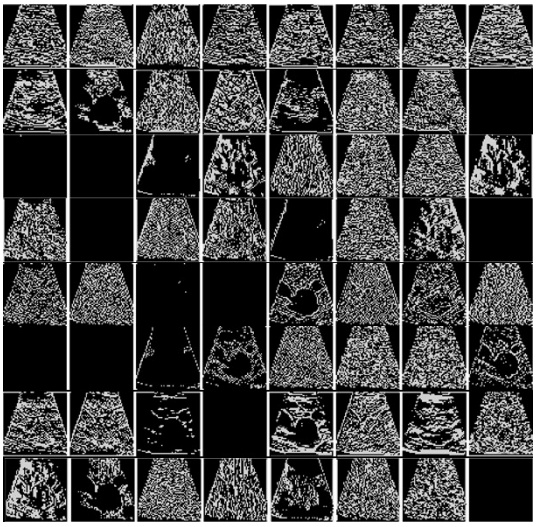


Fig. 10 Feature maps randomly selected from the horizontal, vertical, and diagonal detail components, whose total number is 192.

We conducted a comparative experiment on the network classification performance, and the specific results are shown in Table 3. Table 3 presents the optimal accuracy model and training epoch in the training process of different networks. We observe that ResNet has a better accuracy among the four benchmark networks, and our proposed model based on ResNet outperforms all the other methods, which proves the correctness of choosing ResNet.

We also attempted to replace the pooling layer with DWT on other classical benchmark networks. Because excessive pooling layers will greatly increase the complexity of the network, we only selected and tested several networks with few pooling layers at important positions, and thus VGG was eliminated. The ACC changes during training were collected, and all curves were smoothened. As shown in Fig. 11, replacing

Table 3 Comparative experiment results on the classifier.

Model	ACC (%)	Epoch	Params ($\times 10^6$)
AlexNet	84.45	35	61.10
VGG	86.36	27	68.24
GoogleNet	84.75	38	10.31
ResNet	89.94	38	21.29
ResNet-W (ours)	**91.77**	49	42.65

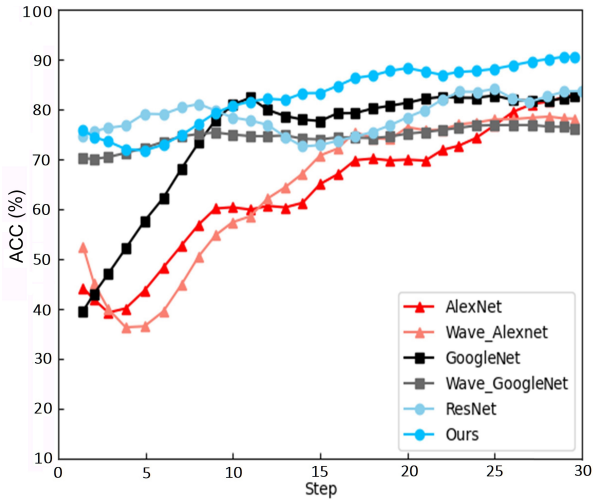


Fig. 11 Performance after replacing the pooling layer with the DWT on classical benchmark networks.

the pooling layer and increasing the number of features enhance the performance of all the networks. ResNet is still the best, which confirms the feasibility of increasing the amount of feature information and the superiority of our proposed method.

