Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Texture branch network for chronic kidney disease screening based on ultrasound images*

Peng-yi HAO[1,5], Zhen-yu XU[1], Shu-yuan TIAN[2], Fu-li WU†‡[1,5], Wei CHEN[3,5], Jian WU[4,5], Xiao-nan LUO[6]

[1]*College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China*
[2]*Tongde Hospital of Zhejiang Province, Hangzhou 310012, China*
[3]*The First Affiliated Hospital, Zhejiang University, Hangzhou 310003, China*
[4]*College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China*
[5]*Real Doctor AI Research Center, Zhejiang University, Hangzhou 310027, China*
[6]*Institute of Artificial Intelligence, Guilin University of Electronic Technology, Guilin 541004, China*
†E-mail: fuliwu@zjut.edu.cn

Received Apr. 24, 2019; Revision accepted June 23, 2019; Crosschecked Aug. 23, 2019; Published online Oct. 12, 2019

Abstract: Chronic kidney disease (CKD) is a widespread renal disease throughout the world. Once it develops to the advanced stage, serious complications and high risk of death will follow. Hence, early screening is crucial for the treatment of CKD. Since ultrasonography has no side effects and enables radiologists to dynamically observe the morphology and pathological features of the kidney, it is commonly used for kidney examination. In this study, we propose a novel convolutional neural network (CNN) framework named the texture branch network to screen CKD based on ultrasound images. This introduces a texture branch into a typical CNN to extract and optimize texture features. The model can automatically generate texture features and deep features from input images, and use the fused information as the basis of classification. Furthermore, we train the base part of the network by means of transfer learning, and conduct experiments on a dataset with 226 ultrasound images. Experimental results demonstrate the effectiveness of the proposed approach, achieving an accuracy of 96.01% and a sensitivity of 99.44%.

Key words: Chronic kidney disease; Ultrasound; Texture branch network; Transfer learning
https://doi.org/10.1631/FITEE.1900210 **CLC number:** TP183; TP391.7

1 Introduction

Chronic kidney disease (CKD) is a condition in which kidney function is declining gradually over time. It is caused mainly by various factors such as diabetes and high blood pressure. It is defined as kidney pathologic lesions or diminished function. In particular, the glomerular filtration rate (GFR) is lower than 60 mL/(min·1.73 m^2) for three months or more, regardless of cause (Levey et al., 2005). Generally, CKD progresses slowly with no obvious symptoms in the early stage. Once it reaches the advanced stage, the risk of complications and premature death will increase greatly (El Nahas and Bello, 2005). Therefore, treatment for CKD focuses on slowing the progression of the kidney damage. Early screening is the most effective measure to help prevent or delay the progression of the disease in CKD patients (Ecder, 2013). Currently, CKD has become

‡ Corresponding author
* Project supported by the Zhejiang Provincial Natural Science Foundation of China (No. LY18F020034), the Zhejiang Provincial Medical Health Science and Technology Project, China (No. 2014KYB320), the National Natural Science Foundation of China (Nos. 61801428 and 61672543), the Zhejiang University Education Foundation, China (Nos. K18-511120-004 and K17-511120-017), and the Major Scientific Project of Zhejiang Lab, China (No. 2018DG0ZX01)
ORCID: Fu-li WU, http://orcid.org/0000-0002-1566-9343

one of the most serious global health problems. It is of great importance to explore an accurate, safe, and cost-effective approach for CKD diagnosis in the general population.

The typical screening techniques are computerized tomography (CT), magnetic resonance imaging (MRI), and ultrasonography. However, due to radiation, contrast agents, and contraindications during the examination of CT and MRI, patients with end-stage CKD are at risk of aggravated kidney impairment. In practice, ultrasound imaging has become the preferred option for renal examination because of its low cost and convenience in operation. In a previous study, radiologists used contrast-enhanced ultrasound to detect the malignancy of the kidney (Chang et al., 2017); the results showed a high sensitivity (96%) but a low specificity (50%). Apparently, for a radiologist, visually analyzing ultrasound images to identify kidney lesions is a time-consuming and challenging task.

With the aim of helping doctors make decisions on the severity of CKD, many researchers have proposed diverse diagnostic support methods (Batra et al., 2016). Machine learning methods have been used to predict the CKD stage, with clinical medical data including attributes like age, weight, blood pressure, and glucose. These advanced classification algorithms involve artificial neural network (ANN) (Jeewantha et al., 2017), naive Bayes (Kunwar et al., 2016), K-nearest neighbor (KNN) (Charleonnan et al., 2016), and support vector machine (SVM) (Ahmad et al., 2017). Although these methods achieved high classification accuracy using non-ultrasonography data, they required the collection of a wide variety of clinical data. Other studies which centered on computer-aided diagnosis (CAD) have been demonstrated to assist doctors in interpreting information from medical images. Several approaches based on ultrasonic image analysis to identify CKD have been presented, for example, applying image processing techniques to calculate the ratio of kidney fibrosis and distinguish the stage of CKD (Ho et al., 2012; Pujari and Hajare, 2014). Features designed from the local binary pattern (LBP) and Nakagami distribution, presented by Hsieh et al. (2014), were classified by SVM for CKD screening. Acharya et al. (2019) extracted higher-order features and elongated quinary patterns from B-mode images, and classified features by SVM to detect CKD. Iqbal et al. (2017) reported texture features from the region of the kidney and explored the differences between normal and diseased regions. So far, the studies on CKD using ultrasound images are limited. Remarkably, texture features of ultrasound images are critical information, and they are often used for the segmentation of kidney (Subramanya et al., 2015) and recognition of renal cysts (Subramanya et al., 2015; Sharma and Virmani, 2016).

