Dovepress

open access to scientific and medical research

ORIGINAL RESEARCH

Fully-Automatic Detection and Diagnosis System for Thyroid Nodules Based on Ultrasound Video Sequences by Artificial Intelligence

Dan Liu[1,*], Ke Yang[2,*], Chunquan Zhang[1,*], Dandan Xiao[1], Yu Zhao[1]

[1]Department of Ultrasound, The Second Affiliated Hospital of Nanchang University, Nanchang, 330006, People's Republic of China; [2]The First in-Patient Department, Jiangxi Provincial People's Hospital, The First Affiliated Hospital of Nanchang Medical College, Nanchang, 330006, People's Republic of China

*These authors contributed equally to this work

Correspondence: Dan Liu, Department of Ultrasound, The Second Affiliated Hospital of Nanchang University, No. 1, Minde Road, Donghu District, Nanchang, Jiangxi, 330006, People's Republic of China, Tel +86-15170076809, Email ultrald@163.com

Background: Interpretation of ultrasound findings of thyroid nodules is subjective and labor-intensive for radiologists. Artificial intelligence (AI) is a relatively objective and efficient technology. We aimed to establish a fully automatic detection and diagnosis system for thyroid nodules based on AI technology by analyzing ultrasound video sequences.

Patients and Methods: We prospectively acquired dynamic ultrasound videos of 1067 thyroid nodules (804 for training and 263 for validation) from December 2018 to January 2021. All the patients underwent hemithyroidectomy or total thyroidectomy. Dynamic ultrasound videos were used to develop an AI system consisting of two deep learning models that could automatically detect and diagnose thyroid nodules. Average precision (AP) was used to estimate the performance of the detection model. The area under the receiver operating characteristic curve (AUC) was used to measure the performance of the diagnostic model.

Results: Location and shape were accurately detected with a high AP of 0.914 in the validation cohort. The AUC of the diagnostic model was 0.953 in the validation cohort. The sensitivity and specificity of junior and senior radiologists were 76.9% vs 78.3% and 68.4% vs 81.1%, respectively. The diagnostic performance of the AI diagnostic model was superior to that of junior radiologists ($P = 0.016$) and was not significantly different from that of senior radiologists ($P = 0.281$).

Conclusion: We established a fully automatic detection and diagnosis system for thyroid nodules based on ultrasound video using an AI approach that can be conveniently applied to optimize the management of patients with thyroid nodules.

Keywords: Thyroid nodule, Ultrasonography, Artificial intelligence, Deep learning, Radiomics

Introduction

Thyroid nodules are exceedingly common and can be detected in approximately 19–68% of the general population using ultrasonography.[1,2] A large proportion of thyroid nodules are benign with no clinical significance or innocent behavior.[1] Therefore, the main current challenge for physicians is to accurately classify thyroid nodules, which can help avoid unnecessary surgery for benign nodules and determine the proper treatment strategies for malignant nodules.

Ultrasonography has become the primary imaging modality for thyroid nodules owing to its advantages of real-time, non-radiation, high resolution, and simplicity. Previous studies have explored the ultrasound image characteristics of different thyroid nodules and revealed that some features are related to risk stratification for malignancy, including taller-than-wide orientation, solid composition, marked hypoechogenicity, microcalcification, irregular margin, and extrathyroidal extension.[3,4] However, interpretation of ultrasound imaging finding is subjective.[5,6] To improve inter- and intra-observer agreement, several committees successively established ultrasound Thyroid Imaging Reporting and Data System (TIRADS) to assist in clinical diagnosis and treatment decisions for thyroid nodules, including Korean-TIRADS, Eu-TIRADS, and ACR-TIRADS.[7–9] Despite their high accuracy, some issues still reflect the lack of clinical

practicality of these systems, which limits their generalization of these classification systems in daily clinical work. First, the evaluation of thyroid nodules is time consuming because of the complexity of the system. Each thyroid nodule is given a total point mainly based on the ultrasound features, such as composition, echogenicity, shape, margin and echogenic foci. The point total determines the nodule's TI-RADS level. The whole analysis and evaluation process takes about 10–15 minutes for each nodule. Second, the interpretation of ultrasonography findings possesses low inter- and intra-observer consistency, which is heavily dependent on the experience of radiologists.[5,6] Third, some benign and malignant thyroid nodules share similar sonographic characteristics perceptible to the human eye, which makes them difficult for radiologists to distinguish.[6] Therefore, a reliable, objective, and individualized method is urgently needed to identify thyroid nodules.

Artificial intelligence (AI) is a revolutionary technology for medical image analysis to achieve the above goals[10] and has been broadly applied in medical imaging, such as computed tomography (CT), magnetic resonance imaging (MRI), X-ray, and ultrasonography.[11–14] Recent research has proven that, with proper analysis methods, AI can achieve satisfactory performance on various clinical issues based on the automatic analysis of ultrasonography.[15–19] B-mode images and videos, shear-wave elastography, and contrast-enhanced ultrasound cine are the common ultrasound imaging techniques used in previous studies. A series of clinical studies have revealed that ultrasonography plays an important role in the development of current medical AI. In recent years, many studies have attempted to analyze the ultrasound images of thyroid nodules using AI technology.[5,20–23] However, there still exists some limitations remain in the proposed AI model. First, current approaches for the diagnosis of thyroid nodules are mainly based on two static ultrasound images in transverse and longitudinal views, rather than a complete ultrasound video sequence, which omits the stereo-structural characteristics of the lesions. Consequently, the classification process may lead to the loss of important comprehensive imaging findings in nodules. Second, manual delineation of nodules in each image was required before establishing the models. However, this process is time- and labor-intensive.

In this study, we aimed to establish a fully automatic detection and diagnosis system for thyroid nodules based on AI technology, by analyzing ultrasound video sequences and comparing the diagnostic precision of the AI system with that of radiologists.

