

Received: 2020.06.23

Accepted: 2020.07.31

Available online: 2020.08.10

Published: 2020.08.15

Visual Interpretability in Computer-Assisted Diagnosis of Thyroid Nodules Using Ultrasound Images

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

ADEFG 1 **Xi Wei**
ACDEF 1 **Jialin Zhu**
CDF 2 **Haozhi Zhang**
BCF 3 **Hongyan Gao**
BCD 4 **Ruiguo Yu**
BCD 4 **Zhiqiang Liu**
BF 2 **Xiangqian Zheng**
AF 2 **Ming Gao**
ABDEF 1 **Sheng Zhang**

1 Department of Diagnostic and Therapeutic Ultrasonography, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center of Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin, P.R. China

2 Department of Thyroid and Neck Cancer, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center of Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin, P.R. China

3 Department of Ultrasonography, Tianjin Xiqing District Women and Children's Health and Family Planning Service Center, Tianjin, P.R. China

4 College of Intelligence and Computing, Tianjin University, Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin Key Laboratory of Advanced Networking, Tianjin, P.R. China

Corresponding Authors: Jialin Zhu, e-mail: zhujialin@tmu.edu.cn, Sheng Zhang, e-mail: zs19620112@126.com

Source of support: This study was supported by the National Natural Science Foundation of China (grant no. 81771852) and the Major Scientific and Technological Project for A New Generation of Artificial Intelligence of Tianjin (grant no. 18ZXZNSY00300)

Background: The number of studies on deep learning in artificial intelligence (AI)-assisted diagnosis of thyroid nodules is increasing. However, it is difficult to explain what the models actually learn in artificial intelligence-assisted medical research. Our aim is to investigate the visual interpretability of the computer-assisted diagnosis of malignant and benign thyroid nodules using ultrasound images.

Material/Methods: We designed and implemented 2 experiments to test whether our proposed model learned to interpret the ultrasound features used by ultrasound experts to diagnose thyroid nodules. First, in an anteroposterior/transverse (A/T) ratio experiment, multiple models were trained by changing the A/T ratio of the original nodules, and their classification, accuracy, sensitivity, and specificity were tested. Second, in a visualization experiment, class activation mapping used global average pooling and a fully connected layer to visualize the neural network to show the most important features. We also examined the importance of data preprocessing.

Results: The A/T ratio experiment showed that after changing the A/T ratio of the nodules, the accuracy of the neural network model was reduced by 9.24–30.45%, indicating that our neural network model learned the A/T ratio information of the nodules. The visual experiment results showed that the nodule margins had a strong influence on the prediction of the neural network.

Conclusions: This study was an active exploration of interpretability in the deep learning classification of thyroid nodules. It demonstrated the neural network-visualized model focused on irregular nodule margins and the A/T ratio to classify thyroid nodules.

MeSH Keywords: **Artificial Intelligence • Image Interpretation, Computer-Assisted • Thyroid Nodule • Ultrasonography**

Full-text PDF: <https://www.medscimonit.com/abstract/index/idArt/927007>

 2728

 6

 7

 41



Background

The evaluation of thyroid nodules according to recommended guidelines for ultrasound examination and reporting has become routine [1]. Deep learning assisted by radiologists' reviews of medical images has also been used in recent years, and its application in medical-aided diagnosis has been developed for use in eye diseases, brain diseases, lung inflammation, and thyroid nodules using medical images including computerized tomography, magnetic resonance imaging, X-ray, and ultrasound [2–5]. In the field of artificial intelligence (AI)-assisted ultrasound diagnosis, several researchers used a convolutional neural network in the ImageNet database to obtain good diagnostic accuracy by connecting feature images [6–8]. In recent years, many improved methods based on a convolutional neural network have gradually emerged in the neural network field [9,10]. In general, the existing AI-assisted medical research models can be divided into 2 categories based on training data: one uses full images for model training and the other uses the region of interest (ROI), which is extracted from the full images, for training [11–17]. However, there is currently no unified method for selecting training data for neural networks.

The accuracy of neural network-assisted diagnosis is better than the performance of doctors in some diseases [18,19]. However, it has been difficult to explain what the models actually learn in AI-assisted medical research. Because of the complex nature of deep learning, it is impossible to know exactly what features are learned by neural networks, which are essentially intelligent classifiers [20]. Thus, AI-assisted diagnosis still lacks a scientific medical explanation, and there is a long way to go before fully realizing AI-assisted medical care [21,22]. In the diagnosis of thyroid nodules by radiologists, ultrasound examination can identify the size, number, location, composition, shape, margin, echogenicity, calcification, and blood flow signals. Among these, the composition, margin, echogenicity, calcification, and anteroposterior/transverse (A/T) ratio (a taller-than-wide shape) are important criteria for radiologists to distinguish between benign and malignant nodules [23–26]. However, whether these criteria are also important for deep learning networks is still unknown.

In this study, we aim to identify the features the models learned in the classification of thyroid nodules on ultrasound images by conducting visualization experiments and to verify that some ultrasound features, such as the A/T ratio or margin, are crucial points for neural networks to distinguish malignant nodules from benign nodules.

Material and Methods

Data introduction

A total of 7 216 thyroid ultrasound images from the database of the Tianjin Medical University Cancer Institute and Hospital were used in this study. Between January 2018 and October 2018, consecutive patients in its 4 medical centers who underwent a diagnostic thyroid ultrasound examination and subsequent surgery were included in the study. All the thyroid nodules were confirmed by postoperative pathological diagnosis. The exclusion criteria were as follows: (1) images from anatomical sites were judged as not having a tumor according to postoperative pathology; (2) nodules with incomplete (one or both orthogonal plane images were missing) or unclear ultrasound images; and (3) cases with incomplete clinicopathological information. This study was approved by the Tianjin Medical University Cancer Institute and Hospital Ethics Committee. The need for medical informed consent from patients was waived because of the retrospective design of the study.

The original ultrasound data were scanned and marked by 3 radiologists from the Tianjin Medical University Cancer Institute and Hospital. The original images were 1024×768 pixels, with a pixel depth of 8. Employing a fully convolutional neural network, we used information on the position of the nodules marked by the doctors to obtain the nodule mask shape of each ultrasound image. The A/T ratio distribution of the nodules in the dataset was classified according to the masks.