In recent years, deep learning techniques have been increasingly employed in CAD, particularly convolutional neural networks (CNNs), since they provide a framework for discriminating feature extraction and classification (Shin et al., 2016). Chen et al. (2015) used a recurrent neural network to detect the fetal standard plane based on ultrasound videos. Up to now, there is little research carried out on deep learning in CKD screening using ultrasonic images, but similar problems have been addressed. For instance, to assess congenital abnormalities of the kidney and urinary tract, Zheng et al. (2019) used pre-trained weights from ImageNet and applied features to an SVM classifier. Dhindsa et al. (2018) suggested a seven-layer CNN to evaluate prenatal hydronephrosis from kidney ultrasound images. Despite great potentials, classical deep learning models do not perform outstandingly on ultrasound images as they do on natural images. The main deficiency is attributed to inadequate medical image data that easily lead to the overfitting of the model. This common issue is related to data acquisition and high-cost professional annotation. To deal with this limitation, we suggest using a pre-trained CNN as an initialization for training (Shen et al., 2015), to transfer the representation abilities from a large-scale dataset. According to the common research, we also suggest combining domain texture features to provide more information for decision making.

In this study, we present a novel CNN framework with a texture branch, called the texture branch network (TBN), for distinguishing CKD based on kidney ultrasound images. The model consists of a base network and a texture branch. The texture branch is organized as a residual structure (He et al., 2016), generating optimized texture features from the input image. At the end of the network, different types of features are fused for classification.

By introducing the TBN model into the CKD screening task, there are two contributions:

1. To the best of our knowledge, this is the first end-to-end approach using deep learning to figure out CKD diagnostic classification based on ultrasound images. Our innovative model, which integrates deep features and domain texture features as multi-level description, achieves an excellent classification accuracy.

2. Employing transfer learning manner to train partial model parameters shows a good performance under the unbalanced, limited-sample ultrasonic image dataset.

2 Method

The proposed TBN model consists of a base network, a texture branch, and dense layers for classification. The base network makes full use of the expressive power of convolution layers to describe the deep features of images. The designed texture features are extracted as input and evolved by a texture branch. The fusion features of two parts are classified by dense layers. The architecture of our model is shown in Fig. 1. The details of this model will be explained in the following subsections.

2.1 Base network

The base network is constructed as the backbone of our CNN framework. To learn effective features from a limited number of samples, an appropriate network structure should be considered. Although complex deep networks have a strong expressive capacity, they are also likely to fall into the dilemma of overfitting. ResNet-34 (He et al., 2016) was chosen for this task in practice due to the moderate model complexity. There are a total of 34 learnable layers contained in the ResNet-34 network. The input of the network is region of interest (ROI) of the kidney, which is resized as a uniform size of 224×224. Each ROI is first processed by a convolutional layer with kernel size 7×7, and each kernel produces a two-dimensional (2D) feature map. The following is a 3×3 max-pooling layer with a stride of two, which reduces the size of feature maps, then four residual blocks, an average pooling operation, and a fully connected layer. In each residual block, two convolutional layers are laid in sequence with kernel size 3×3. The architecture is shown in Table 1. As the input image goes deeper through residual blocks, the size of feature maps decreases gradually while the number of feature maps increases. To adapt the network to our binary classification task of CKD screening, the original fully connected layer of 1000 neurons is replaced by two added fully connected layers with

Table 1 Architectures of residual blocks in ResNet-34*

Layer	Replication	Number of channels	Output size
Residual block 1	3	64, 64	56×56
Residual block 2	4	128, 128	28×28
Residual block 3	6	256, 256	14×14
Residual block 4	3	512, 512	7×7

* Each residual block contains two convolutional layers (3×3, 3×3)