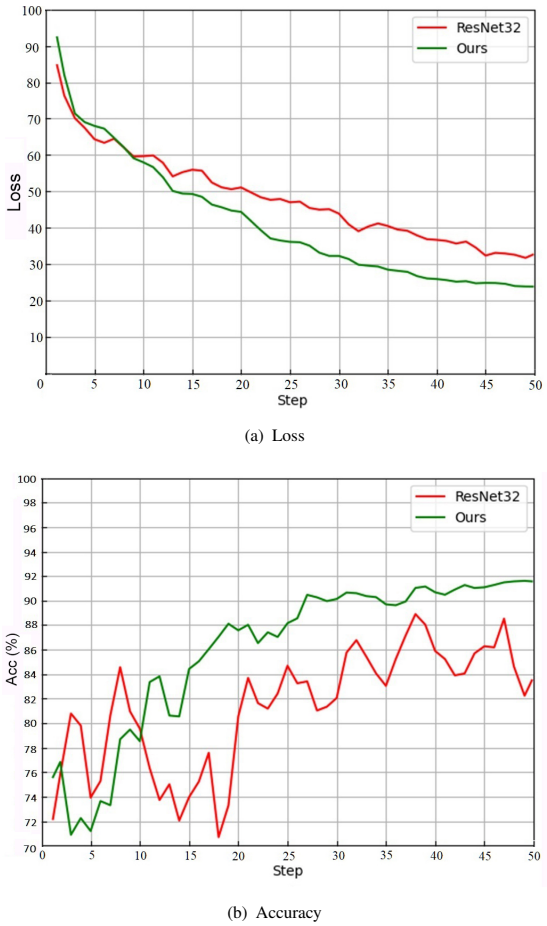
Our improved method is based on a residual network. As shown in Fig. 12, compared with the original network, our model has a more stable training process and better overall performance than the original network in terms of the vibration amplitude and frequency of the curve.

In the first 30 generations, the accuracy and loss values of the original network training model on the test set tend to simultaneously decrease, thus increasing the possibility of network overfitting. This condition further confirms the superiority of our network structure. Based on the above results, a reasonable replacement of the pooling layer by the DWT will increase the number of features, make the training process stable, and improve the performance of the model.

As shown in the PR curve in Fig. 13, the average precision of our final model for different recall points is 0.96, which shows that the diagnostic ability of the model is quite excellent. As shown in the ROC curve in Fig. 14, the worst diagnostic result is that of SFU_3, yet it still reaches 0.90. On the whole, micro-average ROC = 0.98 and macro-average ROC = 0.95, which again proves the superiority of our classifier.

4.3 DL-CAD simulation results

We also simulated the situation when medical institutions tested the DL-CAD model performance on a blockchain system. Three classes of cases, i.e., 600 original



(a) Loss

(b) Accuracy

Fig. 12 Performance of the benchmark models versus our proposed method.

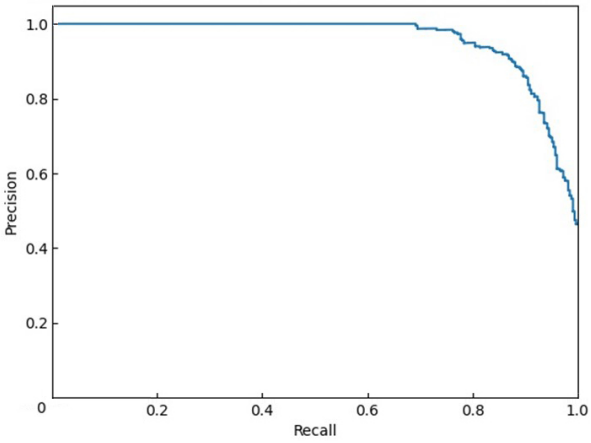


Fig. 13 Micro_F1 over all SFU classes: average precision score = 0.96.

ultrasound images for saving the training cost were adopted and securely transmitted through the IPFS. The generated performance results, including the precision, recall, and F1 of each classification, are shown in Table 4. Besides the general accuracy is 91.11%, which

