Original Article

Thyroid ultrasound image classification using a convolutional neural network

Yi-Cheng Zhu[1#], Peng-Fei Jin[2#], Jie Bao[2], Quan Jiang[3], Ximing Wang[2]

[1]First Clinical Medical College, Soochow University, Suzhou, China; [2]Department of Radiology, First Affiliated Hospital of Soochow University, Suzhou, China; [3]Pudong New Area People's Hospital Affiliated to Shanghai University of Medicine and Health Sciences, Shanghai, China

Contributions: (I) Conception and design: YC Zhu; (II) Administrative support: J Bao; (III) Provision of study materials or patients: YC Zhu, PF Jin; (IV) Collection and assembly of data: Q Jiang; (V) Data analysis and interpretation: YC Zhu, X Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Ximing Wang. Department of Radiology, First Affiliated Hospital of Soochow University, Suzhou, China.
Email: wangximing1998@163.com.

Background: Ultrasound (US) is widely used in the clinical diagnosis of thyroid nodules. Artificial intelligence-powered US is becoming an important issue in the research community. This study aimed to develop an improved deep learning model-based algorithm to classify benign and malignant thyroid nodules (TNs) using thyroid US images.

Methods: In total, 592 patients with 600 TNs were included in the internal training, validation, and testing data set; 187 patients with 200 TNs were recruited for the external test data set. We developed a Visual Geometry Group (VGG)-16T model, based on the VGG-16 architecture, but with additional batch normalization (BN) and dropout layers in addition to the fully connected layers. We conducted a 10-fold cross-validation to analyze the performance of the VGG-16T model using a data set of gray-scale US images from 5 different brands of US machines.

Results: For the internal data set, the VGG-16T model had 87.43% sensitivity, 85.43% specificity, and 86.43% accuracy. For the external data set, the VGG-16T model achieved an area under the curve (AUC) of 0.829 [95% confidence interval (CI): 0.770–0.879], a radiologist with 15 years' working experience achieved an AUC of 0.705 (95% CI: 0.659–0.801), a radiologist with 10 years' experience achieved an AUC of 0.725 (95% CI: 0.653–0.797), and a radiologist with 5 years' experience achieved an AUC of 0.660 (95% CI: 0.584–0.736).

Conclusions: The VGG-16T model had high specificity, sensitivity, and accuracy in differentiating between malignant and benign TNs. Its diagnostic performance was superior to that of experienced radiologists. Thus, the proposed improved deep-learning model can assist radiologists to diagnose thyroid cancer.

Keywords: Thyroid cancer; ultrasonography; deep learning; visual geometry group

Submitted Jul 30, 2021. Accepted for publication Sep 16, 2021.
doi: 10.21037/atm-21-4328
View this article at: https://dx.doi.org/10.21037/atm-21-4328

Introduction

Thyroid nodules (TNs) are a common endocrine problem that can be found in most healthy people (1-3). An ultrasound (US) is a critical diagnostic method for assessing malignant possibility, which provides valuable information for preliminary decision making. An US is commonly used together with fine-needle aspiration (FNA) to achieve better diagnostic accuracy (1-3). However, it is challenging for inexperienced radiologists to accurately identify and interpret US features consistently, and there is moderate

to significant inter-observer and intra-observer differences (4-8). In addition, the classification of thyroid cancer by US is quite time consuming and labor intensive.

In recent decades, machine learning-based solutions have developed and attracted a great deal of attention in medical imaging studies. Medical images contain key information that reflect the underlying physiology of tumors that may be imperceptible to the human eye (9,10). Recently, convolutional neural network (CNN) algorithms have had success in analyzing medical imaging classification tasks, including diabetic retinopathy categorization (11), and skin lesion assessments (12), and in the monitoring of acute neurologic events (13). The potential types and numbers of deep-learning algorithms have also boomed in recent decades (14). Traditional feature extraction methods require a clear understanding and a comprehensive design to define hand-crafted features (15,16). Conversely, CNN models automatically derive and optimize features from data sets based on a defined objective (15,16). In 2014, Simonyan and Zisserman developed a CNN model called the Visual Geometry Group (VGG)-16 model with an accuracy of 92.7% in ImageNet (17).