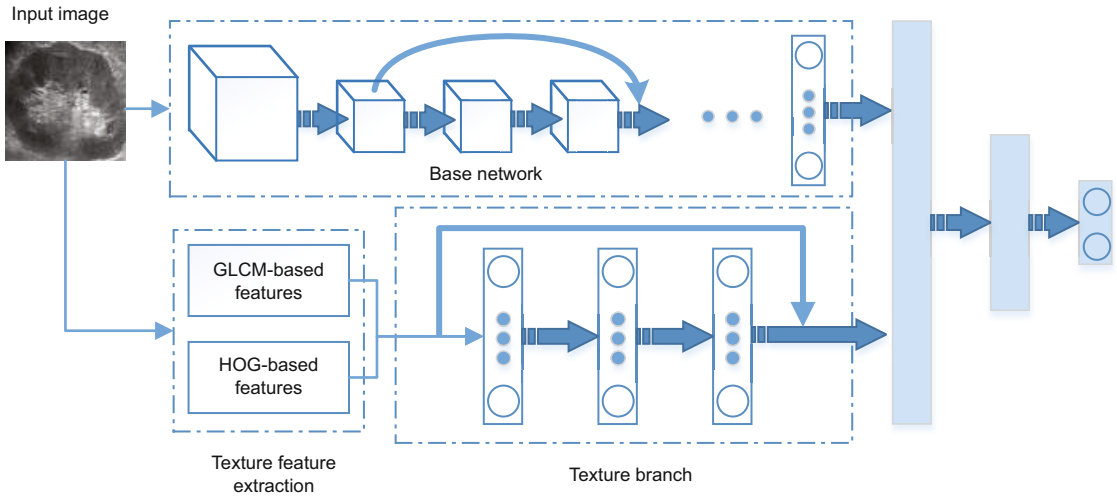


Fig. 1 Architecture of the proposed texture branch network (TBN) model

256 and two neurons, respectively.

To obtain a potent representation ability of CNN for images, even on small-sample medical datasets, the weight parameters of the base network were initialized by a transferred ResNet-34, which was trained on the ImageNet dataset (Oquab et al., 2014) in this study. This transfer learning manner allows the model to be closer to the optimal parameters in the initial training, accelerates the convergence speed of the network, and reduces the risk of overfitting.

2.2 Feature extraction and texture branch

Parenchymal fibrosis is a manifestation of CKD. It increases with disease progression and reveals abnormal texture patterns in kidney ultrasonography. Texture features will be extracted to afford richer information for accurate judgment.

2.2.1 GLCM-based texture feature extraction

The gray-level co-occurrence matrix (GLCM) (Haralick et al., 1973) considers the spatial relationship of pixel gray levels in an image. The calculation of GLCM involves distance and angle to reflect the comprehensive information of the image in the direction and interval. Totally 16 GLCMs are computed under the distance counted 1, 2, 3, 4 and orientations of 0°, 45°, 90°, 135°. It should be noted that for a given matrix $\boldsymbol{P}$, the values of $P_{0,i}$ and $P_{i,0}$ ($i \in [0, N-1]$, N is the maximum gray level of input image and N equals 256 in practice) should be set to 0 to eliminate the interference of non-echo areas. Each GLCM is generally used for a series of second-order texture calculations. We produced the features according to five common measures, contrast, dissimilarity, homogeneity, energy, and correlation, resulting in 80-dimensional features from a total of 16 GLCMs for an analyzed image.

For a given GLCM $\boldsymbol{P}$ with probability $P_{i,j}$, the five different texture measure calculations are defined as follows:

$$\text{Contrast} = \sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2, \tag{1}$$

$$\text{Dissimilarity} = \sum_{i,j=0}^{N-1} P_{i,j}|i-j|, \tag{2}$$

$$\text{Homogeneity} = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}, \tag{3}$$

$$\text{Energy} = \sum_{i,j=0}^{N-1} \sqrt{P_{i,j}^2}, \tag{4}$$

$$\text{Correlation} = \sum_{i,j=0}^{N-1} P_{i,j}\left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{\sigma_i^2\sigma_j^2}}\right], \tag{5}$$

where the mean μ and variance σ^2 of $\boldsymbol{P}$ are calculated as

$$\begin{cases} \mu_i = \sum_{i,j=0}^{N-1} iP_{i,j}, \\ \mu_j = \sum_{i,j=0}^{N-1} jP_{i,j}, \end{cases} \tag{6}$$

$$\begin{cases} \sigma_i^2 = \sum_{i,j=0}^{N-1} P_{i,j}(i-\mu_i)^2, \\ \sigma_j^2 = \sum_{i,j=0}^{N-1} P_{i,j}(j-\mu_j)^2. \end{cases} \tag{7}$$

These features are standardized by the Z-score manner according to different groups.

2.2.2 Histogram of oriented gradient based texture feature extraction

Another widely used texture descriptor is the histogram of oriented gradient (HOG). This feature is based on the statistical information of the gradient in an image, which reflects the edge of a local object (Dalal and Triggs, 2005). To extract the HOG feature, an image is divided into numerous small squared regions called cells. In each cell, an HOG is calculated, and this denotes the texture description of the cell. These cells will form a larger connected region (blocks). The HOG feature of a block is obtained after normalization by concatenating feature vectors of all cells belonging to it. Consequently, the HOG features of an image implemented by combining the gradient statistics of all blocks, and described as magnitude M and direction θ for pixel (x, y), are calculated as

$$M(x,y) = \sqrt{G_x^2(x,y) + G_y^2(x,y)} \tag{8}$$

and

$$\theta(x,y) = \arctan\frac{G_x(x,y)}{G_y(x,y)}, \tag{9}$$

where $G_x(x,y)$ and $G_y(x,y)$ represent the horizontal and vertical gradient values of the image pixel at (x,y), respectively. They are defined as

$$G_x(x,y) = F(x+1,y) - F(x-1,y) \tag{10}$$

and

$$G_y(x,y) = F(x,y+1) - F(x,y-1), \tag{11}$$

where $F(x,y)$ is the pixel value at (x,y) in the image.

