Multi-instance deep learning of ultrasound imaging data for pattern classification of congenital abnormalities of the kidney and urinary tract in children

Supplementary data

We adopted the transfer learning to fine-tune a deep learning classification network based on a pre-trained VGG16 model for learning discriminative features from individual 2D ultrasound images and estimating instance-level classification scores. The VGG16 deep learning model consists of 13 convolutional (conv) layers with a receptive field of 3×3.[1] The stack of convolutional layers is followed by 3 fully connected (FC) layers, each of the first two layers having 4096 channels and the third layer performing 1000-way classification. The final layer is a softmax layer.

The VGG16 model was modified for CAKUT diagnosis at an instance-level, as illustrated by Supplementary Figure 1. First, we modified the last 3 convolution layers into 3 atrous convolutional layers to obtain a denser feature extraction [2]. The atrous convolution stride is 2 and the convolution filter size is 3*3. Second, we modify the first two FC layers as 1 FC layers with the 256 channels to reduce the memory size. Third, the last output classification FC layer is modified as a 2-way output for the CAKUT diagnosis problem. Given a 2D US image x_i, the deep learning model yields p_i and $1-p_i$, encoding probability values for the image to have a positive label (i.e., CAKUT) and a negative label (i.e., control), respectively.

We trained the deep learning model by refining parameters of a pretrained VGG16 model.[2] Particularly, the first 10 convolution layers (denoted as VGG16 conv block) and the three atrous convolution were initialized by adopting parameters of the pretrained VGG16 model, and the FC layers were randomly initialized using Glorot uniform initialization.[3] The model was trained using instance-level softmax loss function. All kidney images of the same individual had the same class label as the individual. In other words, all kidney images of children with CAKUT had a class label of +1, and all images of controls had a class label of -1. Applying the trained model to a 2D kidney image will yield a probability score of CAKUT.

We adopted the commonly used mean pooling operator to compute an overall bag-level classification score and also compared it with the max pooling operator. Particularly, the mean and max pooling operators are defined as

$$P^{\text{mean}} = \frac{1}{I}\sum_{i=1}^{I} p_i,\ P^{\max} = \max_{p_i, i=1,\dots,I} p_i,$$

where p_i is the classification score of an image $x_i, i = 1, \dots, I$, belonging to the same individual.

We evaluated the classification performance of the MIL models that were built on images in the sagittal view, the transverse view, or both views. We also compared the MIL models built using the mean and max pooling operators. Supplementary Table 1 summarizes all evaluation in terms of classification models trained and tested using images in different views.

Supplementary Table 2 summarizes AUC values of different MIL models on testing datasets obtained using mean pooling and max pooling operators. The AUC values obtained using the max pooling operator were close to 0.94, while those obtained using the mean pooling operator were larger than 0.96, indicating that the mean pooling operator could yield better classification performance.

We also obtained classification models using pre-trained deep learning models: ResNet V2 101 version for the ResNet and Inception-ResNet-v2 version for the Inception network, all from the TensorFlow-Slim image classification model library.[4] We used the same 5-fold cross-validation method to estimate their performance. Mean pooling was used to obtained MIL classifiers.

Supplementary Figure 2 and Supplementary Figure 3 show class activation mapping results of representative images obtained by different classification models. Supplementary Figure 4 shows representative CAKUT images misclassified by the classification models trained with the VGG16 model.

References

1. Karen Simonyan AZ. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556.* 2014.

2. Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence.* 2017;40:834-848.

3. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the thirteenth international conference on artificial intelligence and statistics*2010:249-256.

4. Silberman N, Guadarrama S. TensorFlow-Slim image classification model library2016.

Supplementary Figure Legends:

Supplementary Figure 1. Deep learning for CAKUT diagnosis at an instance level.

Supplementary Figure 2. Class activation mapping results of randomly selected CAKUT and control kidney images obtained by different deep learning models trained on images in sagittal view. Regions in warm color contributed more to the classification than those in cold color.

Supplementary Figure 3. Class activation mapping results of randomly selected CAKUT and control kidney images obtained by different deep learning models trained on images in transverse view. Regions in warm color contributed more to the classification than those in cold color.

Supplementary Figure 4. Representative CAKUT images that were misclassified by the deep leaning classifiers.