This study aimed to develop an improved VGG-16 model based on a CNN algorithm to differentiate between benign and malignant TNs automatically in US images. The secondary objective was to compare the results of the machine-learning model with those of junior and senior radiologists.

We present the following article in accordance with the STARD reporting checklist (available at https://dx.doi.org/10.21037/atm-21-4328).

Methods

This retrospective study was approved by the Institutional Review Board of Pudong New Area People's Hospital Affiliated to Shanghai University of Medicine and Health Sciences (the Hospital), and patients' informed consent was waived due to the nature of this retrospective study. All procedures performed in this study involving human participants were in accordance with the Declaration of Helsinki (as revised in 2013).

Study population

Between March 2018 and May 2020, 832 patients, who had been referred to the Hospital for thyroid US examinations, were enrolled in this study. Patients were excluded from the study if they met any of the following exclusion criteria: (I) had no pathological results; (II) had no clear US images; and/or (III) had a history of treatment before the US examination. All the included TNs underwent fine-needle aspiration biopsy (FNAB) and/or surgery pathology within 1 month of the US examinations. Patients with TNs with benign cytologic results were followed-up with for more than 6 months. Ultimately, 592 patients with 600 TNs were included in the internal training, validation, and testing data set with a balanced random selection of 50% benign and 50% malignant TNs. The rest of the patients (n=187) with 200 TNs were included in the external test data set (see *Figure 1*).

US scanning

All thyroid gray-scale US examinations were performed in the Hospital using Siemens S3000, GE Vivid E95, Mindray Resona 7, Philips Epic 7, and Toshiba Apolio 500 US machines with a high-frequency linear probe of 10 MHz. These machines are commonly used to capture US images in real clinical practice. US images from different machines were included to ensure the robustness of the trained CNN model. Both longitudinal and transverse planes of the TNs were recorded by 1 radiologist with more than 10 years' working experience. For example, among the lesions used to develop the CNN model, 299 (49.8%) transverse planes of TNs were obtained. All the images were acquired and stored in red, green and blue (RGB) format.

Building a CNN model

To expand the traditional artificial neural network, the CNN used an additional convolutional layer to optimize the feature used with the best fit to the learning objective. By using the power of supervised learning, the introduction of a convolutional layer allows a classifier to efficiently search for the optimal feature in a large feature space, which provides a great advantage in image classification and has quickly become one of the most popular tools in machine learning. Specifically, convolutional layers consist of a list of filters and activation functions with pooling. The weights of these functions are updated iteratively through back propagation by minimizing the loss between the end value and the actual expectation.

By nature, a CNN model, especially the convolutional layer, requires an enormous amount of data for optimization. However, medical data are known to be