Methods
Study Population
This study was performed in accordance with the principles of the Declaration of Helsinki. The study protocol was approved by the Ethics Commission of the Second Affiliated Hospital of Nanchang University, and informed consent was obtained from all the participants. All the patients consented to their anonymized medical records for analysis and publication for research purposes. From December 2018 to January 2021, 928 potentially eligible patients with 1132 thyroid nodules were recruited in our hospital. The inclusion criteria were as follows:[1] patients with ultrasound-suspected thyroid nodules,[2] those who underwent hemi-or total thyroidectomy, and[3] those with available pathologic and clinical data. The exclusion criteria were as follows:[1] complications with indistinguishable diffuse thyroid parenchymal disease,[2] indeterminate pathology results, and[3] incomplete information or ultra-sound video sequences. Sixty-five nodules from 59 patients were excluded, and 1067 nodules from 869 patients were enrolled. In the final analysis cohort, there were 804 nodules in 652 patients in the training cohort and 263 nodules in 217 patients in the validation cohort. The patient recruitment workflow is shown in Figure 1.

Ultrasonography Examination and Pathologic Ground Truth
One of the two sonographers with over 10 years of experience in thyroid ultrasonography performed preoperative thyroid ultrasonography using a Mindray Resona 7S ultrasound instrument (Mindray Medical International Ltd, Shenzhen, China) equipped with a 5–14 MHz linear array transducer. During ultrasound examination, the sonographer first performed routine thyroid and cervical lymph node ultrasonography. Then, a continuous dynamic video of the target thyroid nodule was acquired in transverse plane, and stored as DICOM file. All patients underwent hemi-or total thyroidectomy within one month of ultrasound examination. Pathologists with over 10 years of experience in thyroid surgery classified surgical specimens as benign or malignant.
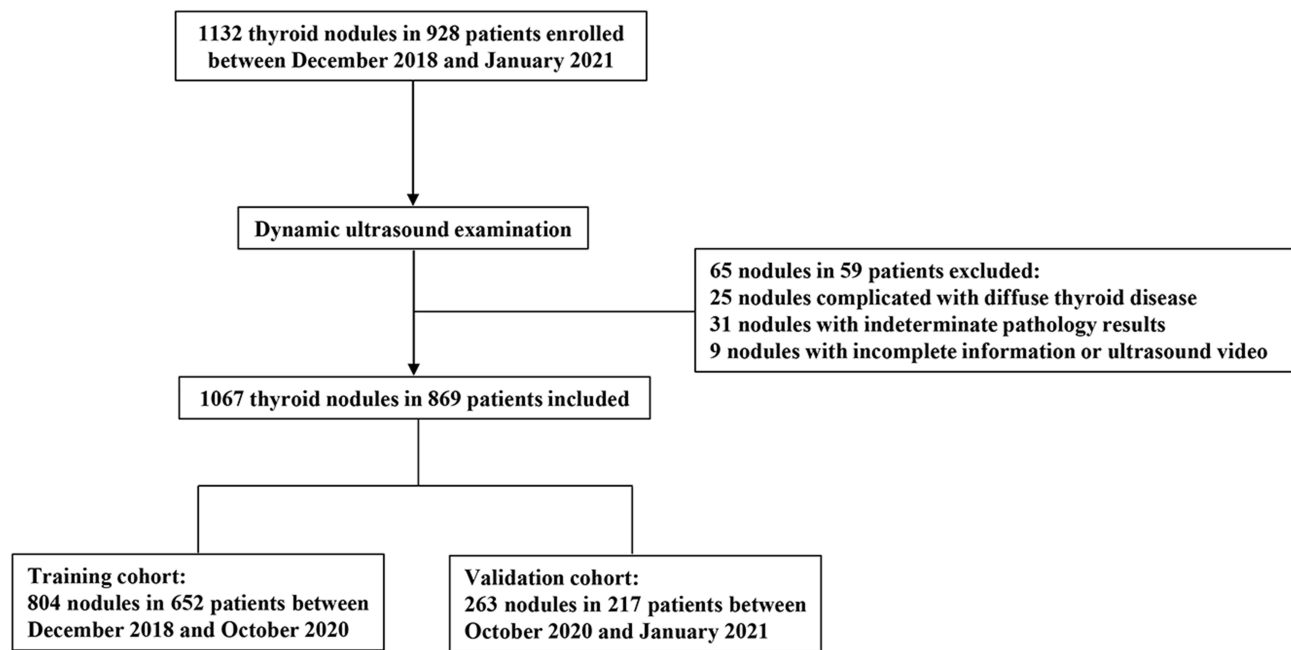
Figure 1 Patient recruitment workflow. In total, 1067 thyroid nodules in 869 patients were enrolled according to the selection criteria.

Nodule Annotation

To determine the region of interest (ROI) in our study, the enrolled ultrasound videos were annotated manually by a sonographer with 10 years of clinical experience. The annotator was blinded to the pathological results and clinical characteristics of patients during the annotation process. The open-source software Labelme was chosen as the annotation tool to improve efficiency (https://github.com/wkentaro/labelme). Specifically, we first separated the frames of the ultrasound videos into images and saved them for annotation. For each video, all frames were loaded into Labelme and the nodule regions were annotated using rectangular boxes (Figure 2a). In this manner, the locations of the nodule regions were saved in a JSON file and parsed in the following training process. The entire annotation process is simple and easy to perform, requiring less than one second per frame for an experienced doctor.

