Program Digest, 2019 IEEE International Ultrasonics Symposium (IUS)
Glasgow, Scotland, October 6-9, 2019

Integrating Clinical Knowledge in a Thyroid Nodule Classification Model Based on

1st Shijie Zhang
Academy for Advanced Interdisciplinary Studies
Peking University
Beijing, China
zhang_shijie@foxmail.com

2nd Huarui Du
E-government and Engineering Center
National Development and Reform Commission
Beijing, China
huaruidu@163.com

3rd Zhuang Jin
Ultrasonography Department
the General Hospital of the People's Liberation Army
Beijing, China
chaoshengke2017@163.com

4th Yaqiong Zhu
Ultrasonography Department
the General Hospital of the People's Liberation Army
Beijing, China
455155808@qq.com

5th Ying Zhang
Ultrasonography Department
the General Hospital of the People's Liberation Army
Beijing, China
zygpq@163.com

6th Fang Xie
Ultrasonography Department
the General Hospital of the People's Liberation Army
Beijing, China
lmsh1225@163.com

7th Mingbo Zhang
Ultrasonography Department
the General Hospital of the People's Liberation Army
Beijing, China
owsifanduizhe@126.com

8th Ziyu Jiao
Ultrasonography Department
the General Hospital of the People's Liberation Army
Beijing, China
jiaoziyujiao@163.com

9th Xiaoqi Tian
Ultrasonography Department
the General Hospital of the People's Liberation Army
Beijing, China
410255527@qq.com

10th Jiabin Zhang
Academy for Advanced Interdisciplinary Studies
Peking University
Beijing, China
zhang_shijie@foxmail.com

11th Hanjing Kong
Academy for Advanced Interdisciplinary Studies
Peking University
Beijing, China
zhang_shijie@foxmail.com

12th Jian An
Academy for Advanced Interdisciplinary Studies
Peking University
Beijing, China
zhang_shijie@foxmail.com

13th Yukun Luo
Ultrasonography Department
the General Hospital of the People's Liberation Army
Beijing, China
lyk301@163.com

14th Jue Zhang
Academy for Advanced Interdisciplinary Studies &
College of Engineering
Peking University
Beijing, China
zhangjue@vip.163.com

***Abstract*—Deep neural network models are currently facing problems due the weak interpretability of predicted results. One solution would involve integrating the clinical knowledge and classifiers. In this study, we utilized mature clinical experiences from TI-RADS, and also trained a multi-label cost-sensitive classification network. The results show that the end-to-end classification model based on cost-sensitive loss function is a more effective way to integrate clinical knowledge into the deep neural network.**

Keywords—Thyroid nodules, Ultrasound imaging, Thyroid cancer

I. Introduction

While the reasoning process for diagnosis results is advocated in evidence-based medicine, deep neural network models, especially Convolutional Neural Network (ConvNet) are currently facing problems due the weak interpretability of predicted results. In recent years, a growing number of researchers have realized that high interpretability is of significant value in practical applications, and have developed models with interpretable knowledge representations (1). However, these knowledge representations are too nebulous to be useful in clinical use. ”let the data speak” is problematic (2). One solution would involve integrating the clinical knowledge and classifiers. However, the accuracy of human inference is lower than the pathological gold standard for the training data sets (3), and thus the performance of classifier is reduced. Therefore, an interpretable classification model based on cost-sensitive loss function is proposed in this paper. The proposed model is an end-to-end model which was trained by pathological results. Several clinical features were used to constrain the intermediate results of the model.

II. Methods

A. Data collection

These experiments were approved by the Research Ethics Committee of The General Hospital of the People's Liberation Army. Data were collected from 843 participants, containing approximately 4368 images. These images were are collected

his work was supported in part by “the National Natural Science Foundation of China 81771834”.

with Siemens ACUSON S2000, Philips iU22, Esaote MyLab Twice , the Siemens ACUSON SEQUOIA 512, the ,Hitachi HI VISION Ascendus, PHILIPS iU Elite and GE Vivid E9 ultrasound systems with a high-frequency probe. Data including the transverse and longitudinal images, as well as the 2202 images wereas marked witha ACR TI-RADS features by an experienced radiologist.

Each of the images selected a region of interest (ROI) and contained a thyroid nodule by experienced radiologists. Aspect ratio $R_a = \frac{w_r}{h_r}$ and echogenicity ratio $R_e = \frac{E_n}{E_t}$ are calculated for each of the images , which w_r is the width of ROI, h_r is the height of ROI ,E_n is the average intensity in ROI and E_t is the average intensity around the ROI.

B. Thyroid Nodule Classification Model

In order to use the clinical experience in the network training, we divided the model into two segments, as Fig. 1. shown. A ConvNet is trained to identify 3 types of 8 clinical features of nodules. The network output is mixed cystic and solid, solid, irregular, smooth, comet-tail artifacts, macro-calcifications, peripheral calcifications, punctate echogenic foci. Then, a fully connected network is applied as a means to learn the correspondence between clinical features of the classification of benign and malignant tumors of the thyroid. Finally, we connected the two models model.
Instead of training from scratch, a ResNet50 network trained on the ILSVRC dataset (4) was used as the pre-trained ConvNet model. As a typical DCNN, the Resnet50 consists of five stage of convolutional layers, and an average pooling layers followed by a 1000-way fully connected layer (5).