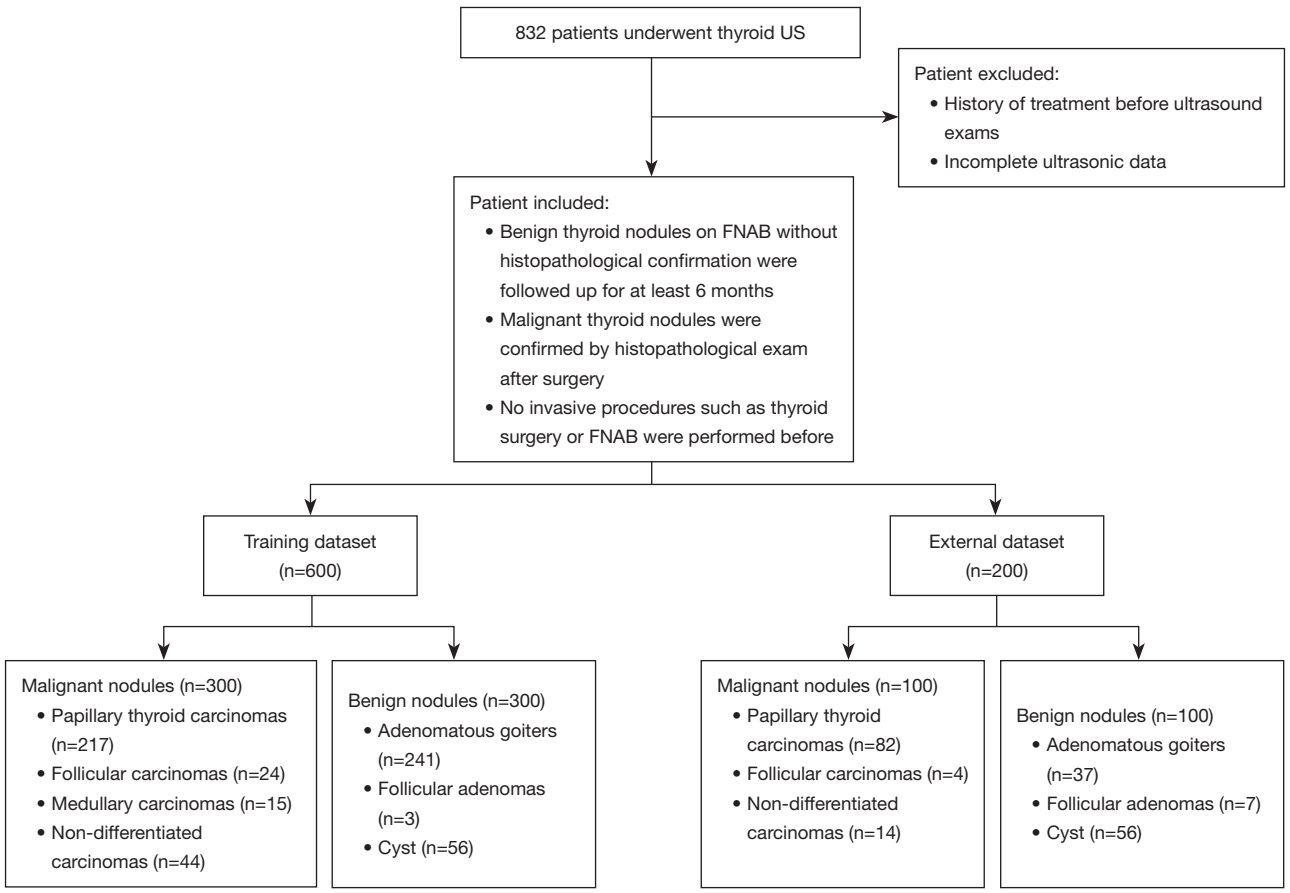

Figure 1 Study workflow of the population. US, ultrasound; FNAB, fine-needle aspiration biopsy.

difficult to collect due the quality and ethical issues involved. Thus, we adopted a geometrical method and singular value decomposition (SVD) method to augment the training data set. The geometric method used rotations and mirrors of the original images to mimic images of different variations. The images were rotated 90, 180, and 270 degrees in our experiment (see *Figure 2*). The SVD method uses the principle of SVD to project the image into another oriented space, which maintains the general appearance of the image but with a slight difference in the values.

The VGG-16T model

The VGG-16 architecture comprises 24 layers (including the flattening and dropout layers), which include 13 convolution layers, 5 pooling layers, and 3 fully connected layers with dropouts in between (for details of the structure and parameters, see *Table 1*). All the convolution layers used the rectified linear unit (ReLU) activation function, which sets the value to 0 for negative input values. Our proposed method was based on the VGG-16 architecture but included additional batch normalization (BN) and dropout layers in addition to the fully connected layers (see *Figure 3*). In the neural network accelerator community (18,19) and other deep-learning models (20-22), BN and the ReLU are commonly used to train deep-learning models (23). A BN can improve training by clearly normalizing inputs of all layers to have 0 means and unit variances. To ensure appropriate distributions, a BN can also lower the difficulty of annealing training rates and initializing parameters. Thus, a BN can improve and accelerate CNN models.