2.2.3 Texture branch

After extracting texture features from an image, a k-dimensional feature vector $\boldsymbol{v}$ can be obtained, and k is set to 80, 324, or 404 for the GLCM feature, HOG feature, or both, respectively.

The texture branch is constructed as a residual structure. It contains three fully connected layers and a skip-connection. The size of the output vector $\boldsymbol{V}_{\text{texture}}$ of this branch is equal to that of input $\boldsymbol{v}$, and the number of neurons in all the three fully connected layers is set to k. Each fully connected layer is followed by consecutive operations of batch normalization, ReLU activation function, and dropout (0.5), to enhance the capability of anti-overfitting for the model. The feature vector $\boldsymbol{v}$ will be added directly to the output of the 3rd fully connected layer via skip-connection. The result of the addition of these two vectors constitutes the output vector $\boldsymbol{V}_{\text{texture}}$. The purpose of this design is to learn an optimized texture feature vector through training.

2.3 Texture branch network model

A deep CNN is capable of classifying images. However, it focuses mainly on complex features and ignores the mining of low-level features. Considering that ultrasonic images present a mass of low-level information such as texture, we introduce the texture branch to a typical CNN structure, forming the TBN model. For any image put into the model, $\boldsymbol{V}_{\text{deep}}$ and $\boldsymbol{V}_{\text{texture}}$ are generated from the base network and texture branch in parallel. After a fusion operation, a combined feature vector $\boldsymbol{V}_{\text{comb}} = [\boldsymbol{V}_{\text{deep}}, \boldsymbol{V}_{\text{texture}}]$ can be obtained, where $[\cdot]$ means the concatenation of vectors.

The fusion operation is connected to dense layers for classification. The dense layers are specifically the two fully connected layers mentioned in Section 2.1, and the second layer has two nodes which correspond to classes of CKD and normal. To measure the classification error, cross entropy loss was chosen as the loss function and we used the initialization algorithm suggested by He et al. (2015) to initialize the weights of fully connected layers in the TBN.

3 Chronic kidney disease screening

Before inputting ultrasound images to the proposed model, preprocessing is needed. The preprocessing phase includes ROI extraction and artificial marker repair. Then the data are augmented for training and classification. Finally, the performance should be examined by separated test images.

3.1 Dataset

A total of 226 ultrasound images were involved in this study, each one containing annotation about the stage of CKD as noted by professional radiologists (stages I–V were classified as diseased and stage 0 as normal). All ultrasound images were acquired at kidney routine clinical care from Tongde Hospital of Zhejiang Province, China. Some of the images bore markings made by radiologists. The personal information of patients involved in the data had been removed. One hundred and eighty subjects presented by the kidney ultrasound images were diagnosed as CKD, covering all stages of conditions from CKD-I to CKD-V. The remaining 46 subjects were normal.

3.2 Preprocessing

A portion of these ultrasound images were artificially marked by radiologists to record the location of renal region, cortical thickness, etc. As Fig. 2a displays, artificial markers occlude the texture region and break the integrity of an image, and this affects adversely image analysis. Hence, these markers are to be removed and repaired. First, we converted the image to a grayscale one, due to the high brightness of the mark. The threshold method was used to determine the suspected marking areas, and these pixels were labeled as

$$m_{i,j} = \begin{cases} 1, & f_{i,j} \geq T, \\ 0, & f_{i,j} < T, \end{cases} \tag{12}$$

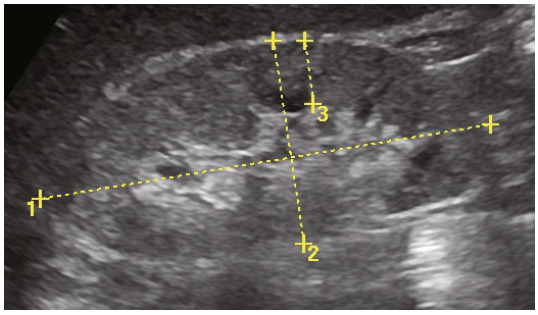
where $\boldsymbol{f}$ is the input gray image, $f_{i,j}$ the pixel value, T the threshold, and $\boldsymbol{m}$ the output binary image. For $f_{i,j} \in S$, where $S = \{(i,j)|m_{i,j} = 1\}$, we adopted a fast marching method (FMM) based inpainting algorithm for restoration (Telea, 2004). For pixel p located on the boundary of region to be inpainted, the approximation $I(p)$ can be given by

$$I(p) = \frac{\sum_{q \in R(p)} w(p,q)\left[I(q) + \bigtriangledown I(q)\boldsymbol{r}\right]}{\sum_{q \in R(p)} w(p,q)}, \tag{13}$$