Development of Thyroid Nodule Detection and Diagnosis Models

The enrolled patients were randomly split into training and validation cohorts (n = 804 vs 263) for model development. In this study, we aimed to develop a diagnostic system that is as automatic as possible. Instead of inputting the doctor-annotated nodule region into the diagnosis system, as in other radiomics research, we explored the potential of combining deep-learning-based nodule detection and diagnosis models, which is a more challenging and meaningful direction to improve the efficiency of clinical diagnosis. The detailed configurations of the models were determined by 5-folded validation in the training cohort. To further improve the generalization of our models to unseen data, we expanded the training cohort by using the Thyroid Digital Image Database (http://cimalab.intec.co/applications/thyroid/) and TNSCUI2020 (https://tn-scui2020.grand-challenge.org/). In this way, the models were able to learn how to detect and diagnose thyroid nodules with sufficient datasets; thus, the automatic prediction results were more reliable.

In the proposed fully automatic AI system, two deep-learning-based models were developed for thyroid nodule detection and diagnosis. Inspired by the diagnosis process of doctor, we firstly determine the potential thyroid nodule region for each video frame by the detection model, and then the prediction nodule regions of an ultrasound video were cropped and inputted to the diagnosis model for the final diagnosis. Specifically, we adopted a Feature Pyramid Network variant of the Faster- Region convolutional neural network (RCNN) algorithm[24] to build the nodule detection model (Figure 2b). In the detection model, the video frames were input as independent training data and the nodule location annotated by the doctor was regarded as the training label. The ResNet-18 algorithm was adopted as the primary classifier for the diagnosis model (Figure 2c). First, we trained the diagnosis model to estimate the probability of malignancy of the nodules predicted by the detection model in each frame of the
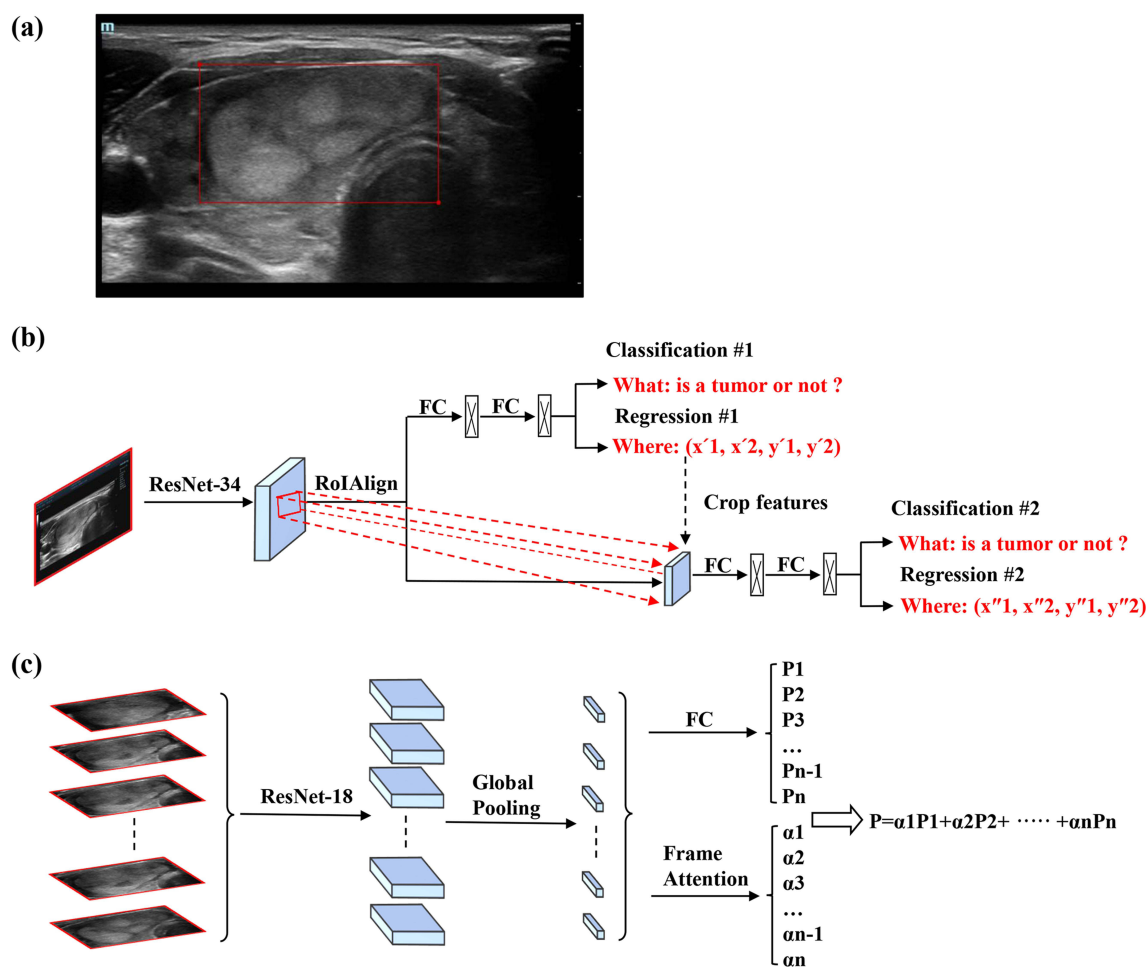
Figure 2 The nodule annotation and the design of detection and diagnosis models. (**a**) An example of nodule ROI annotation in one frame image from the ultrasound video. For each video, all the frames were loaded into Labelme software and the nodule regions were marked by rectangle boxes. (**b**) The network structure of detection model. We adopted the ResNet-34 algorithm to cascadedlly predict whether an image region belongs to thyroid nodule and where the nodule exactly is. (**c**) The network structure of the diagnosis model. Following detection model, the detected thyroid nodule regions from one ultrasound sequence were analyzed by ResNet-18 algorithm and then combined by attention mechanism to extract the stereo information of ultrasound sequences. Pn represented the malignant probability for the nth frame. An represented the importance weight for the nth frame.
Abbreviations: ROI, region of interest; FC, fully connection.

ultrasound video. We then packed all the frames from one video as input and added an attention module at the end of the diagnosis model to estimate the importance weights of each frame in the final diagnosis results. The features of the frames that were beneficial to the correct diagnosis were assigned higher importance weights and played an important role, whereas the contribution of frames with lower weights was suppressed. Figure 3 illustrated the construction of the fully-automatic AI system. More detailed model design and training strategies are provided in the Supplementary Methods.