Data preprocessing

The original thyroid ultrasound images had black bands containing textual information, such as the machine model, nodular crosscut/longitudinal cut, and nodule size, which may have interfered with the training of the neural network to varying degrees. Therefore, preprocessing was required to remove the information on the surrounding black bands. Moreover, the ultrasound images in the dataset were produced by various types of ultrasound scanners; therefore, the positions and sizes of the black bands were different and could not be processed uniformly. We used a convolutional neural network to remove the information on the black bands of the ultrasound images, and, at the same time, the sizes of the images were normalized without any manual annotation or changes to the original scale. For detecting which types of images were the most appropriate for training and testing in deep learning, we tested the diagnostic performance of 3 types: whole images, images with the black edges cut off, and the ROI, which was used to localize and diagnose thyroid nodules.

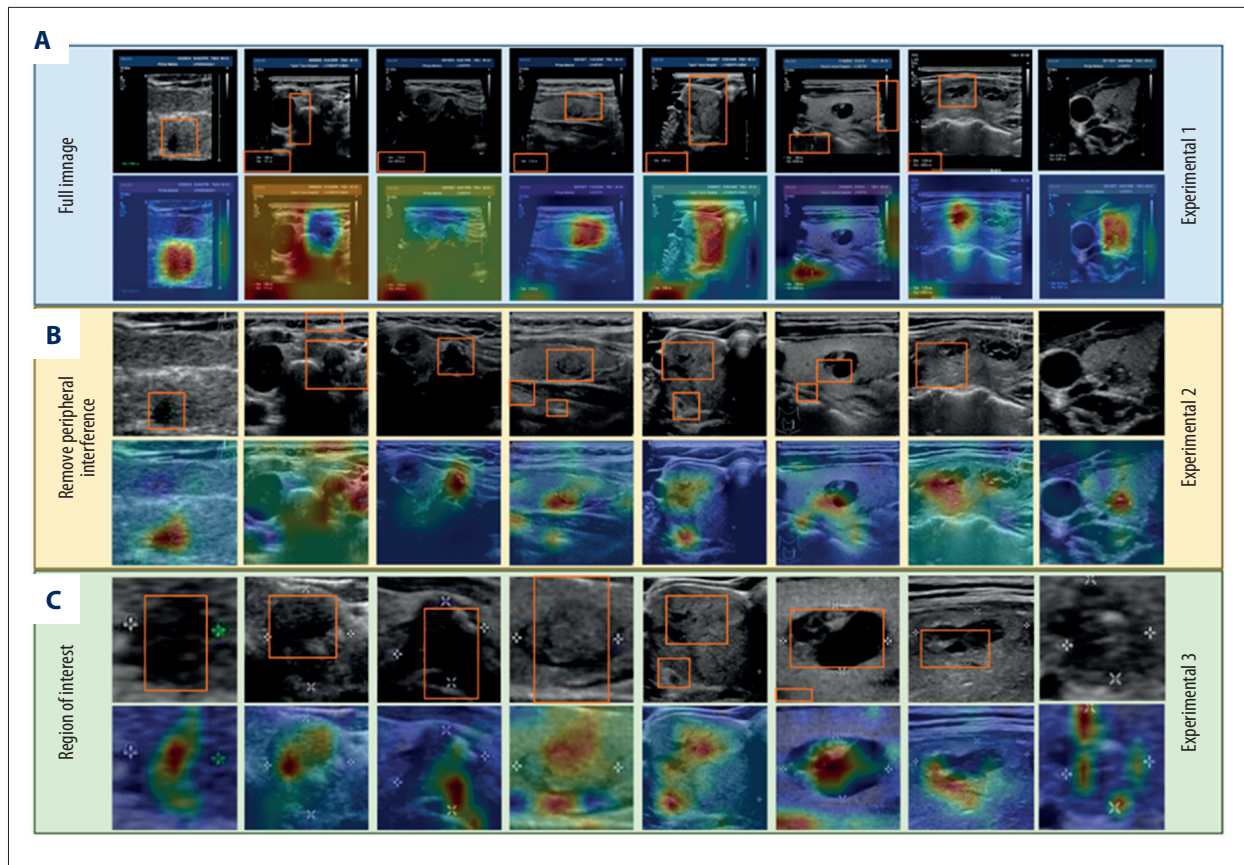


Figure 1. Corresponding visualization results are shown for experiments 1–3. The areas of the orange frames represent the visualization results on the datasets. Each column shows the same ultrasound image with different degrees of segmentation. **(A)** Experiment 1: The model was trained and tested using full images. **(B)** Experiment 2: The model was trained and tested using images after the removal of peripheral interference. **(C)** Experiment 3: The model was trained and tested using the region of interest (ROI).

Visualization method of the neural network

In this study, a boundary visualization model to observe the effect of nodule boundaries during the prediction of neural networks was proposed. The structure of the visualization model is shown in Supplementary Table 1. Sketches of the AI visualization method are shown in Supplementary Figure 1. The batch size was set to 20, and the initial learning rate was set to 0.001 (the learning rate was multiplied by 0.1 per 100 iterations). The training was stopped when the accuracy of the model tended to be stable. The original thyroid ultrasound images were surrounded by textual information, which could have caused some interference in the model, as shown in Figure 1. Therefore, the ultrasound images needed to be cropped and only the ultrasonic imaging region was retained. In addition, in the traditional visualization method, feature maps in the last layer are directly calculated by global average pooling, whereby too much information is lost. In the backward calculation, 1 channel of the feature map corresponds to only 1 weight, as shown in Figure 2A. The boundary visualization model was

proposed to address the above problem by strengthening the global average pooling layer to a fully connected layer. In the backward calculation, each layer of the feature map corresponded to multiple weights, so the visualization was more refined (Figure 2B).

Experimental methods to judge the importance of the A/T ratio of nodules

To explore the sensitivity of the neural network to the A/T ratio, we proposed an experimental method which included 3 main steps. The structure of the model is shown in Supplementary Table 2. First, we defined the original thyroid ultrasound image dataset as $I = \{I_1, I_2, \dots, I_n\}$, where n is the number of images in the dataset. Additionally, $I = B \cup M$, represents benign and malignant images, respectively. Through ROI position extraction, a position set P was obtained, where $P_i = \{x, y, h, w\}$, and x, y, h, w represent the coordinates, and height and width of the ROI, respectively. Meanwhile, the original A/T dataset, $\{B_o, M_o\} = \{\text{Crop}(I_i, x, y, b, b) | i=1, \dots, n, b = \max(h, w)\}$, was