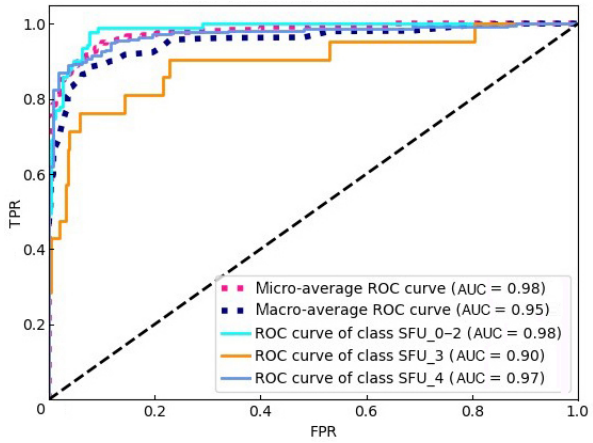


Fig. 14 Classifier performance indicators: ROC. The Area Under the Curve (AUC) is defined by the area enclosed by the ROC curve and coordinate axis. The more convex the ROC curve is and the closer it is to the upper left corner, the greater its diagnostic value is, which evaluates the diagnostic accuracy.

Table 4 Results of the test experiments on the DL-CAD model.

Class	Indicator		
	Precision	Recall	F1
SFU_0–2	0.89	0.99	0.94
SFU_3	0.89	0.95	0.92
SFU_4	0.97	0.79	0.87
Averaged	0.91	0.91	0.91

is almost the same as when training the whole dataset on the local diagnosis model alone without transmission on the IPFS (91.77%). Therefore, the data transmission process through the IPFS will not affect the final diagnosis results, which further confirms the reliability and effectiveness of the DL-CAD model.

5 Discussion

In this paper, we propose a framework that integrates a deep learning diagnosis model and blockchain technology to recognize and grade UPJO on the basis of ultrasound images. Our diagnosis model outperformed benchmark networks in terms of accuracy, reaching 91.77%. This model indicates that our model can highly assist doctors and save their labors. The integration of the blockchain protocol enables secure and desensitized data sharing for institutions and patients, and allows them to access the history information whenever and wherever.

However, limitations still exist in our study. For example, the proposed framework can contain more than one site at each end (patient end and hospital end). However, the real-world dataset that we can use to verify is only one because there is no useable public UPJO ultrasound image dataset, and the data sharing between institutions are restricted. This, however, demonstrates the urgent demands of our work from another point of view. Moreover, although we mentioned IoMT as portable and ultraportable ultrasound scanners at the patients' side, we did not include any specific IoMT in our framework, and the development of ultraportable ultrasound scanners for the kidney is also not sophisticated enough.

In the future, we could link actual IoMT to our system. Because patients of UPJO are always young children, they can be terrified to go to hospitals and are less cooperative with doctors. The changes in their bodies and organs are also constant and unpredictable. Therefore, household ultraportable ultrasound scanners can greatly relieve parents' burden and struggle of visiting doctors. The fast and automatic diagnosis allows parents to monitor children's situations timely, and the blockchain technology facilitates the storage and accessibility of personal medical history at any time. Furthermore, the proposed method provides technical support for further cooperation among multiple parties and perhaps federated learning. Lastly, this system can be generalized for the safe diagnosis of other diseases.

6 Conclusion

We developed a deep learning-based model to diagnose UPJO using ultrasound images. Then, we integrated this model with a constructed blockchain technology IPFS. Finally, we established a deep learning blockchain framework to ensure secure data transfers and fast UPJO diagnosis. Experiments were conducted, and the results verified the superiority and reliability of our system. In the future, more institutions can be included to extend this cooperation. IoMT-like portable and ultraportable ultrasound scanners can also be linked to facilitate the convenient household diagnosis to relieve patients' burdens. Finally, our framework can provide technical support for possible federated learning and generalized diagnosis of other diseases.

Acknowledgment

This study was supported by the National Key R&D Program of China (No. 2020YFB2104402).

References

[1] M. J. Siegel, *Pediatric Sonography*, 4th ed. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2011.

[2] J. M. Smith, D. M. Stablein, R. Munoz, D. Hebert, and R. A. McDonald, Contributions of the transplant registry: The 2006 annual report of the North American Pediatric Renal Trials and Collaborative Studies (NAPRTCS), *Pediatr. Transplant.*, vol. 11, no. 4, pp. 366–373, 2007.

[3] B. A. Warady and V. Chadha, Chronic kidney disease in children: The global perspective, *Pediatr. Nephrol.*, vol. 22, no. 12, pp. 1999–2009, 2007.