Comparison of Diagnostic Performance Between the Diagnosis Model and Radiologists

A junior radiologist with five years of experience and an experienced radiologist with 15 years of experience diagnosed thyroid nodules based on ultrasound videos in the validation cohort. The two radiologists made a tendentious benign or malignant diagnosis based mainly on ACR-TIRADS categories and experience.[25] They were blinded to the pathological information and original ultrasound data acquired during the ultrasound examination. The area under the receiver operating characteristic curve (AUCs) of the two radiologists was compared to that of the AI diagnosis model.
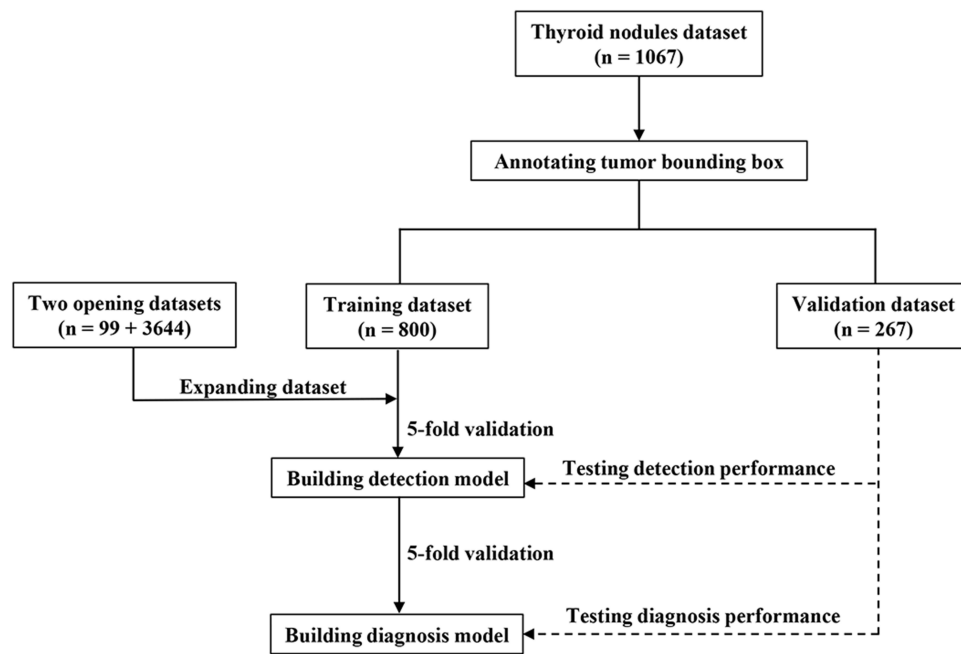
Figure 3 Fully-automatic AI system building flowchart. Two deep learning-based models were developed for thyroid nodules detection and diagnosis in the fully-automatic AI system. In order to improve the generalization of the models to unseen data, the training cohort was expanded by adding two open ultrasound datasets.
Abbreviation: AI, artificial intelligence.

Statistical Analysis

Student's t-test or Mann–Whitney U-test, as appropriate, and the χ2 test were used to compare continuous and categorical variables, respectively. The performance of the diagnostic model was measured using the AUC, specificity, sensitivity, and accuracy. To estimate the performance of the detection model, we computed the intersection over union (IoU) as the overlap ratio between the predicted nodule regions and the manual annotations. IoU was an indicator used to estimate the prediction accuracy of the location information of the target nodule. Detection result with an IoU above one threshold was signified as true predictions. When the IoU threshold was set higher, the requirement for the accuracy of the prediction result was more stringent. In order to obtain a high recall rate, we set the IoU threshold to 0.5 in this study. It meant that if the overlap ratio between the predicted nodule regions and the manual annotations was above 50%, we regarded that the model detected the nodule accurately. Precision was used to measure the rate of true nodule regions among all detected results, and recall was used to measure the rate of true nodule regions among all annotations. Similar to the concept of the receiver operating characteristic curve (ROC) and AUC, when a Precision-Recall curve was plotted, the area under this curve, termed average precision (AP), was used to measure the detection performance considering the balance of Precision and Recall. Data analysis was performed using R software (version 3.4.4; https://www.r-project.org/). The significance level was set at $P < 0.05$.

Results

Baseline Characteristics

The mean age of 869 patients were 47.8±12.6 years (range, 12–82 years). There were 724 (724/928, 78.0%) females and 204 (204/928, 22.0%) males. Among the 1067 nodules, 575 were malignant (575/1067, 53.89%) and 492 were benign (492/1067, 46.11%). There were 471 thyroid nodules in the left lobe, 555 nodules in the right lobe, and 41 nodules in the isthmus. Ninety-six nodules were located near the thyroid capsule (within 2 mm close to the thyroid capsule). The mean maximal nodule size was 1.7±1.3 cm. There were no significant differences in variables between the training and validation cohorts (Table 1).