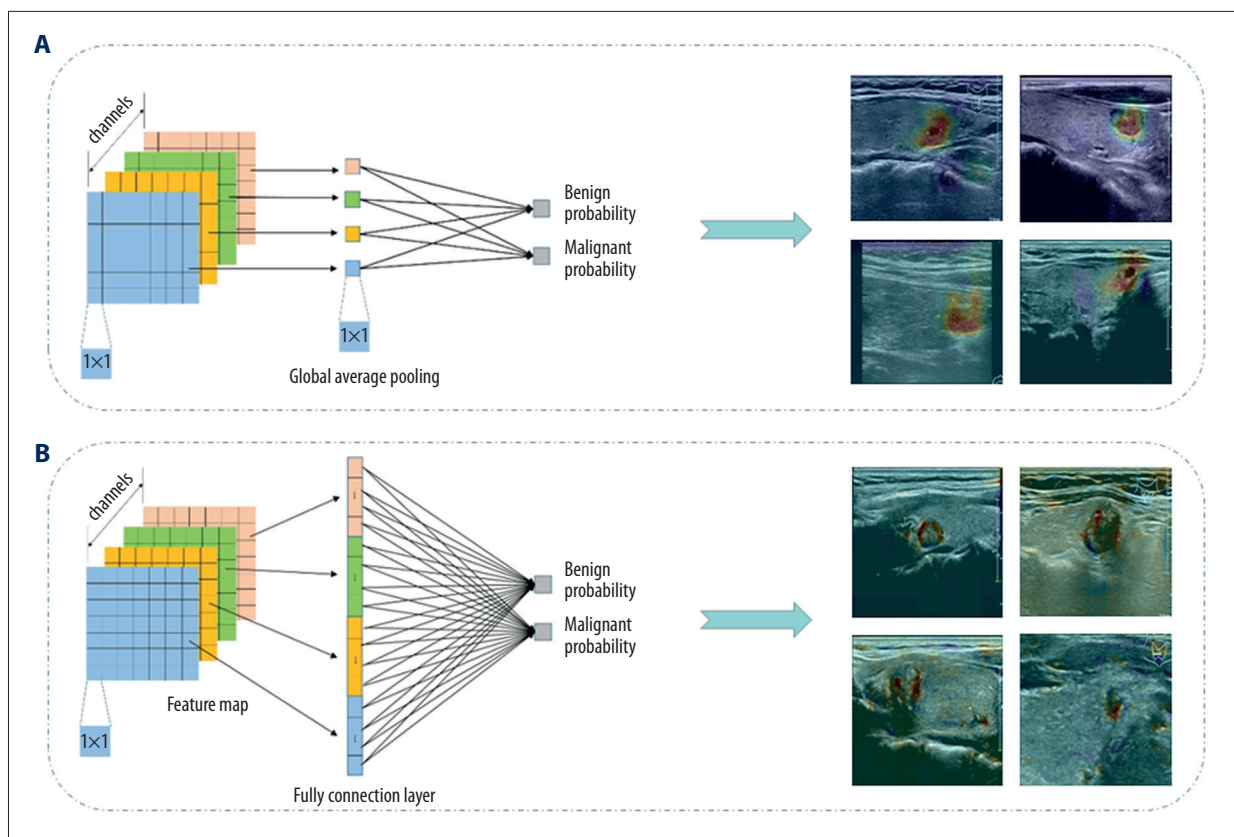


Figure 2. The boundary visualization model was used to localize nodule boundaries during the prediction of neural networks, with sketches of the global average pooling (A) and fully connected layer (B). The visualized reverse calculation result from global average pooling was rough. On the contrary, through the fully connection layer, we made full use of all the data from the feature map to reduce the loss of information, and obtained finer visualization results.

constructed. Second, we obtained the position set PB of benign nodules, where $PB_i = \{x, y - \frac{w}{2}, h, 2w\}$. Similarly, the position set of malignant modules was determined by $PM_i = \{x - \frac{h}{2}, y, 2h, w\}$. The extracted ROI area was expressed as $R = \{Crop(I)_{x',y',h',w'} | I_i \in I\}$. At the same time, the deformed dataset $\{B_d, M_d\} = \{Resize(R_i)_{224} | R_i \in R\}$ was constructed. Third, based on the datasets with the compositions shown in Figure 3, different classification models were trained and tested.

To demonstrate that the A/T ratio is an essential factor in diagnosing benign and malignant nodules and that the neural network used for assisted diagnosis can learn this important feature, the dataset used to train and test the assisted-diagnosis neural network was adjusted accordingly. First, as shown in Figure 3, the A/T ratio of benign nodules was reduced, while that of malignant nodules was increased, demonstrating the influence of the A/T ratio of a nodule. The 3 datasets were (1) DB_{M_o} , which was composed of the original benign and malignant nodule images; (2) DB_{dM_o} , which was composed of the deformed benign nodule images and the original malignant nodule images; and (3) DB_{M_d} , which was composed of the original benign nodule image and the deformed malignant nodule

image. Concurrently, for each dataset, 5 000 images were classified as the training set and 2 214 were used as the test set. Then, diagnosis models 1 through 5 were separately trained by the datasets TRB_{M_o} , TRB_{dM_o} , TRB_{dM_o} , TRB_{M_o} , and TRB_{M_d} based on our neural network. Then these 5 models were tested by TEB_{M_o} , TEB_{dM_o} , TEB_{M_o} , TEB_{M_d} , and TEB_{M_o} , respectively.

Statistical analysis

The diagnostic performance of the network was evaluated by calculating its accuracy, sensitivity, and specificity. The area under the receiver operating characteristic curve with a 95% confidence interval was calculated to evaluate the diagnostic performance of the A/T ratio for distinguishing between benign and malignant thyroid nodules. All statistical results were calculated using MedCalc for Windows v15.8 (MedCalc Software, Ostend, Belgium), and were considered statistically significant when P was less than 0.05.