[4] H. T. Nguyen, C. D. A. Herndon, C. Cooper, J. Gatti, A. Kirsch, P. Kokorowski, R. Lee, M. Perez-Brayfield, P. Metcalfe, and E. Yerkes, et al., The society for fetal urology consensus statement on the evaluation and management of antenatal hydronephrosis, *J. Pediatr. Urol.*, vol. 6, no. 3, pp. 212–231, 2010.

[5] H. Hashim and C. R. J. Woodhouse, Ureteropelvic junction obstruction, *Eur. Urol. Suppl.*, vol. 11, no. 2, pp. 25–32, 2012.

[6] A. K. Ucar and S. Kurugoglu, Urinary ultrasound and other imaging for ureteropelvic junction type hydronephrosis (UPJHN), *Front. Pediatr.*, vol. 8, p. 546, 2020.

[7] American Institute of Ultrasound in Medicine, AIUM practice guideline for the performance of an ultrasound examination in the practice of urology, *J. Ultrasound Med.*, vol. 31, no. 1, pp. 133–144, 2012.

[8] American Institute of Ultrasound in Medicine, AIUM practice guideline for the performance of an ultrasound examination in the practice of urology, *J. Ultrasound Med.*, vol. 31, no. 1, pp. 133–144, 2012.

[9] L. C. Smail, K. Dhindsa, L. H. Braga, S. Becker, and R. R. Sonnadara, Using deep learning algorithms to grade hydronephrosis severity: Toward a clinical adjunct, *Front. Pediatr.*, vol. 8, p. 1, 2020.

[10] S. Turco, P. Frinking, R. Wildeboer, M. Arditi, H. Wijkstra, J. R. Lindner, and M. Mischi, Contrast-enhanced ultrasound quantification: From kinetic modeling to machine learning, *Ultrasound Med. Biol.*, vol. 46, no. 3, pp. 518–543, 2020.

[11] H. Shokoohi, M. A. Lesaux, Y. H. Roohani, A. Liteplo, C. Huang, and M. Blaivas, Enhanced point-of-care ultrasound applications by integrating automated feature-learning systems using deep learning, *J. Ultrasound Med.*, vol. 38, no. 7, pp. 1887–1897, 2019.

[12] K. Dhindsa, L. C. Smail, M. McGrath, L. H. Braga, S. Becker, and R. R. Sonnadara, Grading prenatal hydronephrosis from ultrasound imaging using deep convolutional neural networks, in *Proc. 15th Conf. on Computer and Robot Vision* (*CRV*), Toronto, Canada, 2018, pp. 80–87.

[13] E. S. Blum, A. R. Porras, E. Biggs, P. R. Tabrizi, R. D. Sussman, B. M. Sprague, E. Shalaby-Rana, M. Majd, H. G. Pohl, and M. G. Linguraru, Early detection of ureteropelvic junction obstruction using signal analysis and machine learning: A dynamic solution to a dynamic problem, *J. Urol.*, vol. 199, no. 3, pp. 847–852, 2018.

[14] J. X. He, S. L. Baxter, J. M. Xu, X. T. Zhou, and K. Zhang, The practical implementation of artificial intelligence technologies in medicine, *Nat. Med.*, vol. 25, no. 1, pp. 30–36, 2019.

[15] M. Chen, X. B. Shi, Y. Zhang, D. Wu, and M. Guizani, Deep feature learning for medical image analysis with convolutional autoencoder neural network, *IEEE Trans. Big Data*, vol. 7, no. 4, pp. 750–758, 2021.

[16] R. Hillestad, J. Bigelow, A. Bower, F. Girosi, R. Meili, R. Scoville, and R. Taylor, Can electronic medical record systems transform health care? Potential health benefits, savings, and costs, *Health Affairs*, vol. 24, no. 5, pp. 1103–1117, 2005.