Precision of the Detection Model

The performance of the detection model for thyroid nodules is summarized in Table 2. Our experimental results revealed that the detection model could detect almost all annotated nodules, with recall rates of 0.997 and 0.977 in the training and

Table 1 Baseline Characteristics of Patients in Two Cohorts

Characteristic	Training Cohort	Validation Cohort	P value
Patients number (%)	652(75.03%)	217(24.97%)	–
Nodules number (%)	804(75.35%)	263(24.65%)	–
Age, years	46.6±12.7(12–82)	48.1±12.4(16–78)	0.129
Sex			0.467
Male	164(25.2%)	49(22.6%)	
Female	488(74.8%)	168(77.4%)	
Pathological results			0.154
Benign	381(47.4%)	111(42.2%)	
Malignant	423(52.6%)	152(57.8%)	

Table 2 The Performance of the Thyroid Nodule Detection Model

Cohort Name	AP	Recall
Training cohort	0.993 (0.992–0.994)	0.997 (0.996–0.998)
Validation cohort	0.914 (0.910–0.917)	0.977 (0.975–0.978)

Abbreviation: AP, average precision.

validation cohorts, respectively (Figure 4a). The AP remained at a high level even when the IoU threshold exceeded 0.7, suggesting the high performance of the detection model (Figure 4b). As shown in Figure 4c and Figure 5, the location and shape were detected with high average precision (0.993 and 0.914 in the training and validation cohorts, respectively) when the detection threshold was set to 0.5. The detection process was efficient with an average cost of 2.110 s for a complete ultrasound video. The results suggest that it is reliable to automatically detect thyroid nodules using our detection model for the proposed diagnosis model.

Performance of the Diagnosis Model

The AI model achieved a high diagnostic performance in both the training and validation cohorts (Table 3). In the training cohort, sensitivity, specificity, accuracy, and AUC were 99.9%, 100.0%, 99.9%, and 1.000, respectively. In the validation cohort, sensitivity, specificity, accuracy, and AUC were 83.5%, 94.6%, 87.4%, and 0.953, respectively. Figure 6a shows the ROC curve of the AI diagnosis model for the training and validation cohorts. In addition, the AI model could quickly diagnose thyroid nodules, spending only 0.026 s.
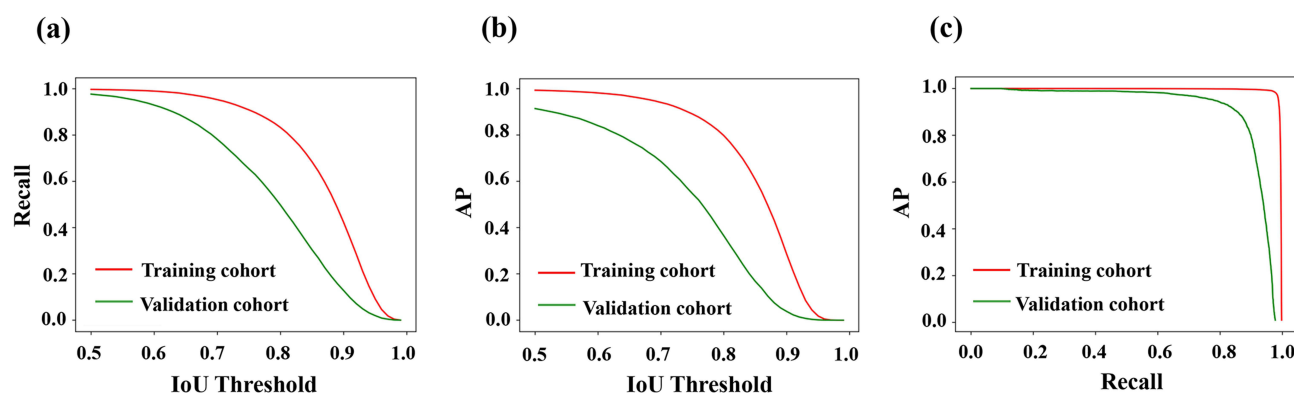

Figure 4 Precision of the detection model. (**a** and **b**) presented the variation of Recall and AP along with IoU threshold. The Recall and AP kept a high level even when the IoU threshold exceeds 0.7, suggesting the high performance of detection model. (**c**) presented the relation between AP and Recall, indicating our detection model can keep high Recall rate while achieve high AP simultaneously.
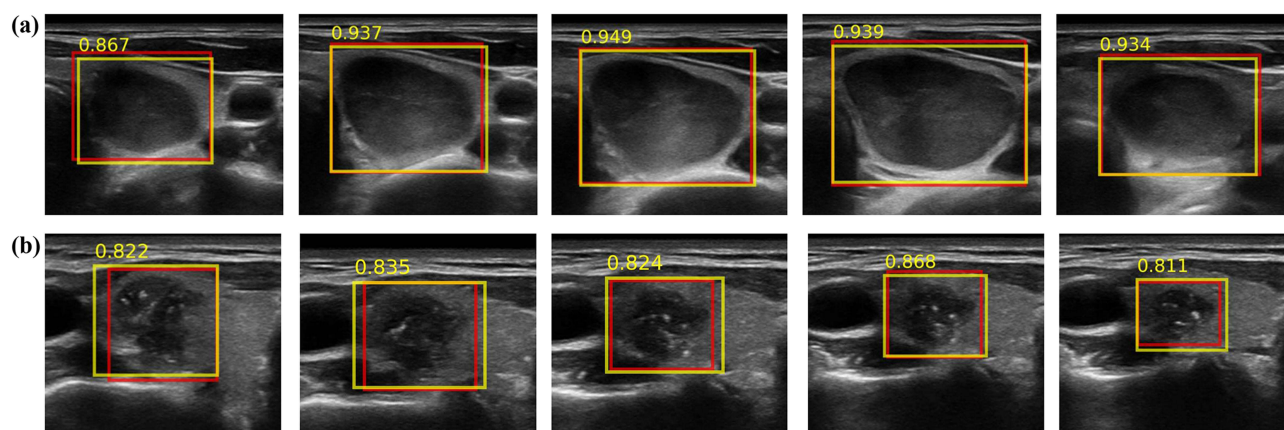**Abbreviations**: IoU, intersection over union; AP, average Precision.