The pooling layers were 2×2 with a stride size of 2, while the convolutional layers were 3×3 with a stride size of 1. The default input image size was 224×224. However, the size of the feature map was reduced by 50% after each pooling layer. The last feature map before the fully connected layers was 7×7, had 512 channels, and was flattened to a 25,088 (7×7×512) dimensional vector.

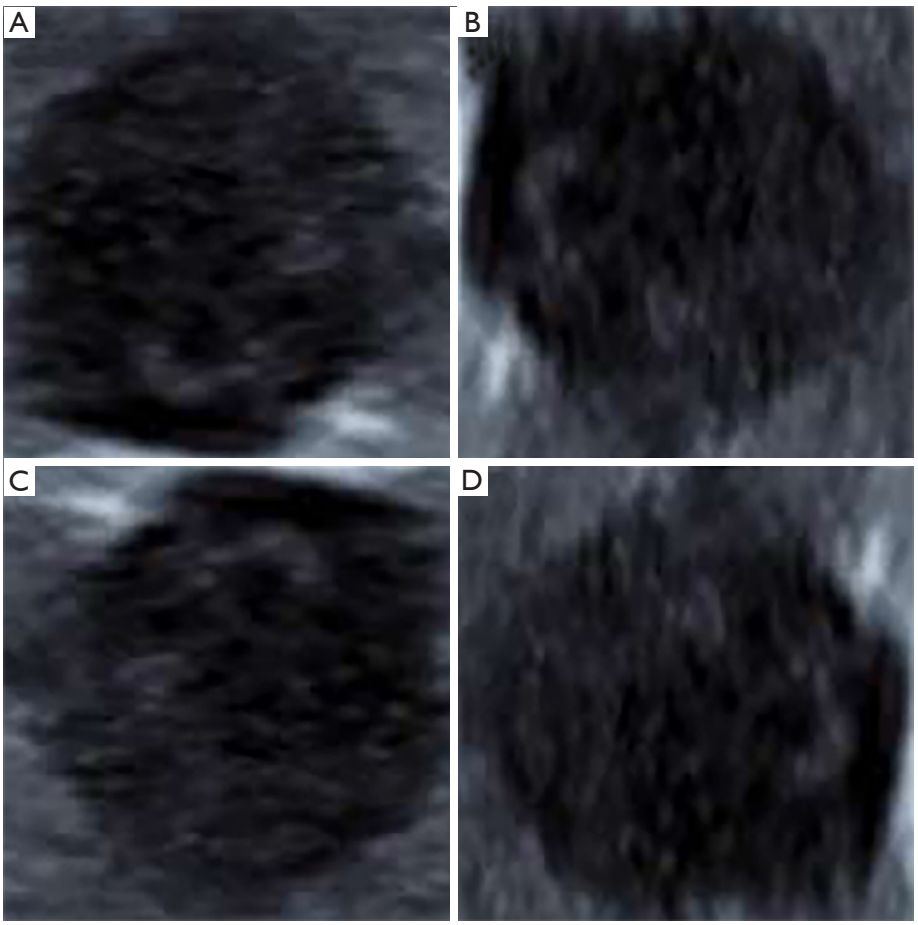

Figure 2 Geometric method of a rotated thyroid US image. (A) The original cropped square of a thyroid nodule; (B) the same image rotated 90° clockwise; (C) the same image rotated 180° clockwise; (D) the same image rotated 270° clockwise. US, ultrasound.

The original output of the VGG-16T model comprised 1,000 classes. To fit the objective of our current research, we modified it into 2 classes (i.e., the benign and malignant TNs classes). We conducted a 10-fold cross-validation to analyze the performance of the VGG-16T model. Ninety percent of the data set was used for training and validation, while the remaining 10% was used to test each fold.

Statistical analysis

We compared the performance of the VGG-16 model, the VGG-16T model, and the radiologists in diagnosing thyroid cancer using receiver operating characteristics (ROCs) curves and the areas under the curves (AUCs). Cytologic and pathologic results were used as gold standards. SPSS (version 24.0, SPSS Inc., Chicago, IL, USA) was used for the data analysis. P values less than 0.05 were considered statistically significant.