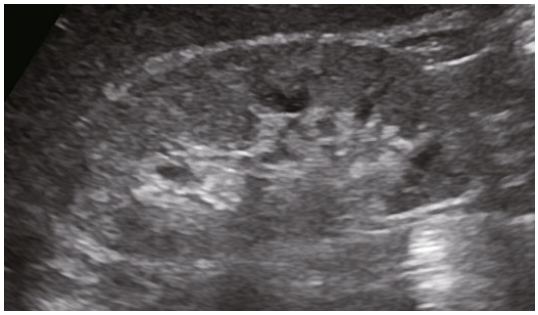
where $R(p)$ is a small neighborhood around p, q is a point in $R(p)$, $\boldsymbol{r}$ is the vector from q to p, and $w(p,q)$ a weighting function defined as

$$w(p,q)=\frac{\boldsymbol{r}\bigtriangledown T(p)}{\|\boldsymbol{r}\|^3(1+|T(p)-T(q)|)}, \quad (14)$$

where $T(k)$ is calculated via FMM depending on whether k is on the boundary of the region to be inpainted. This algorithm computes the approximate value and fills from the periphery of the region to the central area through continuous iteration, until all the regions of markers are restored. As Fig. 2b shows, the markers are entirely repaired.



(a)



(b)

Fig. 2 Removal and reparation of artificial markers in an ultrasound image: (a) image with markers; (b) image after inpainting

The original images show a large echo region, including the kidney and other peripheral organ parts. To eliminate the interference of irrelevant areas and reduce the computational burden of the model, we manually cropped the kidney and uniformed the size to 224×224.

3.3 Data augmentation

Compared with other computer vision tasks, the cost for medical image acquisition and labeling is high, and the available datasets are limited. In addition, the distribution of categories is unbalanced, and the positive samples are nearly four times as many as negative samples. This kind of class-imbalance data is biased to model learning. To this end, the original ultrasound images should be processed to expand the size of examples available. We augmented the data as follows: (1) short-step shifts in random directions; (2) rotation in micro angles (no more than two degrees); (3) random gray-level transformation of pixels ($\leq 3\%$). Based on the particularity of ultrasonic images, it is not feasible to make dramatic adjustments to them, because it may change the texture information and make the labels unavailable.

4 Experiments and results

To verify the effectiveness of the proposed method, we conducted experiments in accordance with the procedure described in the previous section. One hundred and eighty CKD and 46 normal samples were examined. Because of insufficient data, we adopted the assessment approach of five-fold cross validation. These subjects were divided into five equal groups randomly for training and testing. One group was taken out each time as a test set, and the rest as the training set. After data augmentation, 864 images for training were obtained, and the numbers of positive and negative samples in the training set were balanced.

4.1 Evaluation criteria

We set the objects with CKD as positive samples and the normal objects as negative samples. The performance was evaluated by average classification accuracy, sensitivity, specificity, and area under curve (AUC) across the test set. The formulae are as follows:

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}}\times 100\%, \quad (15)$$

$$\text{Sensitivity}=\frac{\text{TP}}{\text{TP}+\text{FN}}\times 100\%, \quad (16)$$

$$\text{Specificity}=\frac{\text{TN}}{\text{TN}+\text{FP}}\times 100\%, \quad (17)$$

where TP, FP, TN, and FN represent the numbers of true positives, false positives, true negatives, and false negatives obtained by classification, respectively. Because of the unbalanced distribution of positive and negative samples, it is not sufficient to simply consider the accuracy. Comprehensively

referring to these indicators makes assessments more meaningful (López et al., 2013).

4.2 Classification

We first attempted to use the conventional machine learning approach to classify ultrasound images with only texture features, to observe the benefits of texture information for ultrasound-guided CKD screening. HOG and GLCM features were selected; in addition, another commonly used texture feature, LBP, was tried. SVM was chosen as the classifier, and the kernel function conformed to a polynomial kernel. The results are shown in Table 2. The method based on LBP features achieved the lowest accuracy, whereas HOG and GLCM texture features had similar good accuracy. The fusion features combining HOG with GLCM showed the optimal performance, and gave an accuracy of 88.48% and AUC score of 0.9096. However, all these methods led to low specificity.

For the proposed deep model, we made a series of experiments and comparisons. All the deep networks were trained using the stochastic gradient descent (SGD) optimizer with momentum of 0.9. We trained the data by batch learning, with the batch size set to 64. Since the size of the training dataset was small, the network fitted rapidly, and we chose 40 epochs for training as the standard procedure. In addition, the initial learning rates of models were defined within the range of [0.001, 0.01]. Because of the different numbers of parameters in diverse models, the learning rate of the beginning phase was adjusted in specific situations. For example, the initial learning rate of TBN with texture branch handling a single type of feature was set at 0.005, and it would be multiplied by 0.2 per eight epochs. To cope with the problem of overfitting, L_2-norm regularization with a weight decay coefficient of 2×10^{-4} was deployed.