Figure 5 Illustration of the location and shape of nodules by detection model in ultrasound videos. (**a**) demonstrated the detection results in five frames of one ultrasound video in a benign thyroid nodule case. Similarly, (**b**) represented the detection results in a malignant nodule case. The red and yellow boxes are the manually annotated and the model predicted thyroid nodule regions. The yellow number at the upper-left corner in the figures is the IoU between the predicted and the annotated bounding boxes. The detection model detected almost all the annotated nodules with high precision.

We analyzed the differences in model features between benign and malignant thyroid nodules by extracting the main descriptive dimensions of ultrasound features using principal component analysis (PCA). The results of the feature analysis are presented in Figure 6b. The patients clustered compactly according to different categories using our model features. This qualitatively demonstrates the ability of the diagnostic model to accurately extract key information for the diagnosis of thyroid nodules.

Comparison Between the AI Diagnosis Model and Radiologists

In the validation cohort, the sensitivity and specificity of junior and experienced radiologists were 76.9% and 78.3% and 68.4% and 81.1%, respectively (Table 3) (Supplementary Figure 1). There were significant differences between the diagnostic performance of the two radiologists ($P = 0.045$). The diagnostic performance of our proposed AI diagnostic model was superior to that of a junior radiologist ($P = 0.016$). The performance of the AI model was not significantly different from that of an experienced radiologist ($P = 0.281$).

Discussion

In this study, we proposed a fully automatic AI detection and diagnosis system for thyroid nodules based on ultrasound video sequences. To the best of our knowledge, there have been few reports on the construction of AI models by analyzing continuous dynamic ultrasound videos of thyroid nodules.[26] A previous study aimed to develop a deep-learning-based risk stratification system for thyroid nodules using ultrasound cine images, which presented a high classification performance with an AUC of 0.88.[26] However, the dataset size in the previous study was limited to 192 nodules. In our study, we included large sample data with 1067 thyroid nodules to construct the models, which would make the performance of the models more stable and reliable. Moreover, the algorithm depended on manual segmentation of the nodules in subsets of the cine frame images in the previous study. In this study, we explored a deep learning

Table 3 The Diagnosis Performance of AI Model and Radiologists

	AUC	Sensitivity, %	Specificity, %	Accuracy, %	PPV, %	NPV, %
AI model (T)	1.000 (0.995–1.000)	99.9 (99.1–100.0)	100.0 (99.0–100.0)	99.9 (99.5–100.0)	100.0 (98.4–100.0)	100.0 (98.2–100.0)
AI model (V)	0.953 (0.920–0.975)	83.5 (76.7–89.1)	94.6 (88.6–98.0)	87.4 (82.8–91.2)	95.5 (90.6–97.9)	80.8 (74.5–85.8)
Junior radiologist (T)	0.727 (0.669–0.780)	76.9 (6.95–83.4)	68.4 (59.0–77.0)	73.2 (67.6–78.6)	77.0 (71.5–81.7)	68.5 (61.3–74.9)
Senior radiologist (T)	0.797 (0.743–0.844)	78.3 (70.9–84.6)	81.1 (72.5–87.9)	79.4 (74.0–84.1)	85.0 (79.3–89.4)	73.2 (66.6–78.9)

Abbreviations: AI, artificial intelligence; T, training cohort; V, validation cohort; AUC, area under the receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value.
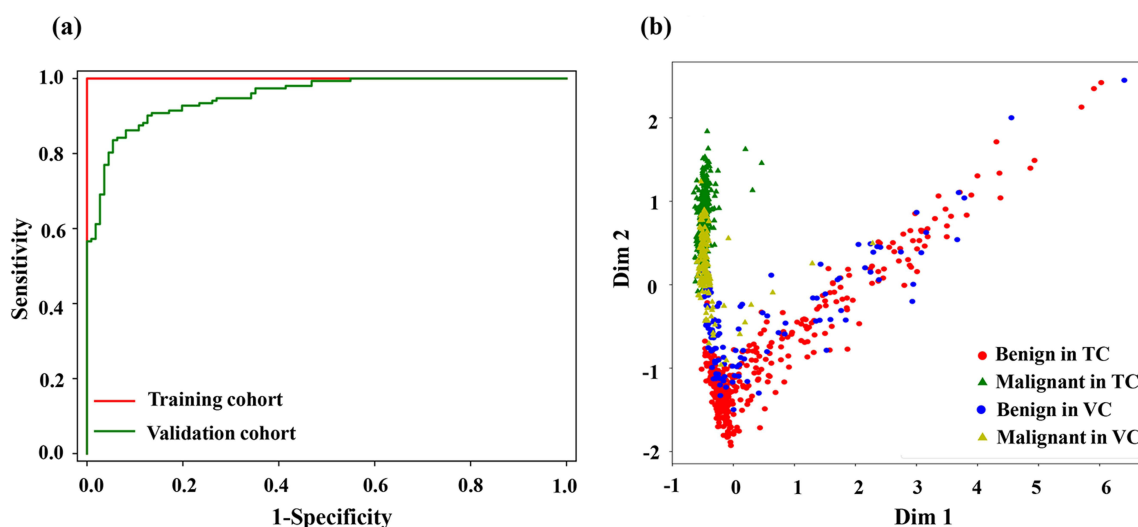
Figure 6 ROC curve and feature analysis of the diagnosis model. (**a**) AUCs of the AI diagnosis model were 1.000 and 0.953 in the training and validation cohorts, respectively. (**b**) The distribution of enrolled patients in the two most important dimensions of diagnosis model features. The features were extracted from the fully connection layers of diagnosis model, and the dimensions were reduced to two by principal components analysis method. The two main feature dimensions were used as figure axis, and marked the patients with different categories and cohorts by different color and shapes for better understanding.
Abbreviations: TC, training cohort; VC, validation cohort; ROC curve, receiver operating characteristic curve; AUC, area under the receiver operating characteristic curve; AI, artificial intelligence.