[17] P. Vimalachandran, H. Wang, and Y. C. Zhang, Securing electronic medical record and electronic health record systems through an improved access control, in *Proc. 4th Int. Conf. on Health Information Science*, Melbourne, Australia, 2015, pp. 17–30.

[18] S. Biswas, K. Sharif, F. Li, I. Alam, and S. Mohanty, DAAC: Digital asset access control in a unified blockchain based E-health system, *IEEE Trans. Big Data*, doi: 10.1109/TBDATA.2020.3037914.

[19] C. J. McDonald, The barriers to electronic medical record systems and how to overcome them, *J. Am. Med. Inf. Assoc.*, vol. 4, no. 3, pp. 213–221, 1997.

[20] S. Selvaraj and S. Sundaravaradhan, Challenges and opportunities in IoT healthcare systems: A systematic review, *SN Appl. Sci.*, vol. 2, no. 1, p. 139, 2020.

[21] S. Rahmadika, M. Firdaus, S. Jang, and K. H. Rhee, Blockchain-enabled 5G edge networks and beyond: An intelligent cross-silo federated learning approach, *Secur. Commun. Networks*, vol. 2021, p. 5550153, 2021.

[22] Z. B. Zheng, S. A. Xie, H. N. Dai, X. P. Chen, and H. Wang, Blockchain challenges and opportunities: A survey, *Int. J. Web Grid Serv.*, vol. 14, no. 4, pp. 352–375, 2018.

[23] R. Kumar, A. A. Khan, J. Kumar, Zakria, N. A. Golilarz, S. M. Zhang, Y. Ting, C. Y. Zheng, and W. Y. Wang, Blockchain-federated-learning and deep learning models for COVID-19 detection using CT imaging, *IEEE Sens. J.*, vol. 21, no. 14, pp. 16301–16314, 2021.

[24] H. N. Dai, Z. B. Zheng, and Y. Zhang, Blockchain for internet of things: A survey, *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8076–8094, 2019.

[25] M. A. Ferrag and L. Shu, The performance evaluation of blockchain-based security and privacy systems for the internet of things: A tutorial, *IEEE Internet Things J.*, vol. 8, no. 24, pp. 17236–17260, 2021.

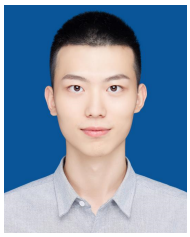
[26] K. Peng, M. J. Li, H. J. Huang, C. Wang, S. H. Wan, and K. K. R. Choo, Security challenges and opportunities for smart contracts in internet of things: A survey, *IEEE Internet Things J.*, vol. 8, no. 15, pp. 12004–12020, 2021.

[27] G. J. Joyia, R. M. Liaqat, A. Farooq, and S. Rehman, Internet of medical things (IoMT): Applications, benefits and future challenges in healthcare domain, *J. Commun.*, vol. 12, no. 4, pp. 240–247, 2017.

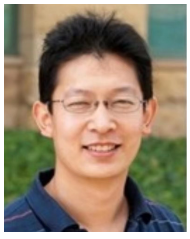
[28] K. D. Krishna, V. Akkala, R. Bharath, P. Rajalakshmi, and A. M. Mohammed, FPGA based preliminary CAD for