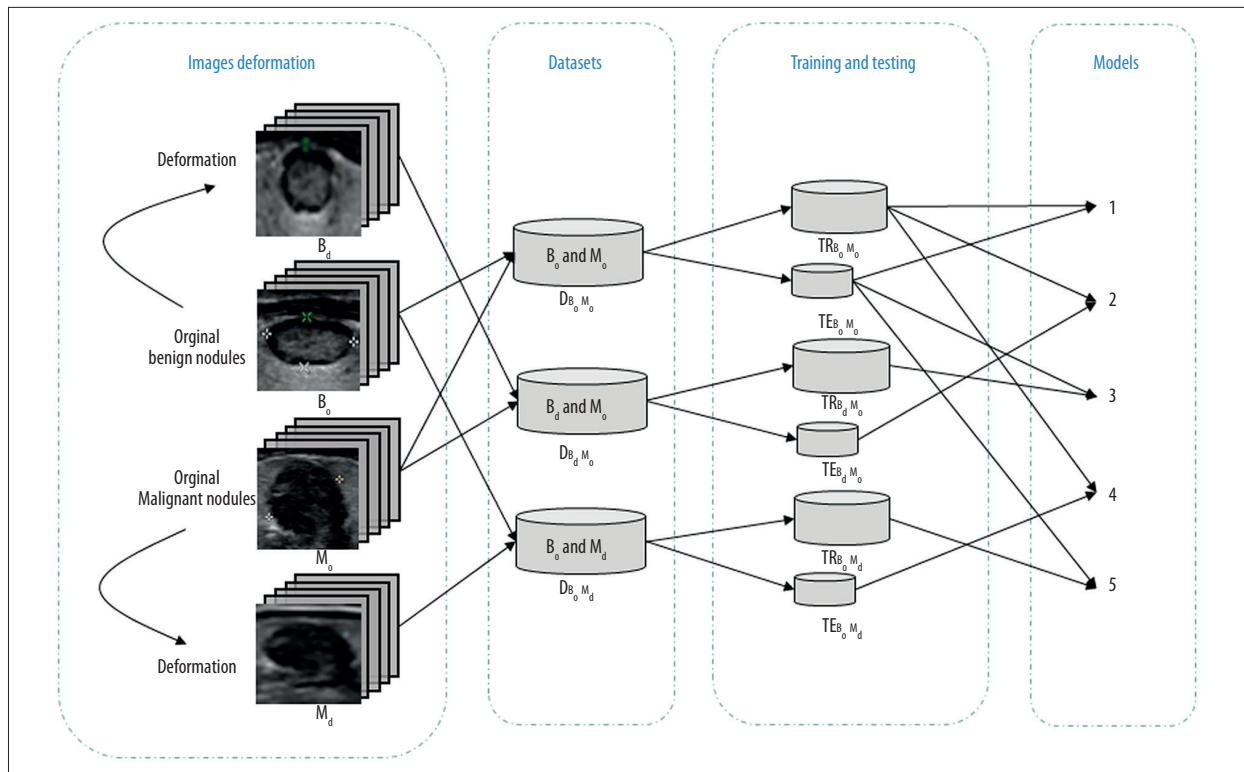


Figure 3. Composition of the dataset including original benign images, original malignant images and their deformed images. Each dataset is divided into a training set and a test set according to the A/T ratio. The models 1-5 were trained and tested. (B_o and M_o are the original image sets. B_d and M_d are the deformed image sets.)

Results

Demographic and pathological features of the patients

This study included 7 216 ultrasound images from 2 489 consecutive patients who underwent diagnostic thyroid ultrasound examination and subsequent surgery. Of the total images, 2 712 were from 1 021 benign nodules, which included nodular goiter, adenomatous goiter, thyroid granuloma, and follicular adenoma, and 4 504 images were from 1 468 malignant nodules, which included papillary thyroid carcinoma, medullary thyroid carcinoma, and follicular thyroid carcinoma. The demographic and pathological features of the patients are shown in Table 1.

Different types of images for the neural network structure training and visualization test models

Through neural network visualization (Figure 1), we found that part of the focus of the neural network trained with full images shifted to the surrounding “interference information”, rather than on nodules or internal organization, which could have greatly affected the generalizability of the model. In this way, the accuracy rate was dramatically decreased if different datasets were replaced; for example, changing the way the peripheral

Table 1. Demographic and pathological features of the patients.

| Dataset | |
|--------------------------|---------------|
| Patients | 2489 |
| Male, n (%) | 614 (24.67%) |
| Female, n (%) | 1875 (75.33%) |
| Age (years) | 45.29±13.45 |
| Malignant nodules | 1468 |
| PTC | 1442 (98.23%) |
| MTC | 15 (1.02%) |
| FTC | 11 (0.75%) |
| Benign nodules | 1021 |
| Nodular goiter | 703 (68.85%) |
| Adenomatous goiter | 142 (13.91%) |
| Thyroid granuloma | 129 (12.63%) |
| Follicular adenoma | 47 (4.61%) |
| Total images | 7216 |
| Malignant images, n (%) | 4504 (62.42%) |
| Benign images, n (%) | 2712 (37.58%) |

PTC – Papillary thyroid carcinoma; MTC – Medullary thyroid carcinoma; FTC – Follicular thyroid carcinoma.

Table 2. The performance of the model whose training sets are using different types of images.

| Experiment | Training set | Accuracy | Sensitivity | Specificity |
|------------|--------------------|----------|-------------|-------------|
| 1 | Full image | 0.8796 | 0.8899 | 0.8609 |
| 2 | Black bars removed | 0.8715 | 0.8480 | 0.8851 |
| 3 | ROI | 0.9315 | 0.9229 | 0.9362 |