strategy to integrate the detection and diagnosis of thyroid nodules by quantitatively analyzing ultrasound cines. Our results revealed that we successfully developed an AI system consisting of two models that can automatically detect and diagnose thyroid nodules accurately and effectively. In the detection model, the location and shape of annotated nodules were detected with a high recall rate. The sensitivity and specificity of the diagnosis model reached 83.5% and 94.6%, respectively, in the validation cohort, showing a higher performance than that of the junior radiologist ($P = 0.016$). This demonstrates that the proposed AI diagnosis model can improve diagnostic accuracy for young inexperienced radiologists in a clinical setting. Furthermore, it holds great potential for shortening the learning curve for junior radiologists from underdeveloped hospitals when compared with the relatively accurate diagnostic results of the AI model. In addition, the diagnosis process of our proposed AI system was time-saving, with only 0.026 s for a thyroid nodule. The results implied that the thyroid nodule AI system could speed up the diagnostic process and help ease the daily workload of radiologists.

The classical deep learning process includes three steps: lesion detection and segmentation, feature extraction, and model construction.[27] Among these, accurate lesion detection is the key process that lays the foundation for model development. However, most current studies have delineated thyroid nodules manually, which is operator-dependent and time-consuming.[21–23] In this study, we innovatively established a detection model to automatically determine thyroid nodule regions using a deep learning approach based on continuous dynamic ultrasound videos, with the motivation of replacing the expertise of radiologists to some extent. The detection model was able to automatically detect almost all annotated thyroid nodules, and the location and shape of lesions were detected precisely and efficiently, with less than 3 s per patient. Moreover, the proposed detection model provides a firm basis for a subsequent automatic thyroid nodule diagnosis system, which is helpful for making the diagnosis more reliable, objective, and reproducible.

In the current study, we attempted to imitate the diagnostic thinking pattern of radiologists by applying the emerging AI method, which was aimed at replicating radiologists' accumulated experience to some extent. In clinical practice, radiologists always make diagnoses based on the overall ultrasonographic characteristics of thyroid nodules. In fact, a continuous dynamic ultrasound video is composed of numerous static images that contain much more information than the two images in the transverse and longitudinal views. In addition, a complementary ultrasound video is not only a superposition of images but also contains other structural information of thyroid nodules in different views. In practical situations, radiologists can determine the overall characteristics of thyroid nodules by interpreting a complete ultrasound

video that thoroughly reflects the stereoscopic structure and spatial characteristics of the lesion. In previous studies,[5,20–23] diagnosis models were mainly built based on static thyroid nodule ultrasound images, which limits their performance. In theory, an AI model based on dynamic ultrasound videos may yield a better classification performance than static images. However, the analysis of ultrasound videos using AI is technically difficult. We applied the ResNet-18 network structure to train the diagnosis model and adopted tips to make the analysis of ultrasound videos effective in this study. First, we introduced the pre-training of ImageNet and open ultrasound datasets to reduce the overfitting risk of model training based on limited ultrasound videos. In addition, we applied an attention mechanism to learn the feature importance among frames, which has proven to be an effective model-building strategy. The attention module is a type of deep learning model structure, which is an example of a biomimetic human-eye attention mechanism. When doctors make a diagnosis by reading an ultrasound video, a common practice is to have a quick glance through the entire ultrasound video, identifying several important parts that have a strong link with the disease they are concerned with. They mainly focused on these important ultrasound features and made the final diagnosis. The above process is so common and effective that there is a deep learning model structure called the attention module, which attempts to simulate it to make the deep learning model recognize key parts automatically. In this study, we designed a deep learning diagnosis model to simulate a human's attention ability and make a reliable and accurate diagnosis. Furthermore, we combined the detection and diagnosis models in a proper manner to make a fully automatic thyroid nodule diagnosis, while allowing the adjustment of radiologists if necessary. Compared with previous studies,[5,20–23] our fully automatic AI system obtained better diagnostic results by focusing on dynamic ultrasound videos combined with a deep learning method, which can provide more information on thyroid nodules and make the system more reliable and accurate. It is worth noting that the process of capturing ultrasound videos is fairly easy to implement in routine thyroid ultrasonography examinations. Therefore, in this study, an established fully automatic detection and diagnosis system based on AI technology that analyzes ultrasound video sequences is highly practical in clinical applications. The proposed system can potentially avoid unnecessary biopsies or surgeries, which may have a positive impact on patients and hospital services.

Limitations

This study had several major limitations. First, this prospective study had a single-center initial design. Multi-center data are required to further verify the clinical application of the proposed fully automatic AI detection and diagnosis system. Second, the subtypes of malignant thyroid nodules were imbalanced in this study, with a majority of papillary carcinomas and fewer cases of other carcinoma subtypes. However, this finding is consistent with the practical clinical situations. Third, the number of nodules in the excluded patients was relatively small compared with the total number of nodules in the patients included in the study. More patients in the exclusion range should be collected in the future. Meanwhile, we refined the network structure and adapted the transfer learning method to apply the established accurate model to the excluded populations. Fourth, even if the diagnostic performance of the AI diagnostic model was superior to that of junior radiologists with statistical significance, its superimposable performances with those of senior radiologists do not reach statistical significance. Fifth, the classification of thyroid nodules is imbalanced in practical clinical situations during surgery, with fewer cases of benign nodules. Last, because of time constraints and the variety of locations and shapes, we did not analyze whether locations and shapes of the nodules would affect the performance of the detection model.

Conclusion

In conclusion, this study established a fully automatic detection and diagnosis system for thyroid nodules based on dynamic ultrasound video sequences, using an artificial intelligence approach that can be conveniently applied to optimize the management of patients with thyroid nodules. Methodologically, the overall stereo information of the nodules in the ultrasound video sequences was extracted using our models to obtain more accurate and reliable diagnostic results. Consequently, the proposed AI system holds great potential for popularization in future clinical application.

