Leveraging deep learning to identify calcification and colloid in thyroid nodules

Chen Chen [a,b,e,f,1], Yuanzhen Liu [b,e,f,1], Jincao Yao [b,c,d], Lujiao Lv [c,d], Qianmeng Pan [f,e], Jinxin Wu [f,e], Changfu Zheng [f,e], Hui Wang [f,e], Xianping Jiang [g], Yifan Wang [b,c,d,e,f,**], Dong Xu [b,c,d,e,f,*]

[a] Graduate School, Wannan Medical College, Wuhu, 241002, China
[b] Department of Diagnostic Ultrasound Imaging & Interventional Therapy, Zhejiang Cancer Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou, 310022, China
[c] Key Laboratory of Head & Neck Cancer Translational Research of Zhejiang Province, Hangzhou, 310022, China
[d] Zhejiang Provincial Research Center for Cancer Intelligent Diagnosis and Molecular Technology, Hangzhou, 310022, China
[e] Taizhou Key Laboratory of Minimally Invasive Interventional Therapy & Artificial Intelligence, China
[f] Taizhou Cancer Hospital, Taizhou, 317502, China
[g] Department of Ultrasound, Shengzhou People's Hospital (the First Affiliated Hospital of Zhejiang University Shengzhou Branch), Shengzhou, 312400, China

ARTICLE INFO

ABSTRACT

Background: Both calcification and colloid in thyroid nodules are reflected as echogenic foci in ultrasound images. However, calcification and colloid have significantly different probabilities of malignancy. We explored the performance of a deep learning (DL) model in distinguishing the echogenic foci of thyroid nodules as calcification or colloid.
Methods: We conducted a retrospective study using ultrasound image sets. The DL model was trained and tested on 30,388 images of 1127 nodules. All nodules were pathologically confirmed. The area under the receiver-operator characteristic curve (AUC) was employed as the primary evaluation index.
Results: The YoloV5 (You Only Look Once Version 5) transfer learning model for thyroid nodules based on DL detection showed that the average sensitivity, specificity, and accuracy of distinguishing echogenic foci in the test 1 group (n = 192) was 78.41%, 91.36%, and 77.81%, respectively. The average sensitivity, specificity, and accuracy of the three radiologists were 51.14%, 82.58%, and 61.29%, respectively. The average sensitivity, specificity, and accuracy of distinguishing small echogenic foci in the test 2 group (n = 58) was 70.17%, 77.14%, and 73.33%, respectively. Correspondingly, the average sensitivity, specificity, and accuracy of the radiologists were 57.69%, 63.29%, and 59.38%.
Conclusions: The study demonstrated that DL performed far better than radiologists in distinguishing echogenic foci of thyroid nodules as calcifications or colloid.

* Corresponding author. No.1 East Banshan Road, Gongshu District, Hangzhou 310022, China
** Corresponding author. No.1 East Banshan Road, Gongshu District, Hangzhou 310022, China
E-mail addresses: cc929202@163.com (C. Chen), yuanzhen_0128@aliyun.com (Y. Liu), yaojc@zjcc.org.cn (J. Yao), yoalkg@163.com (L. Lv), 13357691918@163.com (Q. Pan), wjx77888@163.com (J. Wu), 13958675431@139.com (C. Zheng), 15258673285@139.com (H. Wang), jiangxianping2023@163.com (X. Jiang), wangyf@zjcc.org.cn (Y. Wang), xudong@zjcc.org.cn (D. Xu).
[1] These authors contributed equally to this work.

1. Introduction

In recent years, the detection rate of thyroid nodules has increased and can be as high as 50% or more in the population [1–4]. Of nodules detected by touch or ultrasound, 5–15% are malignant [5,6]. Therefore, the identification of benign and malignant thyroid nodules is very important and determines the treatment plan for the patient [7]. According to the American Thyroid Association guidelines, ultrasound is highly recommended for patients with suspected thyroid nodules [8]. On ultrasound, some features of thyroid nodules, such as calcification, hypoechogenicity, and being taller than wide, are highly correlated with malignancy [9–11]. Calcification is common in both benign and malignant nodules, but a higher percentage of thyroid cancer nodules are combined with calcification compared to benign disease [12,13].

Calcifications generally appear as echogenic foci on ultrasound images. However, nodules with echogenic foci on ultrasound images are often diagnosed as colloid on pathological examination [14]. In the existing literature, echogenic foci on ultrasound images are often poorly defined and are simply referred to as calcifications [15]. Therefore, differentiating between echogenic foci with calcification and colloid pathological results is particularly important. Experienced radiologists can differentiate this based on the characteristics of the ultrasound images. Microcalcifications or psammoma bodies appear as echogenic foci on ultrasound and are not accompanied by acoustic shadowing. Echogenic foci with comet-tail artifacts may be caused by colloid aggregation [16]. A large amount of colloid deposition is more suggestive of a benign nodule [17]. However, some studies have shown that malignant nodules also present echogenic foci with comet-tail artifacts [15,18,19]. Therefore, a certain probability of misjudgment exists if the echogenic focal component is judged subjectively based on the experience of the radiologist alone. Many recent studies have shown that combining medical imaging and artificial intelligence (AI) to build relevant assisted diagnostic models can help radiologists make diagnoses [20,21]. For example, an AI model developed by Chen et al. [22] predicted benign and malignant thyroid nodules with a higher area under the receiver-operator characteristic curve (AUC) and sensitivity than junior physicians and a higher specificity than senior physicians. Yao et al. [23] developed a multimodal deep learning model to predict cervical lymph node metastasis of papillary thyroid cancer from ultrasound images, which has an AUC value of more than 0.87 and can provide a basis for selecting treatment options. Zhao et al. [24] investigated the efficacy of deep learning models in diagnosing Hashimoto's thyroiditis. Therefore, based on these studies, it is reasonable to believe that AI can distinguish echogenic foci in ultrasound images of thyroid nodules that are calcification or colloid by pathological histology. To our knowledge, existing studies have never used AI models to differentiate calcification and colloid in thyroid nodules. The use of such methods to differentiate calcification and colloid could help in the determination of benign and malignant thyroid nodules.