Table 2 Classification performance of texture features under machine learning methods

Method	ACC (%)	AUC	SE (%)	SP (%)
LBP+SVM	67.72	0.7534	68.89	63.04
HOG+SVM	83.18	0.8850	87.79	65.22
GLCM+SVM	83.14	0.8746	87.22	67.39
HOG+GLCM+SVM	88.48	0.9096	92.22	73.91

ACC: accuracy; AUC: area under curve; SE: sensitivity; SP: specificity. LBP: local binary pattern; SVM: support vector machine; HOG: histogram of oriented gradient; GLCM: gray-level co-occurrence matrix

To verify the superiority of the TBN model, we compared the performance of TBN models with those of several other methods, including those using only the base network in various training ways, texture branch with different texture features, and a method presented by Zheng et al. (2019). The results are summarized in Table 3. The method suggested by Zheng et al. (2019) was applied to screen congenital anomalies of the kidney and urinary tract in their work, in which HOG feature and pre-trained AlexNet feature were classified by SVM. We implemented this method and tested it on our dataset. In practice, ResNet-34 was selected as the base network for our task because of its moderate model complexity, while other deeper CNNs such as ResNet-101 and DenseNet would fall into overfitting on our dataset. ResNet-34 (S) means training from scratch, while ResNet-34 (T) implies transfer learning. It can be seen that the CNN under the transfer learning method provides a better performance than training from scratch, for the current application. Therefore, we took transfer learning as the default configuration for TBN to train the partition of the basic network. We evaluated various texture branches handling texture features based on HOG, GLCM, and a fusion of them, which are labeled as texture branch (H), texture branch (G), and texture branch (HG), respectively. TBN models with diverse texture branch were tested in the same way. In general, the TBN model received more comprehensive image information and avoided the negative effects of an inadequate training dataset. Thus, the accuracy of various TBN models is higher than that of ResNet-34 or texture branch alone. Among these methods, the greatest accuracy was given by TBN-G, achieving an

Table 3 Comparison of the performance of TBN with those of different texture branch and other related methods

Method	ACC (%)	AUC	SE (%)	SP (%)
ResNet-34 (S)	85.40	0.8914	91.67	60.87
ResNet-34 (T)	91.15	0.9498	96.67	69.57
Texture branch (H)	79.20	0.8344	82.22	67.39
Texture branch (G)	78.31	0.8208	79.44	73.91
Texture branch (HG)	79.65	0.8467	80.56	76.08
Zheng et al. (2019)	85.40	0.8818	90.00	67.39
TBN-H	93.36	0.9409	96.67	80.43
TBN-G	**96.01**	**0.9710**	**99.44**	82.44
TBN-HG	95.13	0.9688	97.22	**86.96**

ACC: accuracy; AUC: area under curve; SE: sensitivity; SP: specificity. The highest values are in bold

accuracy of 96.01%, an AUC score of 0.9710, and a sensitivity of 99.44%. The performance of the TBN-H method was lower than that of the former, and the accuracy of the TBN-HG model which combines the two texture features was also slightly inferior to that in the optimal case, but it showed the highest specificity of 86.96%.

5 Discussion

In this study, we demonstrate that a CNN framework with a texture branch can be applied to CKD screening based on ultrasound images. As an additional structure for texture feature extraction and optimization, the texture branch offers a typical CNN model which diversifies information. Compared to the performance of SVM using fused texture features and ResNet-34 under transfer learning, the accuracy of identifying CKD was improved by 7.53% and 4.86%, respectively in the TBN-G model; in terms of specificity, the results revealed benefits of 13.05% and 17.39% improvement in the TBN-HG model, respectively.

Notably, the specificity of classification revealed benefits (17.39% improvement compared to pure ResNet-34) in the TBN-HG method. Although it seems that the accuracy is lower than that of the TBN-G model that extracts only GLCM-based features, considering the small number of normal samples in our dataset, it can be predicted that the performance of a model using multi-texture features will be better than that using single-texture features if the data volume can be improved and the classes tend to be balanced.

As mentioned in Section 1, the gold standard for diagnosis and staging of CKD is on the basis of GFR, which is estimated by biochemical testing. Therefore, it is challenging to judge CKD based on only pathological lesions reflected in ultrasound images, even for professional radiologists. After all, the extent of lesion within the renal parenchyma does not perfectly match the differentiation in renal echogenicity, especially in the early stage of CKD. To confirm that, we counted the samples misclassified by the TBN-HG model and found that most of the false-negative samples were in the early stage (labeled as CKD-I). Some typical wrongly classified cases are displayed in Fig. 3, It can be seen that originally healthy images but classified as diseased (Figs. 3c and 3d) contain some abnormal shadowing areas appearing around the kidney. These abnormal echoes are called ultrasound artifacts, which are commonly encountered but affect the quality of an ultrasound image, particularly the visual characteristics of the texture. These artifact noises may mislead the model to CKD cases diagnosed with renal cysts, which also have some regular hypoechoic areas.