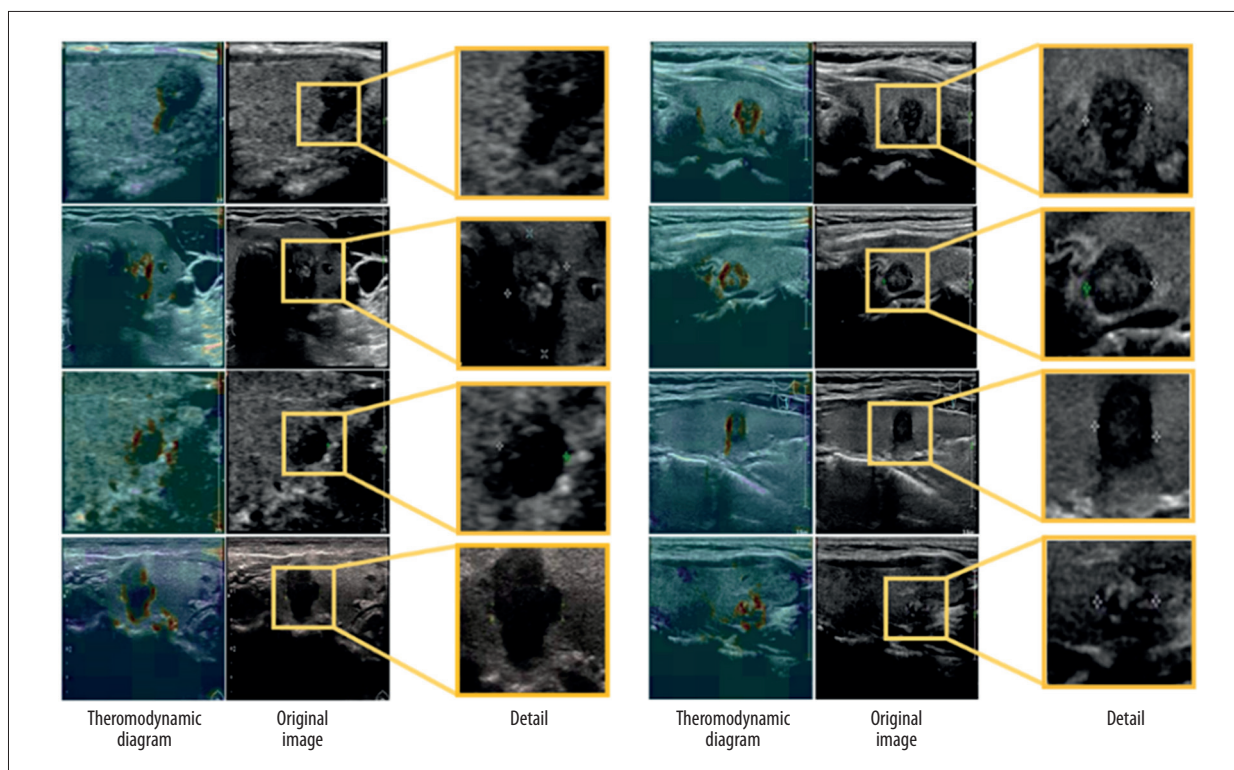


Figure 4. The display of the neural network visualization results. The first column is the visualization results, the second column is the original image, and the third column is an enlarged image of the most distinguishing area.

information was presented led to a change in the models. As shown in Table 2, the dataset of experiment 1 had an accuracy rate of 87.96% from using full images to train and test the model; whereas, the dataset of experiment 3, which was obtained by segmenting the ROI of the ultrasound images, performed best among the 3 models. The accuracy of the experiment 3 was 6% and 5.19% higher, respectively, than those of the other 2 experiments. As shown in Figure 1, the model was too focused in experiments 1 and 2 on the surroundings or multiple signs in the visualization result.

Experimental results of neural network precise visualization

According to the description of the Thyroid Imaging Reporting and Data System (TI-RADS), malignant nodules have the characteristics of unclear borders and irregular margins. With the

global average pooling layer, the visualized reverse calculation yielded a rough result. On the contrary, through a fully connected layer, we made full use of all the data of the feature map to reduce the loss of information and obtain finer visualization results. The results of the neural network visualization method on malignant nodules focused on the nodule boundary area, which means that the boundary information was also an important indicator in the diagnosis of malignances. Figure 4 shows the experimental results, including screenshots of the visualization results, the original images, and detailed images of the nodules. According to the experimental results, the neural network paid special attention to malignant thyroid nodules with irregular and unclear margins; that is, the red spot areas of the precise visualization results shown in Figure 4. The AI visual results indicated that irregular margin (red spot area of malignant nodules) was a crucial feature differentiated by the neural network.

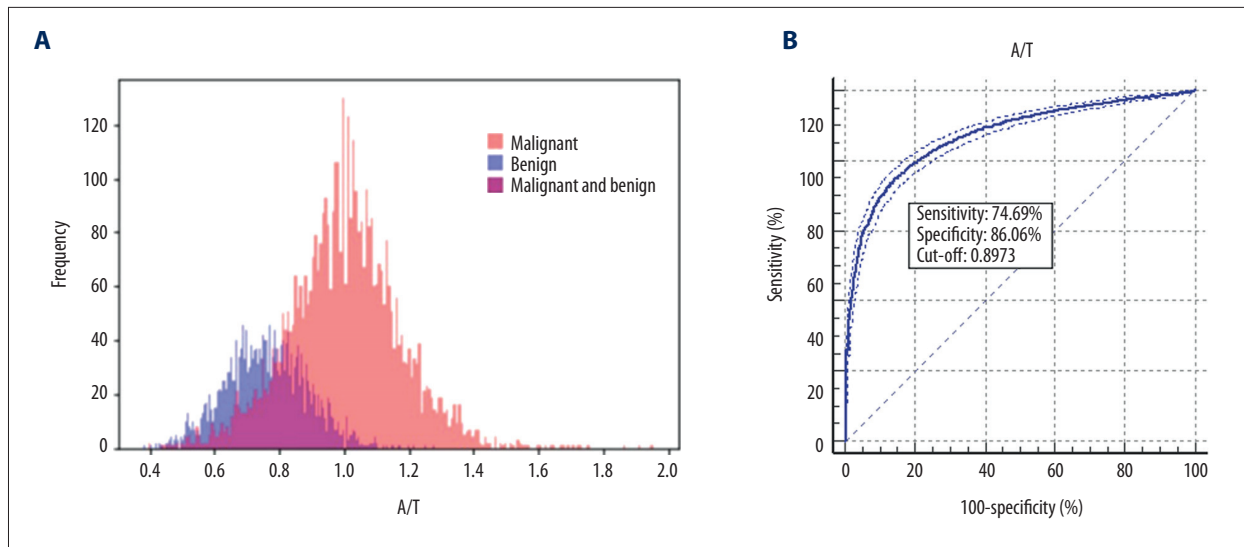


Figure 5. The plot of the A/T ratio experiments of thyroid nodules is shown in **A**: the pink area represents the distribution area of malignant nodules, the blue area represents the distribution area of benign nodules, and the purple area in the middle is the overlap area of both malignant and benign nodules. The receiver operating characteristic (ROC) curve of the A/T ratio in the diagnosis of thyroid nodules is demonstrated in **B**.