In this study, a deep learning model was developed based on ultrasound images. The efficacy of the model in distinguishing echogenic foci of thyroid nodules as calcification or colloid was investigated and compared with the judgment of radiologists.

2. Materials and methods

2.1. Data source

This retrospective study was approved by the Ethics Committee of Zhejiang Cancer Hospital, and informed consent was waived (IRB-2020-287). All images and data were anonymized. The image acquisition was performed by five radiologists from Zhejiang Cancer Hospital from March 2021 to August 2022. Thyroid ultrasound images were derived from ultrasound devices manufactured by companies such as Siemens, Toshiba, and Philips. The patients were examined in the supine position with full exposure of the neck. The images were reviewed by three highly qualified radiologists, each with more than ten years of experience in the thyroid field.

The inclusion criteria for thyroid nodules in the training dataset, validation dataset and test 1 dataset were:

1) Patients had clear preoperative ultrasound images of thyroid examination.
2) The pathological results of thyroid nodules were with colloid and/or calcification.
3) Each patient had complete clinical information.

The exclusion criteria for thyroid nodules were:

1) The patient's preoperative thyroid examination ultrasound image was incomplete or substandard.
2) The patient had undergone two or more neck surgeries.
3) The patient's pathological results were unclear.
4) The patient's clinical information was incomplete.

These three datasets included 1069 nodules with a total of 3496 ultrasound images. All nodules had corresponding postoperative pathological results. Thyroid nodules with colloid only were classified as the colloid group, and those with calcification were classified as the calcification group, according to the pathological results.

Thyroid nodules in the test 2 dataset were selected based on the following inclusion criteria: 1) Patients with clear preoperative ultrasound images displaying small echogenic foci within the thyroid nodules; 2) Pathological confirmation of the thyroid nodules; 3) Each patient had complete clinical information. The exclusion criteria remained consistent with the previous three datasets. One senior radiologist made the diagnosis of "microcalcification" or "colloid" based on the available pathological findings, which was then

reviewed by another senior radiologist. In case of consensus, the diagnosis was used as the gold standard. In case of disagreement, the diagnosis of a third senior radiologist served as the gold standard. Ultimately, a collection of 108 ultrasound images from 58 nodules was obtained in the test 2 dataset.

2.2. Data preprocess

All thyroid data were converted from the original DICOM to high-quality JPG. Considering the rarity of medical data and the sparseness of features, data augmentation operations such as splicing, optical change, geometric change, rotation, flip, scaling, cropping, translation, and jitter were performed to expand the data and improve the generalization ability of the deep learning model [25,26]. After removing the images to test the generalization ability of the final model, the thyroid data were divided into a training dataset (80%) and a validation dataset (20%). The image size was adjusted to 640x640 pixels, and the image normalization operation was performed to use it for the training of the object detection model.

2.3. Model building

The block diagram of the overall process of this study is shown in Fig. 1. In our study, the YoloV5 object detection model was used to establish the model that detects calcification and colloid, and the model automatically learned the feature difference between calcification and colloid from shallow to deep from the input dataset. This study compared the detection efficiency of five models of YoloV5 series (YoloV5 is the improved object detection model of YoloV4 [27]) on thyroid echogenic foci nodules. The five detection models are all composed of Input, Backbone, Neck, and Head. The Backbone part extracts feature information from two types of echogenic foci nodule data and sends it to the next layer. The difference between the five models is that the feature extraction part of Backbone is different in depth. The Neck part upsamples and downsamples the features extracted by Backbone and performs feature fusion, and the obtained features are more discriminative. The Head end judges the target area in the image to obtain the final target area and category.

The five object detection models were named YoloV5n, YoloV5s, YoloV5m, YoloV5l, and YoloV5x, and the model in this study was trained using the augmented thyroid data. Two models had initialized weights using the COCO dataset of this task to transfer learning, the optimal weighting parameters using stochastic gradient descent optimizer (SGD), bounding box return loss (Loss_GIoU), classification loss (Loss_Cls), and loss of confidence (Loss_Obj) together. The error update of the model in training is completed as a comprehensive loss function, as shown in Equations (1)–(3). A total of 300 epochs were trained iteratively, with the initial learning rate of the SGD optimizer as 0.01, the cyclic learning rate as 0.002, and the weight decay coefficient as 0.0005. In the process of model training, when the performance index on the training dataset and the validation dataset was no longer improved, the learning rate was adjusted according to the weight decay coefficient until the iteration was completed. This model was trained using a computer with a GeForce GTX 3060Ti graphics processor and a Core I9–12900 KF central processing unit.