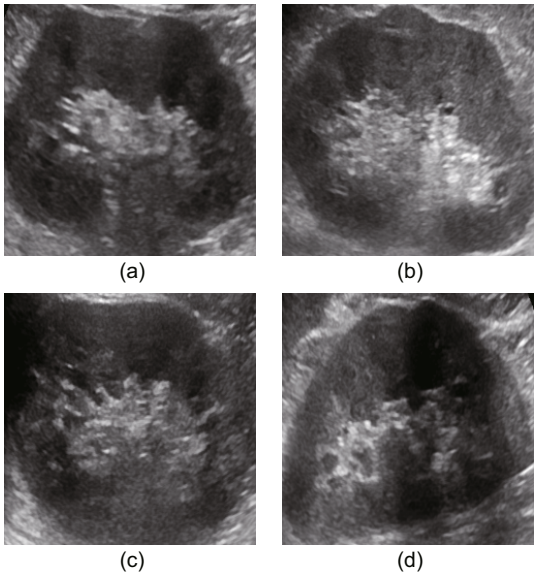


Fig. 3 Examples misclassified by the TBN-HG model: (a) false negative sample (CKD-I actually); (b) false negative sample (CKD-II actually); (c, d) false positive samples (normal actually)

The unbalanced distribution of positive and negative samples increases the risk of bias on classification. Consequently, it is essential to capture more images and category balance should also be considered in the data acquisition phase. With a diversified data distribution, the learning capacity of deep models can be better performed, theoretically.

In future work, we intend to explore effective screening methods in larger datasets, maintaining a high sensitivity of classification while continuing to improve specificity. We also plan to explore the CKD staging challenge, which is a multi-classification task.

6 Conclusions

We presented a novel TBN model for ultrasound-guided CKD screening, which used a texture branch to extract texture descriptors and supplement the deep features from a CNN.

Our results showed the great potential of the deep learning method in the CKD classification problem. In addition, under the limitation of an unbalanced small-sample dataset, the scheme of fusing texture features and deep features, combined with the training approach of transfer learning, showed an excellent classification accuracy. Due to high sensitivity, the proposed method has the potential to be applied as computer-aided screening for CKD.

Compliance with ethics guidelines

Peng-yi HAO, Zhen-yu XU, Shu-yuan TIAN, Fu-li WU, Wei CHEN, Jian WU, and Xiao-nan LUO declare that they have no conflict of interest.

References

Acharya UR, Meiburger KM, Koh JEW, et al., 2019. Automated detection of chronic kidney disease using higher-order features and elongated quinary patterns from B-mode ultrasound images. *Neur Comput Appl*, in press. https://doi.org/10.1007/s00521-019-04025-y

Ahmad M, Tundjungsari V, Widianti D, et al., 2017. Diagnostic decision support system of chronic kidney disease using support vector machine. Proc 2nd Int Conf on Informatics and Computing, p.1-4. https://doi.org/10.1109/IAC.2017.8280576

Batra A, Batra U, Singh V, 2016. A review to predictive methodology to diagnose chronic kidney disease. Proc 3rd Int Conf on Computing for Sustainable Global Development, p.2760-2763.

Chang EH, Chong WK, Kasoji SK, et al., 2017. Diagnostic accuracy of contrast-enhanced ultrasound for characterization of kidney lesions in patients with and without chronic kidney disease. *BMC Nephrol*, 18, Article 266. https://doi.org/10.1186/s12882-017-0681-8

Charleonnan A, Fufaung T, Niyomwong T, et al., 2016. Predictive analytics for chronic kidney disease using machine learning techniques. Proc Management and Innovation Technology Int Conf, p.MIT-80-MIT-83. https://doi.org/10.1109/MITICON.2016.8025242

Chen H, Dou Q, Ni D, et al., 2015. Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks. Proc 18th Int Conf on Medical Image Computing and Computer-Assisted Intervention, p.507-514. https://doi.org/10.1007/978-3-319-24553-9_62

Dalal N, Triggs B, 2005. Histograms of oriented gradients for human detection. Proc IEEE Computer Society Conf on Computer Vision and Pattern Recognition, p.886-893. https://doi.org/10.1109/CVPR.2005.177

Dhindsa K, Smail LC, McGrath M, et al., 2018. Grading prenatal hydronephrosis from ultrasound imaging using deep convolutional neural networks. Proc 15th Conf on Computer and Robot Vision, p.80-87. https://doi.org/10.1109/CRV.2018.00021

Ecder T, 2013. Early diagnosis saves lives: focus on patients with chronic kidney disease. *Kidney Int Suppl*, 3(4):335-336. https://doi.org/10.1038/kisup.2013.70

El Nahas AM, Bello AK, 2005. Chronic kidney disease: the global challenge. *Lancet*, 365(9456):331-340. https://doi.org/10.1016/S0140-6736(05)17789-7

Haralick RM, Shanmugam K, Dinstein I, 1973. Textural features for image classification. *IEEE Trans Syst Man Cybern*, SMC-3(6):610-621. https://doi.org/10.1109/TSMC.1973.4309314