kidney on IoT enabled portable ultrasound imaging system, in *Proc. IEEE 16th Int. Conf. on e-Health Networking, Applications and Services* (*Healthcom*), Natal, Brazil, 2015, pp. 257–261.
[29] H. S. Zhao, J. P. Shi, X. J. Qi, X. G. Wang, and J. Y. Jia, Pyramid scene parsing network, in *Proc. 2017 IEEE Conf. Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017, pp. 6230–6239.
[30] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, CBAM: Convolutional block attention module, in *Proc. 15th European Conf. Computer Vision*, Munich, Germany, 2018, pp. 3–19.
[31] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, Deep residual learning for image recognition, in *Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 770–778.
[32] B. Vidakovic, *Discrete Wavelet Transformation*. Hoboken, NJ, USA: John Wiley & Sons, 2008.
[33] D. Mazières, Self-certifying file system, PhD dissertation, Massachusetts Institute of Technology, Cambridge, UK, 2000.
[34] J. Benet, IPFS-content addressed, versioned, p2p file system, arXiv preprint arXiv:1407.3561, 2014.
[35] J. A. Smith and O. C. Jensen, Portable ultrasound scanner, U.S. Patent US20110118562A1, 2011.
[36] R. Bharath, D. Chandrashekar, V. Akkala, D. Krishna, H. Ponduri, P. Rajalakshmi, and U. B. Desai, Portable ultrasound scanner for remote diagnosis, in *Proc. 17th Int. Conf. on E-Health Networking*, *Application & Services*, Boston, MA, USA, 2016, pp. 211–216.
[37] S. K. Fernbach, M. Maizels, and J. J. Conway, Ultrasound grading of hydronephrosis: Introduction to the system used by the Society for Fetal Urology, *Pediatr. Radiol.*, vol. 23, no. 6, pp. 478–480, 1993.
[38] A. Onen, An alternative grading system to refine the criteria for severity of hydronephrosis and optimal treatment guidelines in neonates with primary UPJ-type hydronephrosis, *J. Pediatr. Urol.*, vol. 3, no. 3, pp. 200–205, 2007.
[39] L. Xu and J. H. Xiang, ComboLoss for facial attractiveness analysis with squeeze-and-excitation networks, arXiv preprint arXiv:2010.10721, 2020.
[40] M. A. Rahman and Y. Wang, Optimizing intersection-over-union in deep neural networks for image segmentation, in *Proc. 12th Int. Symp. on Visual Computing*, Las Vegas, NV, USA, 2016, pp. 234–244.

Yu Guan received the BEng degree in information security from Beijing University of Technology, China in 2016. Currently, she is a PhD candidate at the Faculty of Information Technology, Beijing University of Technology, China. Her research interests are big data and data mining applied to the medical area.



Pengceng Wen received the BEng degree in electronic information from Chongqing Jiaotong University, China in 2020. He is now a master student at the Faculty of Information Technology, Beijing University of Technology, China. His main research interest is data mining applied to the medical area.



Jianqiang Li received the BEng degree in mechatronics from Beijing Institute of Technology, Beijing, China in 1996, and the MEng and PhD degrees in control science and engineering from Tsinghua University, Beijing, China in 2001 and 2004, respectively. He worked as a researcher at the Digital Enterprise Research Institute, National University of Ireland in 2004–2005. He worked at NEC Labs China as a researcher from 2005 to 2013, and at the Department of Computer Science, Stanford University China, as a visiting scholar in 2009–2010. He joined Beijing University of Technology, Beijing, China in 2013 as a Beijing Distinguished Professor. His research interests are Petri nets, enterprise information systems, business processes, data mining, information retrieval, semantic web, privacy protection, and big data. He has over 40 publications and 37 international patent applications (19 of them have been granted in China, the US, or Japan). He served as a PC member in multiple international conferences and organized the IEEE workshop on medical computing.

Jinli Zhang received the PhD degree in computer science from Beijing University of Technology, Beijing, China in 2020. She worked as a researcher at Drexel University, USA in 2019–2020. She joined Beijing University of Technology, Beijing, China in 2020 as a postdoctoral researcher. She has published 15 SCI journals and 4 invention patents. Her research interests are in artificial intelligence, machine learning, and data/text/web mining. She is a member of IEEE.

Xianghui Xie received the bachelor and master of medicine degrees from Capital Medical University, China in 1994 and 2002, respectively. He was assigned to Beijing Children's Hospital Affiliated to Capital Medical University in 2002, and has been engaged in clinical, teaching, scientific research, and hospital management of pediatric urology. He now works in the Capital Institute of Pediatrics Affiliated Children's Hospital. He has published a total of 35 articles in various academic core journals and edited many books related to pediatric urology. His research interests are pediatric urology, and correction of birth structural defects and deformities.