$$Loss\_GIoU = \sum_{i=0}^{S^2} \sum_{j=0}^{D} \frac{|A \cap B|}{|A \cup B|} - \frac{|C - (A \cup B)|}{|C|} \tag{1}$$
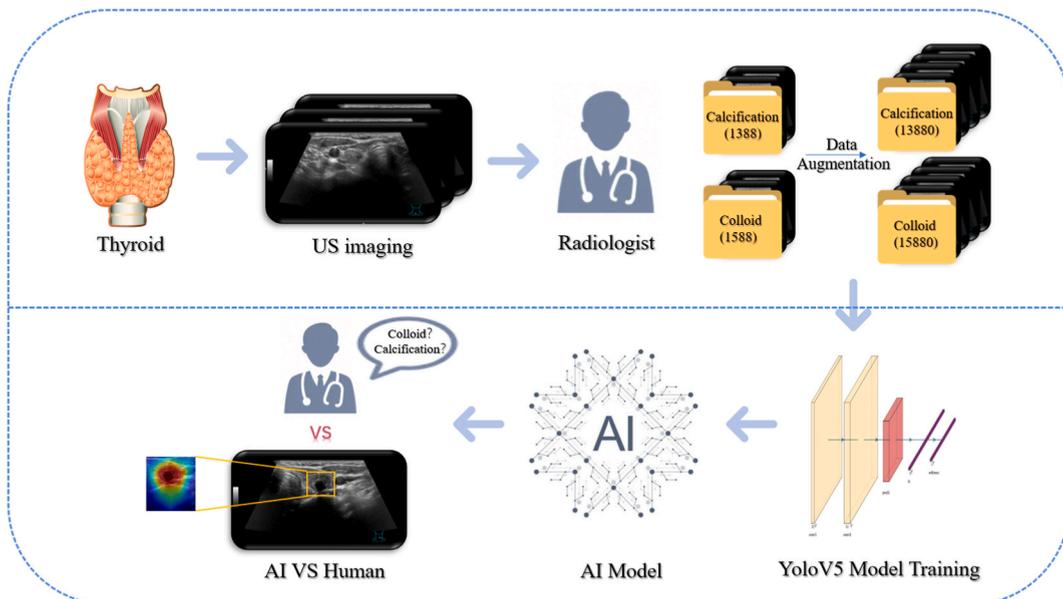

Fig. 1. Overall schematic for differentiating echogenic foci as calcification or colloid in thyroid nodules using a deep learning model.

$$Loss\_Cls = \sum_{i=0}^{S^2} \sum_{j=0}^{D} \sum_{c \subset cls} I_{ij}^{obj} \left[-\hat{p}_i(c)\ln(p_i(c)) - \left(1 - \hat{p}_i(c)\right)\ln(1 - p_i(c)) \right] \tag{2}$$

$$Loss\_Obj = \lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^{D} I_{ij}^{obj} \left[-\hat{C}_i \ln C_i - \left(1 - \hat{C}_i\right)\ln(1 - C_i) \right] + \lambda_{nobj} \sum_{i=0}^{S^2} \sum_{j=0}^{D} I_{ij}^{nobj} \left[-\hat{C}_i \ln C_i - \left(1 - \hat{C}_i\right)\ln(1 - C_i) \right] \tag{3}$$

* where S^2 is the number of divided grids, $\sum_{i=0}^{S^2} \sum_{j=0}^{D} \frac{|A \cap B|}{|A \cup B|} - \frac{|C - (A \cup B)|}{|C|}$ is the number of predicted target frames in each grid, $\sum_{i=0}^{S^2} \sum_{j=0}^{D} \sum_{c \subset cls} I_{ij}^{obj}[-\hat{p}_i(c)\ln(p_i(c)) - (1 - \hat{p}_i(c))\ln(1 - p_i(c))]$ is the area pixel of the real region, $\lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^{D} I_{ij}^{obj}[-\hat{C}_i \ln C_i - (1 - \hat{C}_i)\ln(1 - C_i)] + \lambda_{nobj} \sum_{i=0}^{S^2} \sum_{j=0}^{D} I_{ij}^{nobj}[-\hat{C}_i \ln C_i - (1 - \hat{C}_i)\ln(1 - C_i)]$ is the area pixel of the region predicted by the model, S^2 is the area pixel of the smallest outer rectangle of D and A, B is the judgment whether the C th bounding box of the A th grid has a target that needs to be predicted, B is whether there is a target object that does not need to be detected and predicted in the I_{ij}^{obj} boundary frame of the j th grid, i and I_{ij}^{nobj} are the weight parameters of whether there is a target in the grid, j and i are the confidence values of the predicted and actual targets, λ_{obj} is the class of targets predicted by the bounding box, λ_{nobj} is the probability size of belonging to C_i when the target is detected by the $\hat{C}_i$ th grid, and c is the actual probability size of being $p_i(c)$ when the target is detected by the c th network.

2.4. Performance evaluation and statistical analysis

Quantitative data are expressed as mean ± standard deviation. The object detection model for this study was built using the deep learning framework Pytorch 1.9.0, based on the Python 3.8 language. The AUC was used as the main evaluation index of the test dataset and compared for sensitivity (Se), specificity (Sp), Yordon index (YI), accuracy (Acc), positive predictive value (PPV), negative predictive value (NPV), and F1-score (F1) [28,29]. Python 3.8 and SciKit-Learn were used to plot receiver-operating characteristic curves and calculate model performance indicators. The consistency between the model results and the true results was measured using the Kappa coefficient.