He KM, Zhang XY, Ren SQ, et al., 2015. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. Proc IEEE Int Conf on Computer Vision, p.1026-1034. https://doi.org/10.1109/ICCV.2015.123

He KM, Zhang XY, Ren SQ, et al., 2016. Deep residual learning for image recognition. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.770-778. https://doi.org/10.1109/CVPR.2016.90

Ho CY, Pai TW, Peng YC, et al., 2012. Ultrasonography image analysis for detection and classification of chronic kidney disease. Proc 6th Int Conf on Complex, Intelligent, and Software Intensive Systems, p.624-629. https://doi.org/10.1109/CISIS.2012.180

Hsieh JW, Lee CH, Chen YC, et al., 2014. Stage classification in chronic kidney disease by ultrasound image. Proc 29th Int Conf on Image and Vision Computing New Zealand, p.271-276. https://doi.org/10.1145/2683405.2683457

Iqbal F, Pallewatte AS, Wansapura JP, 2017. Texture analysis of ultrasound images of chronic kidney disease. Proc 17th Int Conf on Advances in ICT for Emerging Regions, p.1-5. https://doi.org/10.1109/ICTER.2017.8257787

Jeewantha RA, Halgamuge MN, Mohammad A, et al., 2017. Classification performance analysis in medical science: using kidney disease data. Proc Int Conf on Big Data Research, p.1-6. https://doi.org/10.1145/3152723.3152724

Kunwar V, Chandel K, Sabitha AS, et al., 2016. Chronic kidney disease analysis using data mining classification techniques. Proc 6th Int Conf on Cloud System and Big Data Engineering, p.300-305. https://doi.org/10.1109/CONFLUENCE.2016.7508132

Levey AS, Eckardt KU, Tsukamoto Y, et al., 2005. Definition and classification of chronic kidney disease: a position statement from kidney disease: improving global outcomes (KDIGO). *Kidney Int*, 67(6):2089-2100. https://doi.org/10.1111/j.1523-1755.2005.00365.x

López V, Fernández A, García S, et al., 2013. An insight into classification with imbalanced data: empirical results and current trends on using data intrinsic characteristics. *Inform Sci*, 250:113-141. https://doi.org/10.1016/j.ins.2013.07.007

Oquab M, Bottou L, Laptev I, et al., 2014. Learning and transferring mid-level image representations using convolutional neural networks. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.1717-1724. https://doi.org/10.1109/CVPR.2014.222

Pujari RM, Hajare VD, 2014. Analysis of ultrasound images for identification of chronic kidney disease stages. Proc 1st Int Conf on Networks & Soft Computing, p.380-383. https://doi.org/10.1109/CNSC.2014.6906704

Sharma K, Virmani J, 2016. Classification of renal diseases using first order and higher order statistics. Proc 3rd

Int Conf on Computing for Sustainable Global Development, p.425-430.

Shen W, Zhou M, Yang F, et al., 2015. Multi-scale convolutional neural networks for lung nodule classification. Proc 24th Int Conf on Information Processing in Medical Imaging, p.588-599. https://doi.org/10.1007/978-3-319-19992-4_46

Shin HC, Roth HR, Gao MC, et al., 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imag*, 35(5):1285-1298. https://doi.org/10.1109/TMI.2016.2528162

Subramanya MB, Kumar V, Mukherjee S, et al., 2015. SVM-based CAC system for B-mode kidney ultrasound images. *J Dig Imag*, 28(4):448-458. https://doi.org/10.1007/s10278-014-9754-4

Telea A, 2004. An image inpainting technique based on the fast marching method. *J Dig Imag*, 9(1):23-34. https://doi.org/10.1080/10867651.2004.10487596

Zheng Q, Furth SL, Tasian GE, et al., 2019. Computer-aided diagnosis of congenital abnormalities of the kidney and urinary tract in children based on ultrasound imaging data by integrating texture image features and deep transfer learning image features. *J Pediatr Urol*, 15(1):75.e1-75.e7. https://doi.org/10.1016/j.jpurol.2018.10.020

Fu-li WU, corresponding author of this invited paper, is an associate professor at Zhejiang University of Technology. He received his PhD degree from Zhejiang University, China in 2005. From 2006 to 2007, he was a post-doctor of the University of Bedfordshire. From 2012 to 2013, he was a visiting scholar of the University of California-Davis. His research interests include computer graphics, data visualization, and medical image processing.

Prof. Jian WU provides conditions for the smooth progress of project research and gives guidance for the technologies of this invited paper. He received a bachelor's degree and a PhD from College of Computer Science and Technology, Zhejiang University, China. He is the director of Real Doctor AI Research Center of Zhejiang University, member of CCF Youth Working Committee, CCF TCSC, and CCF TCAPP, one of the 151 Talents Program from Zhejiang Province, and a member of the key field innovation team of the Ministry of Science and Technology. His research interests focus on artificial intelligence in medicine, services computing, and so on.