Table 3. The experimental results of changing the A/T ratio of benign or malignant nodules.

| Model | Training set | Test set | Accuracy | Sensitivity | Specificity |
|-------|---------------|---------------|----------|-------------|-------------|
| 1 | $TR_{B_o}M_o$ | $TE_{B_o}M_o$ | 0.9272 | 0.9492 | 0.8928 |
| 2 | $TR_{B_o}M_o$ | $TE_{B_d}M_o$ | 0.7662 | 0.9536 | 0.4697 |
| 3 | $TR_{B_d}M_o$ | $TE_{B_o}M_o$ | 0.6227 | 0.9868 | 0.0466 |
| 4 | $TR_{B_o}M_o$ | $TE_{B_o}M_d$ | 0.8348 | 0.8049 | 0.8823 |
| 5 | $TR_{B_o}M_d$ | $TE_{B_o}M_o$ | 0.8317 | 0.8056 | 0.873 |

Results of A/T ratio experiments of thyroid nodules

As shown in Supplementary Table 3, 96.30% of the nodules were malignant while only 3.70% (85/2 296) of the nodules were benign when the A/T ratio was ≥ 1 , indicating that the A/T ratio effectively distinguished malignant nodules. However, when the A/T ratio was < 1 , the differentiated benign and malignant nodules were 53.39% and 46.61%, respectively, in which case the benign nodules could not be obviously distinguished (Figure 5A). According to the receiver operating characteristic curve, the best cutoff value for the A/T ratio was 0.90 in predicting benign and malignant nodules, with an accuracy, sensitivity, and specificity of 0.872, 74.69%, and 86.06%, respectively (Figure 5B).

The results of models 1–5 are shown in Table 3. The data show that the accuracy of the diagnosis of benign and malignant thyroid nodules reached up to 92.72% in model 1. Models 2–4 were tested to verify the sensitivity and specificity of the A/T

ratios of nodules. When the A/T ratio was changed in the training set or testing set, the diagnostic accuracy of the models decreased by 9.24–30.45% (Table 3, Figure 6). The specificity of models 2 and 3 decreased significantly, and the sensitivity and specificity decreased in all models.

Discussion

At present, there are many studies on and applications of medically aided diagnoses with neural networks, and a high classification accuracy rate has been achieved [27–31]. However, the neural network is essentially a black box, which means that its users do not know which features in the images are important to the neural network in making a prediction [32]. Thus, they cannot give a reasonable explanation regarding the reliability of the prediction [33,34]. In this study, by designing comparative experiments and adopting neural network visualization methods, we showed that the basis of neural network

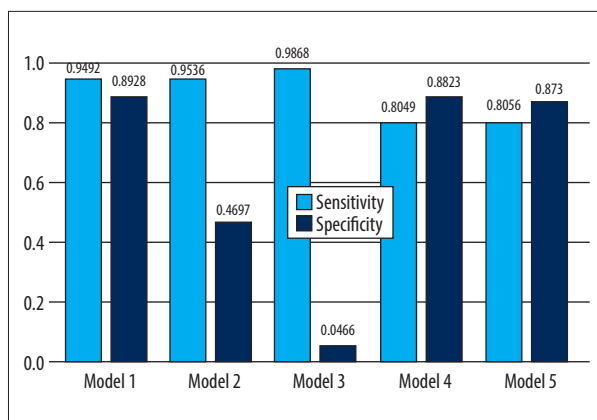


Figure 6. Comparison of the sensitivity and specificity in Models 1–5.

diagnosis of thyroid nodules includes the A/T ratio and margin information. The results showed that an A/T ratio of 0.9 was the best cutoff value for differentiating between benign and malignant nodules. The precise AI visualization of ultrasound made features the successful localization of irregular margins of nodules, indicating the margin feature is an important aspect of neural networks within the black box operation.

The diagnostic accuracy of deep learning can currently reach more than 92%, which is better than the performance of doctors [35–37]. This study found that the datasets used in some studies on deep learning in medical image-assisted diagnosis are based on full images as inputs, which may result in the model learning interference information, leading to poor robustness of the deep learning model without good generalizability among experiments. Therefore, it is recommended to extract the ROI or remove boundary interference information by segmenting ultrasound images before training. In this study, it was proposed that the ROI should be segmented first so that more in-depth research could be carried out with good results when using thyroid ultrasound images for auxiliary medical research.

In 2019, deep learning algorithms achieved a high accuracy in the visual localization of lung disease by lung ultrasound, hip fractures by pelvic radiograph, and preclinical pulmonary fibrosis identification by computerized tomography scan [38–40].

Several experiments in our study were used to visualize the neural network, which can intuitively show the nodular feature area with a higher weight within the neural network. In the medical imaging field, margins, calcifications, and the A/T ratios of thyroid nodules are significant factors in the diagnosis of thyroid cancer. A study by Moon et al. [23] showed that an A/T ratio of ≥ 1 is the best predictor of malignant tumors. Yoon et al. [41] concluded that the feature of a taller-than-wide shape depends on the compressibility of a malignant thyroid nodule on transversal ultrasound images. According to our statistics, the best cutoff value of the A/T ratio to distinguish between benign and malignant nodules was approximately 0.90, which may be related to the inconsistency of the pathological types of the collected nodules in our research.

There were 2 limitations in our study. First, the study used only margins and the A/T ratio in neural network training. Further studies are needed to explore which aspects the neural network has learned in addition to the margin and A/T ratio, including other ultrasound features of thyroid nodules, such as echogenicity, composition, and calcification. In addition, further studies are needed to determine whether neural networks can learn microscopic features at the pixel-level that humans cannot observe. Second, the pathological types of nodules included in our study were limited. Different pathological types of nodules should be used in AI-model training for the visualization of generalized performance.

Conclusions

In this study, the A/T ratio and margin information of thyroid nodules were crucial criteria for distinguishing between malignant and benign nodules. These criteria are also important factors in the auxiliary diagnosis of neural network models, which have significant value in medicine. Further, according to our visualization experiment results, we conclude there is risk in using the entire ultrasound image to train the neural network; therefore, we suggest segmenting images first.

Conflict of interest

None.