3. Results

3.1. Baseline characteristics

A total of 1127 nodules with echogenic foci in the thyroid from 1013 patients were used in this study. There were 664 cases of malignant nodules and 463 cases of benign nodules. All nodules were confirmed by histopathology. Of these, 877 nodules were used for training the model. The test 1 dataset included 192 nodules and the test 2 dataset included 58 nodules. The mean age of the patients in the data used for model training was 50.0 ± 11.8 years, with 177 males and 600 females. The mean age of the patients in the test 1 dataset was 49.6 ± 12.3 years, with 38 males and 140 females. The mean age of the patients in the test 2 dataset was 49.8 ± 13.3 years, with 17 males and 41 females. Table 1 shows the basic information about the subjects and the distribution of the four datasets.

3.2. Comparison between the model and radiologists

The YoloV5 series models outperformed the radiologists. The YoloV5 series model had a mean AUC of 0.8241, a mean sensitivity of 0.7841, and a mean specificity of 0.9136 for identifying echogenic foci in thyroid nodules. Among them, the YoloV5x model had the best generalization ability for the test 1 dataset with an AUC of 0.9135, a sensitivity of 0.8750, and a specificity of 0.9318. The mean AUC for radiologists to identify echogenic foci in thyroid nodules was 0.6050, the mean sensitivity was 0.5114, and the mean specificity was 0.8258 (Table 2, Fig. 2A). Thus, the YoloV5 series model identified echogenic foci of thyroid nodules with superior performance to that of radiologists.

Table 1

Baseline characteristics of study participants and number of cases and images in the four datasets used in our study.

Final pathological	Training dataset		Validation dataset		Test 1 dataset		Test 2 dataset	
	Colloid	Calcification	Colloid	Calcification	Colloid	Calcification	Colloid	Microcalcification
No. of patients	277	356	56	88	96	82	32	26
No. of nodules	336	366	84	91	104	88	32	26
No. of original images	1270	1110	318	278	252	268	59	49
No. of images	12704	11104	3176	2776	/	/	/	/
Male (%)	177 (22.8%)				38 (21.3%)		17 (29.3%)	
Female (%)	600 (77.2%)				140 (78.7%)		41 (70.7%)	
Mean age	50.0 ± 11.8 years				49.6 ± 12.3 years		49.8 ± 13.3 years	
Mean nodule size	12.9 ± 11.5 mm				12.3 ± 10.7 mm		10.9 ± 9.8 mm	

Table 2

Performance of YoloV5 series model compared with radiologists in differentiating echogenic foci as calcification or colloid in thyroid nodules.

	YoloV5n	YoloV5s	YoloV5m	YoloV5l	YoloV5x	Radiologists Average	Model Average
Acc	0.7292	0.7292	0.7813	0.8229	0.8281	0.6129	0.7781
Se	0.7273	0.7614	0.7386	0.8182	0.8750	0.5114	0.7841
Sp	0.8636	0.8295	0.9659	0.9773	0.9318	0.8258	0.9136
PPV	0.6957	0.6837	0.7738	0.8000	0.7778	0.5887	0.7462
NPV	0.7600	0.7766	0.7870	0.8431	0.8817	0.5983	0.8097
F1	0.7111	0.7204	0.7558	0.8090	0.8235	0.5473	0.7640
YI	0.5909	0.5959	0.7045	0.7955	0.8068	0.3372	0.6987
Kappa	0.4564	0.4593	0.5579	0.6440	0.6571	0.2122	0.5550
AUROC	0.7484	0.7615	0.8288	0.8681	0.9135	0.6050	0.8241
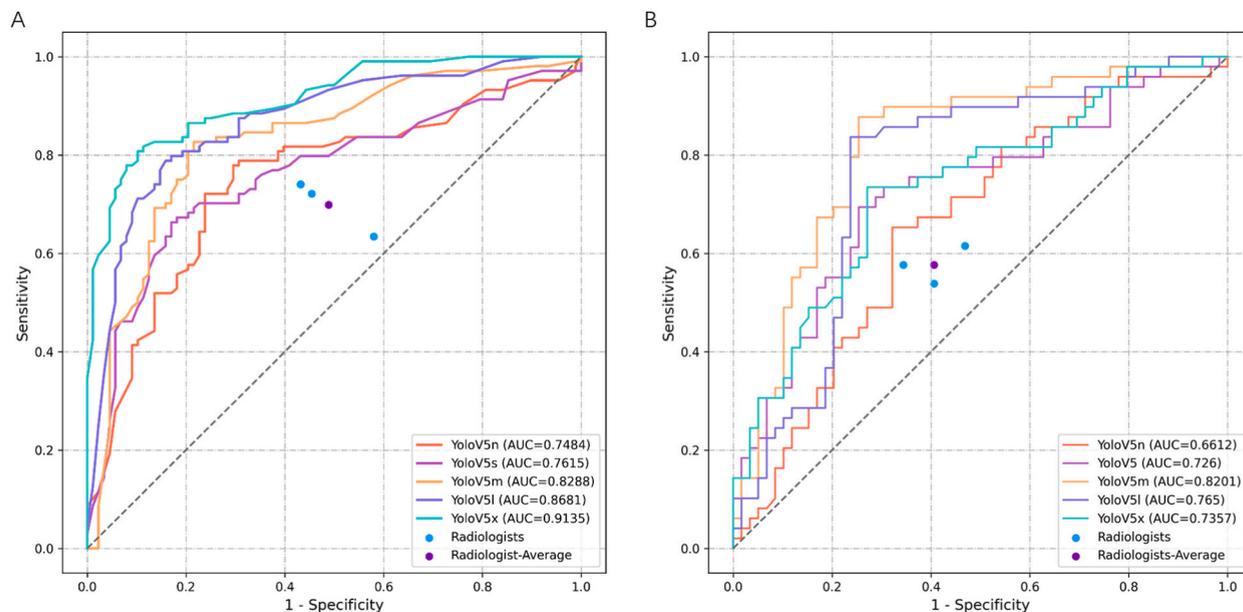
Fig. 2. Performance results for YoloV5 series model for differentiating echogenic foci in thyroid nodules as calcification or colloid (A), and performance results for YoloV5 series model for differentiating small echogenic foci in thyroid nodules as microcalcification or colloid (B).