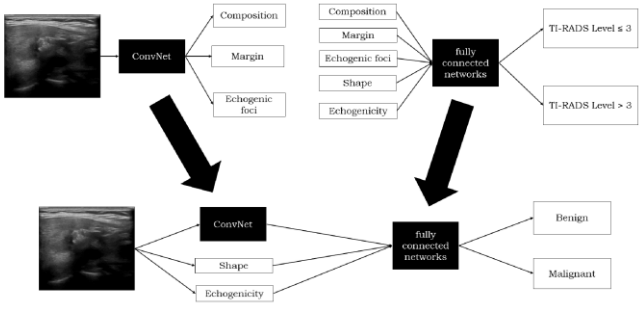

Fig. 1. The TI-RADS network structure diagram. Networks are divided into two parts: the Feature extraction network and the TI-RADS grading network.

In the training phase, both benign\malignant and clinical features were used as labels to train the network end-to-end. The data only labeled as benign or malignant was used to train the network alternately to improve the performance.

A key problem for the training phase was that clinical features differ widely in their occurrence frequency. Clinical features differ widely in their occurrence frequency. Currently, the commonly used methods of imbalanced data are image generation in fewer classes. However, in multi-label classification, it is necessary to generate images to make all classes in equalizing balance which is difficult to implement and usually has a negative performance impact. Therefore, we proposed a multi-label cost-sensitive classification network for learning clinical features. As show in Fig. 2. , we converted the multi-label classification at the top of the network into various binary classifications. Each feature is a binary classification with a softmax.
To solve the impact of imbalanced data, one solution is to make the lost function cost-sensitive:

$$L\big(Y, f(x)\big) = \lambda\big(Y, f(x)\big)\big(Y, f(x)\big)^2 \qquad (1)$$

where $\lambda\big(Y, f(x)\big) > 0$ is the cost of network output $f(x)$and the ground truth Y. $Y = \{Y_1, \dots, Y_n\}$, where n is the number of the network output. The function $f_i \epsilon [0,1]$ represents the real-valued score of the neural network output , $i \in \{1, \dots, n\}$.We defined a cost matrix $C = \{C_1, \dots, C_n\}$, where $Ci = \{C_i^1, C_i^2\}$ is a two-dimensional vector, where C_i^1 is false positive cost of f_i and C_i^2 is false negative cost of f_i. Ideally, when $Y_i = 0$, as the fi grows, C_i^1dominates the λ, and $\lambda = C_i^1$ while $f_i = 1, \lambda = 0$, while $f_i = 0$; when $Y_i = 1$,as the f_i grows, C_i^2dominate the λ, and $\lambda = C_i^2$ while $f_i = 0, \lambda = 0$, while $f_i = 1$. Thus, the cost λ is deifine as:

$$\lambda_i = C_i^1 (f_i - Y_i) + C_i^2 (Y_i - f_i). \qquad (2)$$

In order to ensure $\lambda_i \geq 0$, $(f_i - Y_i)$ and $(Y_i - f_i)$ plus a constant , thus cost λ difine as:

$$\lambda_i = C_i^1 (f_i - Y_i + 1) + C_i^2 (Y_i - f_i + 1). \quad (3)$$

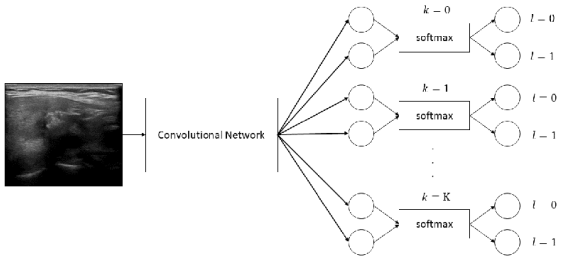

Fig. 2. Feature extraction network. Each feature is a binary classification with a softmax.

III. EXPERIMENTS

Below are the results of a multicenter experiment on a dataset consisting of 14,867 images from 5,131 patients. 14,220 images from 4,794 patients comprised the training data set, while 647 images from 337 patients comprised the test data set. The proposed model offered significantly improved performance (accuracy: 0.813, sensitivity: 0.833, specificity: 0.799, AUC: 0.858), compared to clinical feature detection trained separately (accuracy: 0.753, sensitivity: 0.784, specificity: 0.728, AUC: 0.785) and radiologists (average accuracy: 0.764, sensitivity: 0.741, specificity: 0.845). These results show that the end-to-end classification model based on cost-sensitive loss function is a more effective way to integrate clinical knowledge into the deep neural network.

TABLE I. COMPARISON OF THE DIAGNOSTIC PERFORMANCE

	Accuracy	***Sensitivity***	***Specificity***
Radiologist	76.4%	74.1%	84.5%
Clinical Feature Detection Trained Separately	75.3%	78.4%	72.8%
Proposed Model	81.3%	83.3%	79.9%

IV. CONCLUSION

In conclusion, we proposed a new strategy for integrate clinical knowledge into the deep neural network in thyroid nodules classification based on ultrasound images. With this novel strategy, tests show that the proposed model performance has exceeded the clinical labels from radiologist.

REFERENCES

[1] Q. Zhang and S.-C. Zhu, "Visual interpretability for deep learning: a survey," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, no. 1, pp. 27–39, 2018.

[2] Z. Obermeyer and E. J. Emanuel, "Predicting the future - big data, machine learning, and clinical medicine," *The New England journal of medicine*, vol. 375, no. 13, p. 1216—1219, September 2016. [Online]. Available: http://europepmc.org/articles/PMC5070532

[3] R. Gilles, R. Bénédicte, B. Claude, R. Agnès, B. P. Marie, and L. Laurence, "Prospective evaluation of thyroid imaging reporting and data system on 4550 nodules with and without elastography," *European Journal of Endocrinology*, vol. 168, no. 5, pp. 649–655, 2013.

[4] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and F. F. Li, "Imagenet: a large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 248–255.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778.