Ethics Approval and Consent to Participate

This study was approved by the Institutional Review Board of The Second Affiliated Hospital of Nanchang University.

Acknowledgments

Author Contributions

All authors participated in the drafting, revising, and critical review of the article. They gave final approval on the manuscript to be published. They also agreed on the journal to which the article has been submitted, and they all made a significant contribution to the work reported, whether that be in the conception, study design, execution, acquisition of data, analysis, and interpretation, or in all these areas.

Funding

Disclosure

All authors have no conflicts of interest to declare.

References

1. Durante C, Grani G, Lamartina L. et al. The diagnosis and management of thyroid nodules: a review. *JAMA*. 2018;319(9):914–924. doi:10.1001/jama.2018.0898
2. Haugen BR, Alexander EK, Bible KC, et al. 2015 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*. 2016;26(1):1–133. doi:10.1089/thy.2015.0020
3. Moon WJ, Jung SL, Lee JH, et al. Benign and malignant thyroid nodules: US differentiation--multicenter retrospective study. *Radiology*. 2008;247:762–770.
4. Remonti LR, Kramer CK, Leitão CB, et al. Thyroid ultrasound features and risk of carcinoma: a systematic review and meta-analysis of observational studies. *Thyroid*. 2015;25(5):538–550. doi:10.1089/thy.2014.0353
5. 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(3):695–701. doi:10.1148/radiol.2019181343
6. Hoang JK, Middleton WD, Farjat AE, et al. Interobserver variability of sonographic features used in the American College of radiology thyroid imaging reporting and data system. *Am J Roentgenol*. 2018;211(1):162–167. doi:10.2214/AJR.17.19192
7. Shin JH, Baek JH, Chung J, et al. Korean Society of Thyroid Radiology (KSThR) and Korean Society of Radiology. Ultrasonography Diagnosis and Imaging-based management of thyroid nodules: revised Korean society of thyroid radiology consensus statement and recommendations. *Korean J Radiol*. 2016;17(3):370–395. doi:10.3348/kjr.2016.17.3.370
8. 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(5):587–595. doi:10.1016/j.jacr.2017.01.046
9. 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(5):225–237. doi:10.1159/000478927
10. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5(1):4006. doi:10.1038/ncomms5006
11. Wang S, Liu Z, Rong Y, et al. Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer. *Radiother Oncol*. 2019;132:171–177. doi:10.1016/j.radonc.2018.10.019
12. Zhang Q, Peng Y, Liu W, et al. Radiomics based on multimodal MRI for the differential diagnosis of benign and malignant breast lesions. *J Magn Reson Imaging*. 2020;52(2):596–607. doi:10.1002/jmri.27098
13. Zhang HX, Sun ZQ, Cheng YG, et al. A pilot study of radiomics technology based on X-ray mammography in patients with triple-negative breast cancer. *J Xray Sci Technol*. 2019;27(3):485–492. doi:10.3233/XST-180488
14. Zhang Q, Xiao Y, Dai W, et al. Deep learning based classification of breast tumors with shear-wave elastography. *Ultrasonics*. 2016;72:150–157. doi:10.1016/j.ultras.2016.08.004
15. Kuo CC, Chang CM, Liu KT, et al. Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning. *NPJ Digit Med*. 2019;2(1):29. doi:10.1038/s41746-019-0104-2
16. Wang K, Lu X, Zhou H, et al. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. *Gut*. 2019;68(4):729–741. doi:10.1136/gutjnl-2018-316204
17. Ta CN, Kono Y, Eghtedari M, et al. Focal liver lesions: computer-aided diagnosis by using contrast-enhanced US cine recordings. *Radiology*. 2018;286(3):1062–1071. doi:10.1148/radiol.2017170365
18. Liu D, Liu F, Xie X, et al. Accurate prediction of responses to transarterial chemoembolization for patients with hepatocellular carcinoma by using artificial intelligence in contrast-enhanced ultrasound. *Eur Radiol*. 2020;30(4):2365–2376. doi:10.1007/s00330-019-06553-6
19. Liu F, Liu D, Wang K, et al. Deep learning radiomics based on contrast-enhanced ultrasound might optimize curative treatments for very-early or early-stage hepatocellular carcinoma patients. *Liver Cancer*. 2020;9(4):397–413. doi:10.1159/000505694

20. 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(4):546–552. doi:10.1089/thy.2016.0372

21. Wu H, Deng Z, Zhang B, et al. Classifier model based on machine learning algorithms: application to differential diagnosis of suspicious thyroid nodules via sonography. *AJR Am J Roentgenol*. 2016;207(4):859–864. doi:10.2214/AJR.15.15813

22. Ma J, Wu F, Zhu J, et al. A pre-trained convolutional neural network based method for thyroid nodule diagnosis. *Ultrasonics*. 2017;73:221–230. doi:10.1016/j.ultras.2016.09.011

23. Zhang B, Tian J, Pei S, et al. Machine learning-assisted system for thyroid nodule diagnosis. *Thyroid*. 2019;29(6):858–867. doi:10.1089/thy.2018.0380

24. Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell*. 2017;39(6):1137–1149. doi:10.1109/TPAMI.2016.2577031

25. Peng S, Liu Y, Lv W, et al. Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study. *Lancet Digit Health*. 2021;3(4):e250–e259. doi:10.1016/S2589-7500(21)00041-8

26. Yamashita R, Kapoor T, Alam MN, et al. Toward reduction in false-positive thyroid nodule biopsies with a deep learning-based risk stratification system using US cine-clip images. *Radiol Artif Intell*. 2022;4(3):e210174. doi:10.1148/ryai.210174

27. Wang ZW, She Q, Aljosa S. ACTION-net: multipath excitation for action recognition. IEEE Conference on Computer Vision and Pattern Recognition; 2021.