Table 3

Performance of YoloV5 series model compared with radiologists in differentiating small echogenic foci as microcalcification or colloid in thyroid nodules.

	YoloV5n	YoloV5s	YoloV5m	YoloV5l	YoloV5x	Radiologists Average	Model Average
Acc	0.6481	0.7130	0.7778	0.7963	0.7315	0.5938	0.7333
Se	0.6271	0.6949	0.6949	0.7627	0.7288	0.5769	0.7017
Sp	0.6735	0.7347	0.8776	0.8367	0.7347	0.6329	0.7714
PPV	0.6981	0.7593	0.8723	0.8491	0.7678	0.5372	0.7893
NPV	0.6000	0.6667	0.7049	0.7455	0.6923	0.5555	0.6819
F1	0.6346	0.6990	0.7818	0.7885	0.7129	0.1707	0.7234
YI	0.3006	0.4259	0.5725	0.5994	0.4635	0.1696	0.4731
Kappa	0.2975	0.3271	0.5608	0.5933	0.4611	0.5853	0.4677
AUROC	0.6612	0.7260	0.8201	0.7650	0.7357	0.5862	0.7416

Additionally, we employed the YoloV5 series models to diagnose small echogenic foci in thyroid nodules, which holds significant clinical significance. Within the test 2 dataset, 25 nodules were malignant and 33 were benign. The YoloV5 series models achieved a mean AUC of 0.7416, mean sensitivity of 0.7017, and mean specificity of 0.7714 in identifying small echogenic foci. Among these, the YoloV5m model demonstrated the best generalization ability, with an AUC of 0.8201, sensitivity of 0.6949, and specificity of 0.8776. The mean AUC for radiologists in identifying small echogenic foci in thyroid nodules was 0.5862, with a mean sensitivity of 0.5769 and mean specificity of 0.6329 (Table 3, Fig. 2B). Consequently, the YoloV5 series models outperformed radiologists in identifying small echogenic foci of thyroid nodules. The corresponding interpretation of the DL models and radiologists are described in Table 4.

Table 4

Number of microcalcifications/colloid in the gold standard, and the corresponding interpretation of DL and radiologists.

	microcalcifications		colloid	
Number of images	49		59	
interpretation	microcalcifications	colloid	microcalcifications	colloid
YoloV5n	33	16	22	37
YoloV5s	36	13	18	41
YoloV5m	43	6	18	41
YoloV5l	41	8	14	45
YoloV5x	36	13	16	43
Number of nodules	26		32	
Radiologist 1	14	12	13	19
Radiologist 2	16	10	15	17
Radiologist 3	15	11	11	21
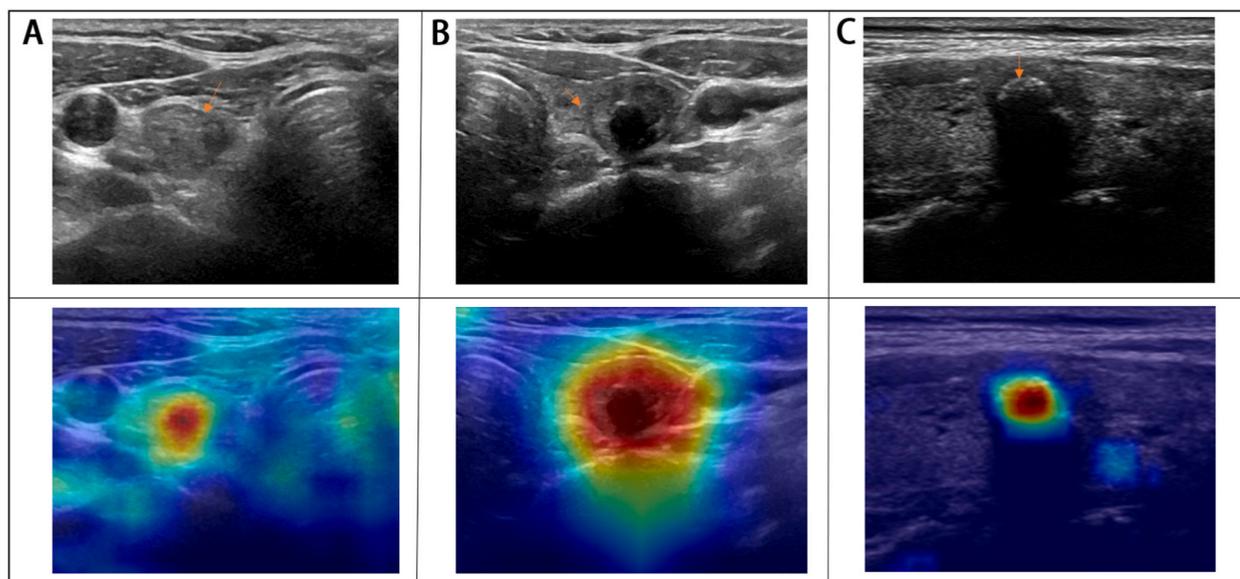
Fig. 3. Radiologist-labeled thyroid nodule region and the heat map generated by the object detection model on the thyroid nodule. The pathological results of A and B were calcification, the figures read by the radiologists were colloid, and the figures read by the model were calcification. The pathological result of C was colloid, the figure read by the radiologists was calcification, and the figure read by the model was colloid.