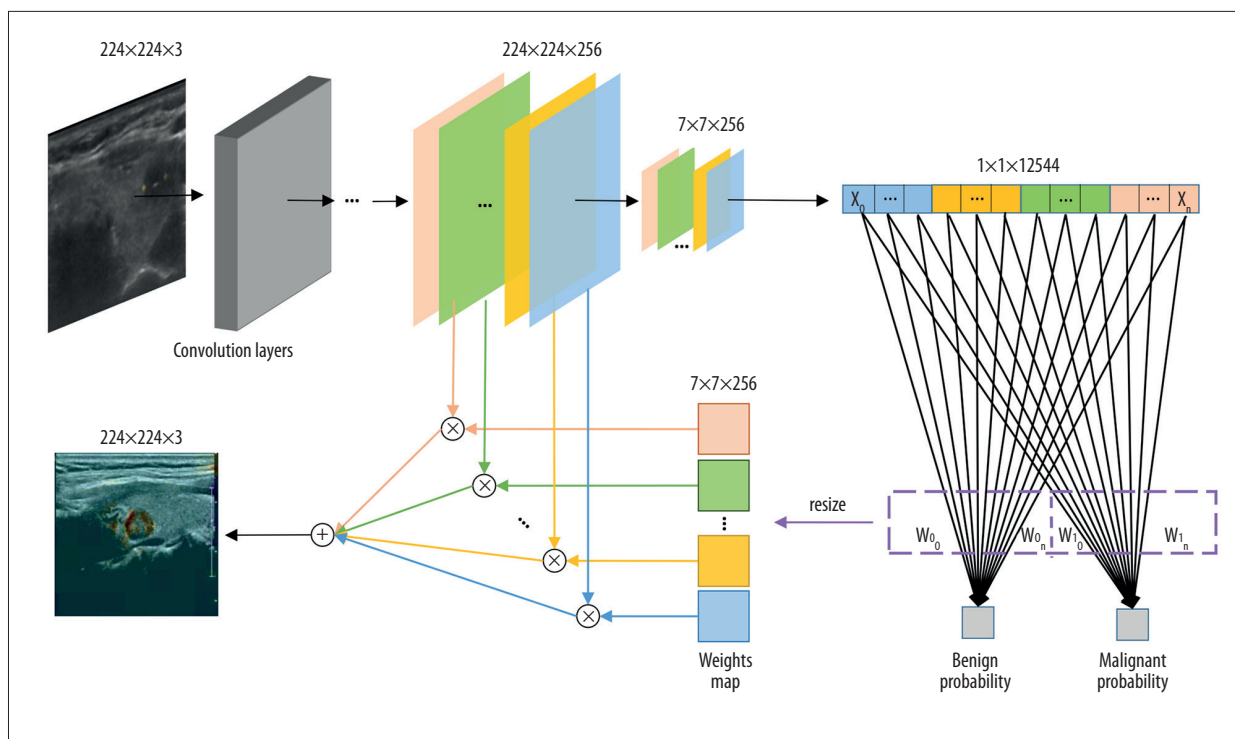
Supplementary Data

Supplementary Table 1. The structure of the visualization model.

| Layer name | Output size | Activation |
|------------------|-------------|------------|
| conv1_1 | 224×224×64 | Relu |
| conv1_2 | 224×224×64 | Relu |
| pool1 | 112×112×64 | |
| conv2_1 | 112×112×128 | Relu |
| conv2_2 | 112×112×128 | Relu |
| pool2 | 56×56×128 | |
| conv3_1 | 56×56×256 | Relu |
| conv3_2 | 56×56×256 | Relu |
| conv3_3 | 56×56×256 | Relu |
| conv3_4 | 56×56×256 | Relu |
| pool3 | 28×28×256 | |
| conv4_1 | 28×28×512 | Relu |
| conv4_2 | 28×28×512 | Relu |
| conv4_3 | 28×28×512 | Relu |
| conv4_4 | 28×28×512 | Relu |
| pool4 | 7×7×512 | |
| Fully connection | 25088 | |
| Softmax | 2 | |

Supplementary Table 2. The structure of the model in A/T ratio experiment.

| Layer name | Output size | Activation |
|--------------------|-------------|------------|
| conv1_1 | 224×224×64 | Relu |
| conv1_2 | 224×224×64 | Relu |
| pool1 | 112×112×64 | |
| conv2_1 | 112×112×128 | Relu |
| conv2_2 | 112×112×128 | Relu |
| pool2 | 56×56×128 | |
| conv3_1 | 56×56×256 | Relu |
| conv3_2 | 56×56×256 | Relu |
| conv3_3 | 56×56×256 | Relu |
| conv3_4 | 56×56×256 | Relu |
| pool3 | 28×28×256 | |
| conv4_1 | 28×28×512 | Relu |
| conv4_2 | 28×28×512 | Relu |
| conv4_3 | 28×28×512 | Relu |
| conv4_4 | 28×28×512 | Relu |
| pool4 | 14×14×512 | |
| conv5_1 | 14×14×512 | Relu |
| conv5_2 | 14×14×512 | Relu |
| conv5_3 | 14×14×512 | Relu |
| conv5_4 | 14×14×512 | Relu |
| pool5 | 7×7×512 | |
| Fully connection 1 | 25088 | |
| Fully connection 2 | 25088 | |
| Fully connection 3 | 2 | |
| Softmax | 2 | |



Supplementary Figure 1. Sketches of visualization AI method. The input of the model was the thyroid ultrasound image after data preprocessing, and then the fully connected operation was performed after the feature extraction through the convolution neural network. During training, the features that were useful for classification were gradually strengthened. Therefore, by inversely calculating the weights of the fully connected layers and the feature maps, it was possible to visualize the areas in the images which had a significant effect on the classification task.

Supplementary Table 3. A/T ratio statistics of the dataset.

| Range of A/T | Malignant | Benign | Total |
|--------------|----------------|----------------|-------|
| <1 | 2 627 (53.39%) | 2 293 (46.61%) | 4 920 |
| 0.3–0.4 | 2 | 1 | |
| 0.4–0.5 | 10 | 38 | |
| 0.5–0.6 | 53 | 220 | |
| 0.6–0.7 | 135 | 605 | |
| 0.7–0.8 | 311 | 817 | |
| 0.8–0.9 | 656 | 659 | |
| 0.9–1.0 | 1 126 | 287 | |
| ≥1 | 2 211 (96.30%) | 85 (3.70%) | 2 296 |
| 1.0–1.1 | 1 066 | 70 | |
| 1.1–1.2 | 648 | 12 | |
| 1.2–1.3 | 303 | 3 | |
| 1.3–1.4 | 134 | 0 | |
| 1.4–1.5 | 34 | 0 | |
| ≥1.5 | 26 | 0 | |

References:

1. Tessler FN, Middleton WD, Grant EG et al: ACR thyroid imaging, reporting and data system (TI-RADS): White paper of the ACR TI-RADS Committee. *J Am Coll Radiol*, 2017; 14: 587–95
2. Rajaraman S, Antani S: Training deep learning algorithms with weakly labeled pneumonia chest X-ray data for COVID-19 detection. medRxiv. 2020; 20090803
3. Li X, Zhang S, Zhang Q et al: Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: A retrospective, multicohort, diagnostic study. *Lancet Oncol*, 2019; 20: 193–201
4. Naceur MB, Saouli R, Akil M, Kachouri R: Fully automatic brain tumor segmentation using end-to-end incremental deep neural networks in MRI images. *Comput Methods Programs Biomed*, 2018; 166: 39–49
5. Shaffer K: Deep learning and lung cancer: AI to extract information hidden in routine CT scans. *Radiology*, 2020; 296: 225–26
6. Lim KJ, Choi CS, Yoon DY et al: Computer-aided diagnosis for the differentiation of malignant from benign thyroid nodules on ultrasonography. *Acad Radiol*, 2008; 15: 853–58
7. Ma J, Wu F, Jiang T et al: Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks. *Int J Comput Assist Radiol Surg*, 2017; 12: 1895–910
8. Chang CY, Lei YF, Tseng CH, Shih SR: Thyroid segmentation and volume estimation in ultrasound images. *IEEE Trans Biomed Eng*, 2010; 57: 1348–57
9. Zhang X, Jiang L, Yang D et al: Urine sediment recognition method based on multi-view deep residual learning in microscopic image. *J Med Syst*, 2019; 43: 325
10. Angrick M, Herff C, Mugler E et al: Speech synthesis from ECoG using densely connected 3D convolutional neural networks. *J Neural Eng*, 2019; 16: 036019
11. Tajbaksh N, Shin JY, Gurudu SR et al: Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Trans Med Imaging*, 2016; 35: 1299–312
12. Chen J, You H, Li K: A review of thyroid gland segmentation and thyroid nodule segmentation methods for medical ultrasound images. *Comput Methods Programs Biomed*, 2020; 185: 105329
13. Huang W: Segmentation and diagnosis of papillary thyroid carcinomas based on generalized clustering algorithm in ultrasound elastography. *J Med Syst*, 2019; 44: 13
14. Koundal D, Sharma B, Guo Y: Intuitionistic based segmentation of thyroid nodules in ultrasound images. *Comput Biol Med*, 2020; 121: 103776
15. Koundal D: Computer-aided diagnosis of thyroid nodule: A review. *Int J Comput Sci Eng Survey (IJCSSES)*, 2012; 3: 67–83
16. Koundal D, Gupta S, Singh S: Automated delineation of thyroid nodules in ultrasound images using spatial neutrosophic clustering and level set. *Appl Soft Comput*, 2016; 40: 86–97
17. Koundal D, Gupta S, Singh S: Computer aided thyroid nodule detection system using medical ultrasound images. *Biomed Signal Proces*, 2018; 40: 117–30
18. Buda M, Wildman-Tobriner B, Hoang JK et al: Management of thyroid nodules seen on US images: Deep learning may match performance of radiologists. *Radiology*, 2019; 292: 695–701
19. Zech JR, Badgeley MA, Liu M et al: Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Med*, 2018; 15: e1002683
20. Walsh SLF, Calandriello L, Silva M, Sverzellati N: Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: A case-cohort study. *Lancet Respir Med*, 2018; 6: 837–45
21. Ghosh A, Kandasamy D: Interpretable artificial intelligence: Why and when. *Am J Roentgenol*, 2020; 214: 1137–38
22. Lee JH, Baek JH, Kim JH et al: Deep learning-based computer-aided diagnosis system for localization and diagnosis of metastatic lymph nodes on ultrasound: A pilot study. *Thyroid*, 2018; 28: 1332–38
23. Moon HJ, Kwak JY, Kim EK, Kim MJ: A taller-than-wide shape in thyroid nodules in transverse and longitudinal ultrasonographic planes and the prediction of malignancy. *Thyroid*, 2011; 21: 1249–53
24. Russ G, Bonnema SJ, Erdogan MF et al: European Thyroid Association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: The EU-TIRADS. *Eur Thyroid J*, 2017; 6: 225–37
25. Shin JH, Baek JH, Chung J et al: Ultrasonography diagnosis and imaging-based management of thyroid nodules: Revised Korean Society of Thyroid Radiology Consensus Statement and Recommendations. *Korean J Radiol*, 2016; 17: 370–95
26. Gharib H, Papini E, Garber JR et al: American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules – 2016 Update. *Endocr Pract*, 2016; 22: 622–39
27. Choi YJ, Baek JH, Park HS et al: A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of thyroid nodules on ultrasound: Initial clinical assessment. *Thyroid*, 2017; 27: 546–52
28. Yoo YJ, Ha EJ, Cho YJ et al: Computer-aided diagnosis of thyroid nodules via ultrasonography: Initial clinical experience. *Korean J Radiol*, 2018; 19: 665–72
29. Jeong EY, Kim HL, Ha EJ et al: Computer-aided diagnosis system for thyroid nodules on ultrasonography: Diagnostic performance and reproducibility based on the experience level of operators. *Eur Radiol*, 2019; 29: 1978–85
30. Reverter JL, Vazquez F, Puig-Domingo M: Diagnostic performance evaluation of a computer-assisted imaging analysis system for ultrasound risk stratification of thyroid nodules. *Am J Roentgenol*, 2019; 11: 1–6
31. Wang L, Yang S, Yang S et al: Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network. *World J Surg Oncol*, 2019; 17: 12
32. Liberati EG, Ruggiero F, Galuppo L et al: What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. *Implement Sci*, 2017; 12: 113
33. Holzinger A, Langs G, Denk H et al: Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov*, 2019; 9: e1312
34. Moxey A, Robertson J, Newby D et al: Computerized clinical decision support for prescribing: Provision does not guarantee uptake. *J Am Med Inform Assoc*, 2010; 17: 25–33
35. Zhang B, Tian J, Pei S et al: Machine learning-assisted system for thyroid nodule diagnosis. *Thyroid*, 2019; 29: 858–67
36. Wei X, Gao M, Yu R et al: Ensemble deep learning model for multicenter classification of thyroid nodules on ultrasound images. *Med Sci Monit*, 2020; 26: e926096
37. Chambara N, Ying M: The diagnostic efficiency of ultrasound computer-aided diagnosis in differentiating thyroid nodules: A systematic review and narrative synthesis. *Cancers (Basel)*, 2019; 11: 1759
38. van Sloun RJG, Demi L: Localizing B-Lines in lung ultrasonography by weakly supervised deep learning, *in-vivo* results. *IEEE J Biomed Health Inform*, 2020; 24: 957–64
39. Cheng CT, Ho TY, Lee TY et al: Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs. *Eur Radiol*, 2019; 29: 5469–77
40. Mathai SK, Humphries S, Kropski JA et al: MUC5B variant is associated with visually and quantitatively detected preclinical pulmonary fibrosis. *Thorax*, 2019; 74: 1131–39
41. Yoon SJ, Yoon DY, Chang SK et al: “Taller-than-wide sign” of thyroid malignancy: Comparison between ultrasound and CT. *Am J Roentgenol*, 2010; 194: W420–24