3.3. Heat maps generated by CAM

The heat map generated by the class activation mapping method shows the results of YoloV5 object detection model identification. The probability of the model identifying the thyroid region in the complete ultrasound image is shown as red, yellow, or green, in decreasing order. The red area in Fig. 3 indicates that this area has the highest probability of being a thyroid nodule in the whole figure. The pathological results of Fig. 3A and B were calcification, the figures read by the radiologists were colloid, and the figures read by the model were calcification. The pathological result of Fig. 3C was colloid, the figure read by the radiologists was calcification, and the figure read by the model was colloid (Fig. 3).

4. Discussion

Of thyroid nodules, 19.8–32.1% are accompanied by calcification, which is the most important ultrasound feature in their assessment [30]. Calcification exists in 8–32% of benign nodules and 26–54% of malignant nodules [31–33]. This suggests that calcification is more common in malignant nodules than benign nodules. Among them, microcalcifications and malignancy are highly correlated [34,35]. Microcalcifications in ultrasonography can have many pathological results, including psammomatous calcifications, dystrophic calcifications, and colloid deposits [36]. Calcifications generally appear as echogenic foci on ultrasound images, but not all pathological results of echogenic foci show calcifications. In clinical practice, some radiologists often equate echogenic foci with calcifications. Wang et al. [14] explored the relationship between ultrasound echogenic thyroid nodules and calcification in paraffin-wax sections. They found that 209 of 366 malignant nodules showing calcification on ultrasound were confirmed with calcification, and 127 of 414 benign nodules showing calcification on ultrasound were confirmed with calcification. In our study,

50.67% of the nodules with ultrasound showing calcification had pathological results with calcification.

Nodules that were pathologically confirmed to be colloid only were also shown as calcification in the ultrasound report. For radiologists, distinguishing between nodules with echogenic foci in the thyroid as calcifications or colloid deposits is a major difficulty. Colloid versus calcification has opposite clinical significance [37]. Intracystic echogenic foci with comet-tail artifacts reliably predict benign nodules with a risk of malignancy of <1–2% [38]. Ahuja et al. [39] showed similar results by studying 300 cases of thyroid nodules. They also found that the companion comet-tail artifact may be related to the presence of colloid. Therefore, colloid deposits may have some correlation with benign nodules. Su et al. [40] found better performance of the combined Thyroid Imaging Reporting and Data System (TIRADS) to assess different echogenic foci for the diagnosis of benign and malignant thyroid nodules. However, determining the presence or absence and type of echogenic foci is highly subjective.

This study was dedicated to developing a deep learning detection and classification model for differentiating nodules that show the same echogenic foci on ultrasound, which have pathological results of calcification or colloid only. This could help determine the benignity or malignancy of thyroid nodules. To our knowledge, this is the first time that a deep learning model has been used to identify echogenic foci in thyroid nodules. Our results show that the YoloV5 series models outperformed radiologists. The YoloV5 series model had a mean AUC of 0.8241, a mean sensitivity of 0.7841, and a mean specificity of 0.9136 for identifying echogenic foci in thyroid nodules. Among them, the YoloV5x model had the best generalization ability with the test dataset, with an AUC of 0.9135, a sensitivity of 0.8750, and a specificity of 0.9318. Our three radiologists have been working for 5, 8, and 15 years. The mean AUC for the radiologists to identify echogenic foci in thyroid nodules was 0.6050, the mean sensitivity was 0.5114, and the mean specificity was 0.825. The AUC, accuracy, sensitivity, and specificity of the model were higher than the average of radiologists. The sensitivity of the model was much higher than the sensitivity of the radiologists. This illustrates the ability of the model to identify calcifications much better than radiologists. The specificity of the model was 0.9773, indicating a low probability of the model misclassifying colloid as calcification. Because the pathological results of colloid are more suggestive of benign nodules, the model may, to some extent, reduce some unnecessary biopsies. Moreover, based on clinical experience, distinguishing between microcalcifications and colloid in thyroid nodules is challenging for radiologists, as both types manifest as small echogenic foci on ultrasound images. Hence, we employed DL models to differentiate between microcalcifications and colloid. The results indicated that our models outperformed radiologists in identifying calcification and colloid, including microcalcification and colloid. When using the YoloV5m model with the best generalization ability to identify small echogenic foci, 31 nodules were diagnosed as colloid, of which 20 were confirmed to be benign through pathology. By diagnosing colloid, this model successfully classified 20 out of 33 benign nodules in the test 2 dataset as benign. The thyroid ultrasound images used in our study were from different ultrasound devices. This helps increase the heterogeneity of the data and improves the generalization ability of the model, which can be used as an objective opinion to assist the radiologist in making a diagnosis. However, such AI models are not suitable for clinical use alone, and the final clinical decision must be made by the radiologist. The results predicted by our model can, to a degree, help radiologists determine the benignity or malignancy of thyroid nodules.

AI methods have various applications in the field of medicine. For instance, the DL detection model in this study integrates AI's lesion detection and classification technology into one model, achieving lesion localization and type classification. In DL segmentation models, the boundaries of lesions can be segmented, and diagnosis can be performed based on the outlines, shapes of lesions, and overall information obtained from the segmentation. Another mainstream direction of AI is machine learning. This method can model data such as patient's clinical information, images, and different omics. Machine learning can eliminate the manual delineation step of deep learning, achieving unsupervised learning, enabling the model to learn the intrinsic rules of data through continuous iterations.

Our study has several limitations. First, the process of scanning for thyroid nodules by a radiologist is dynamic. The ultrasound images used for training, validation, and testing of our model are all static, and the effective information obtained is far less than that obtained by dynamic scanning. Second, despite the numerous advantages of AI models, certain limitations persist. Due to the complexity, black box nature, abstractness, and nonlinearity of AI models, the process of human interpretation of AI decisions remains an unresolved problem. This is not conducive to radiologists learning how to distinguish between calcification and colloid on ultrasound images. Third, the ultrasound images of the thyroid gland that we obtained had clear pathological results. However, in the real world, most thyroid nodules detected by screening do not undergo pathological examination. This may affect the generalizability of the model. Finally, due to the retrospective nature of this study, definitive requirements for pathologists to make a pathological diagnosis were absent. In future endeavors, we plan to conduct a prospective, multicenter, large-sample study that will propose clearer inclusion criteria for the pathological diagnosis of cases in line with the objectives of the prospective investigation.

In summary, the model developed in this study was trained and tested with 30,680 static images of 1069 nodules. It could identify whether echogenic foci in thyroid nodules were calcification or colloid and may improve the accuracy of the differential diagnosis of thyroid nodules by radiologists.

Ethics statement

The research has been carried out in accordance with the World Medical Association Declaration of Helsinki. And this retrospective study was approved by the Ethics Committee of Zhejiang Cancer Hospital, and informed consent was waived (IRB-2020-287)

Funding statement

Data Availability Statement

Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

Author contributions

Chen Chen: conceived and designed the experiments; performed the experiments; analyzed and interpreted the data; wrote the paper.

Yuanzhen Liu: performed the experiments; analyzed and interpreted the data; contributed reagents, materials, analysis tools or data.

Jincao Yao：contributed reagents, materials, analysis tools or data.

Lujiao Lv: contributed reagents, materials, analysis tools or data.

Qianmeng Pan: performed the experiments.

Jinxin Wu: contributed reagents, materials, analysis tools or data.

Changfu Zheng: performed the experiments.

Hui Wang: performed the experiments.

Xianping Jiang: performed the experiments.

Yifan Wang: contributed reagents, materials, analysis tools or data.

Dong Xu: conceived and designed the experiments; contributed reagents, materials, analysis tools or data.

Declaration of competing interest

The authors declare that they have no known competing financial interests.

Acknowledgements

None.

References

[1] K. Kobaly, C.S. Kim, S.J. Mandel, Contemporary management of thyroid nodules, Annu. Rev. Med. 73 (2022) 517–528.
[2] R.-M. Feng, et al., Current cancer situation in China: good or bad news from the 2018 Global Cancer Statistics? Cancer Commun. 39 (1) (2019) 22.
[3] S. Vaccarella, et al., Worldwide thyroid-cancer Epidemic? The increasing impact of overdiagnosis, N. Engl. J. Med. 375 (7) (2016) 614–617.
[4] G. Grani, et al., Contemporary thyroid nodule evaluation and management, J. Clin. Endocrinol. Metabol. 105 (9) (2020) 2869–2883.
[5] J.H. Shin, et al., Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean society of thyroid radiology consensus statement and recommendations, Korean J. Radiol. 17 (3) (2016) 370–395.
[6] R. Wong, S.G. Farrell, M. Grossmann, Thyroid nodules: diagnosis and management, Med. J. Aust. 209 (2) (2018) 92–98.
[7] E.K. Alexander, G.M. Doherty, J.A. Barletta, Management of thyroid nodules, Lancet Diabetes Endocrinol. 10 (7) (2022) 540–548.
[8] B.R. Haugen, et al., American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer, Thyroid : Official Journal of the American Thyroid Association 26 (1) (2015), 2016.
[9] F.N. Tessler, et al., ACR thyroid imaging, reporting and data System (TI-RADS): white paper of the ACR TI-RADS committee, J. Am. Coll. Radiol. : JACR 14 (5) (2017) 587–595.
[10] J.P. Brito, et al., The accuracy of thyroid nodule ultrasound to predict thyroid cancer: systematic review and meta-analysis, J. Clin. Endocrinol. Metabol. 99 (4) (2014) 1253–1263.
[11] L.B. Ferreira, et al., Molecular aspects of thyroid calcification, Int. J. Mol. Sci. 21 (20) (2020).
[12] L. Yin, et al., Relationship between morphologic characteristics of ultrasonic calcification in thyroid nodules and thyroid carcinoma, Ultrasound Med. Biol. 46 (1) (2020) 20–25.
[13] M.L.C. Khoo, et al., Thyroid calcification and its association with thyroid carcinoma, Head Neck 24 (7) (2002) 651–655.
[14] G. Wu, et al., Do hyperechoic thyroid nodules on B-ultrasound represent calcification? J. Int. Med. Res. 41 (3) (2013) 848–854.
[15] H. Malhi, et al., Echogenic foci in thyroid nodules: significance of posterior acoustic artifacts, AJR. American Journal of Roentgenology 203 (6) (2014) 1310–1316.
[16] V.T. Hoang, C.T. Trinh, A review of the pathology, diagnosis and management of colloid goitre, Eur. Endocrinol. 16 (2) (2020) 131–135.
[17] V.A. LiVolsi, Z.W. Baloch, The pathology of hyperthyroidism, Front. Endocrinol. 9 (2018) 737.
[18] M.D. Beland, et al., Nonshadowing echogenic foci in thyroid nodules: are certain appearances enough to avoid thyroid biopsy? J. Ultrasound Med. : Official Journal of the American Institute of Ultrasound In Medicine 30 (6) (2011) 753–760.
[19] D.T. Ginat, et al., Pearls and pitfalls of thyroid nodule sonography and fine-needle aspiration, Ultrasound Q. 26 (3) (2010) 171–178.
[20] B. Zhang, et al., Machine learning-assisted System for thyroid nodule diagnosis, Thyroid : Official Journal of the American Thyroid Association 29 (6) (2019) 858–867.
[21] K.-Y. Chen, et al., Computerized detection and quantification of microcalcifications in thyroid nodules, Ultrasound Med. Biol. 37 (6) (2011) 870–878.
[22] Y. Chen, et al., An artificial intelligence model based on ACR TI-RADS characteristics for US diagnosis of thyroid nodules, Radiology 303 (3) (2022) 613–619.
[23] J. Yao, et al., DeepThy-Net: a multimodal deep learning method for predicting cervical lymph node metastasis in papillary thyroid cancer, Adv. Intell. Syst. 4 (10) (2022) 2200100.
[24] W. Zhao, et al., Convolutional neural network-based computer-assisted diagnosis of hashimoto's thyroiditis on ultrasound, J. Clin. Endocrinol. Metabol. 107 (4) (2022) 953–963.

[25] H. Zhang, et al., Mixup: beyond Empirical Risk Minimization, Learning, 2017.

[26] T. DeVries, G.W. Taylor, Improved Regularization of Convolutional Neural Networks with Cutout, Computer Vision and Pattern Recognition, 2017.

[27] W.C.Y. Bochkovskiy A, Liao H Yolov4, Optimal Speed and Accuracy of Object Detection, Computer Vision and Pattern Recognition, 2020.

[28] W. Leisenring, T. Alonzo, M.S. Pepe, Comparisons of predictive values of binary medical diagnostic tests for paired designs, Biometrics 56 (2) (2000) 345–351.

[29] Y. Wu, Joint comparison of the predictive values of multiple binary diagnostic tests: an extension of McNemar's test, J. Biopharm. Stat. 33 (1) (2023) 31–42.

[30] Y. Bai, et al., Survival impact of psammoma body, stromal calcification, and bone formation in papillary thyroid carcinoma, Mod. Pathol. : an Official Journal of the United States and Canadian Academy of Pathology, Inc 22 (7) (2009) 887–894.

[31] Y. Li, et al., Efficacy and safety of radiofrequency ablation for calcified benign thyroid nodules: results of over 5 years' follow-up, BMC Med. Imag. 22 (1) (2022) 75.

[32] Z. Lu, et al., Clinical value of using ultrasound to assess calcification patterns in thyroid nodules, World J. Surg. 35 (1) (2011) 122–127.

[33] G. Chen, et al., *Retrospective analysis of thyroid nodules by clinical and pathological characteristics, and ultrasonographically detected calcification correlated to thyroid carcinoma in South China.* European Surgical Research. Europaische Chirurgische Forschung, Recherches Chirurgicales Europeennes 42 (3) (2009) 137–142.

[34] D.G. Na, et al., Thyroid imaging reporting and data System risk stratification of thyroid nodules: categorization based on solidity and Echogenicity, Thyroid : Official Journal of the American Thyroid Association 26 (4) (2016) 562–572.

[35] V. Triggiani, et al., Microcalcifications and psammoma bodies in thyroid tumors, Thyroid : Official Journal of the American Thyroid Association 18 (9) (2008) 1017–1018.

[36] A.M. Tahvildari, et al., Sonographic-pathologic correlation for punctate echogenic reflectors in papillary thyroid carcinoma: what are they? J. Ultrasound Med. : Official Journal of the American Institute of Ultrasound In Medicine 35 (8) (2016) 1645–1652.

[37] C. Maxwell, J.A. Sipos, Clinical diagnostic evaluation of thyroid nodules, Endocrinol Metab. Clin. N. Am. 48 (1) (2019) 61–84.

[38] E.J. Ha, et al., Korean thyroid imaging reporting and data System and imaging-based management of thyroid nodules: Korean society of thyroid radiology consensus statement and recommendations, 2021, Korean J. Radiol. 22 (12) (2021) 2094–2123.

[39] A. Ahuja, et al., Clinical significance of the comet-tail artifact in thyroid ultrasound, J. Clin. Ultrasound : JCU (J. Clin. Ultrasound) 24 (3) (1996) 129–133.

[40] S.M. Ha, et al., Echogenic foci in thyroid nodules: diagnostic performance with combination of TIRADS and echogenic foci, BMC Med. Imag. 19 (1) (2019) 28.