Results

Basic characteristics of patients and nodules

In total, 800 TNs (training cohort: n=600; external test cohort: n=200) were found in 779 patients, including 400 (50.0%) benign and 400 (50.0%) malignant TNs. Among the 779 patients, 212 were male and 567 were female; the mean age of the patients was 54.6±6.2 years (range, 28–67 years). The 300 malignant TNs in the training cohort comprised 217 papillary thyroid carcinomas, 24 follicular carcinomas, 15 medullary carcinoma, and 44 nondifferentiated carcinomas. The 300 benign TNs in the training cohort comprised 241 adenomatous goiters, 3 follicular adenomas, and 56 cysts (see *Table 2*).

Table 1 Structure of the VGG-16 model

Layer name	Output size
conv2d_1	224×224×64
conv2d_2	224×224×64
max_pooling2d_1	112×112×64
conv2d_3	112×112×128
conv2d_4	112×112×128
max_pooling2d_2	56×56×128
conv2d_5	56×56×256
conv2d_6	56×56×256
conv2d_7	56×56×256
max_pooling2d_3	28×28×256
conv2d_8	28×28×512
conv2d_9	28×28×512
conv2d_10	28×28×512
max_pooling2d_4	14×14×512
conv2d_11	14 x14×512
conv2d_12	14×14×512
conv2d_13	14 x14×512
max_pooling2d_5	7×7×512
flatten_1	25,088
dense_1	4,096
dropout_1	4,096
dense_2	4,096
dropout_2	4,096
dense_3	2

VGG, Visual Geometry Group.

Diagnostic performance of the VGG-16T model in the training and external test cohorts

The results for the training data set showed that the VGG-16 model had 81.67% sensitivity, 81.00% specificity, and 81.34% accuracy. The improved VGG-16T algorithm had 87.43% sensitivity, 85.43% specificity, and 86.43% accuracy. Among the 10-fold results, Model 7 had the highest sensitivity (92.33%), specificity (90.33%), and accuracy (91.33%) (see *Table 3*). Thus, Model 7 was used on the external cohorts. The VGG-16T model in the external cohorts had 85.00% sensitivity, 79.00% specificity, and 82.00% accuracy.

Comparison between the VGG-16T model and human observers

The 3 radiologists (1 with 5 years' experience, 1 with 10 years' experience, and 1 with 15 years' experience), who were blind to the cytology data, performed differential diagnoses using the US images of the external cohort. In the external set, the AUC of the VGG-16T model was 0.829 [95% confidence interval (CI): 0.779–0.879]. The VGG-16T model achieved the highest sensitivity of 85.00%, followed by the radiologist with 10 years' experience (76.00%), the radiologist with 15 years' experience (75.00%), and the radiologist with 5 years' experience (71.00%) (see *Table 4*). Additionally, the VGG-16T model achieved the highest specificity (79.00%), followed by the radiologist with 15 years' experience (71.00%), the radiologist with 10 years' experience (69.00%), and the radiologist with 5 years' experience (61.00%) (see *Figure 4*). On average, the VGG-16T model performed significantly better than the 3 radiologists, especially in terms of sensitivity.

Discussion

It is widely believed that artificial intelligence (AI) will not be able to take the place of radiologists in the upcoming years. However, medical experts can still benefit from AI-based solutions to extend their understandings of TNs in their routine practice. Previous studies have adopted different classifier models, such as naïve bayes classifiers, support vector machines, and histogram analyses, in thyroid US imaging to increase classification accuracy (24,25). Additionally, the findings of the present study provided further evidence that a CNN-based solution could also be applied to US thyroid diagnoses. In the internal cohort, the VGG-16T model had a sensitivity of 87.43%, a specificity of 85.43%, and an accuracy of 86.43%. Compared to the junior and senior radiologists, the VGG-16T model was superior at differentiating between benign and malignant TNs. The proposed algorithm indicates that adding a BN before the ReLU can accelerate and improve the training results. The VGG-16T model had a higher sensitivity (81.67%), a higher specificity (81.00%), and a higher accuracy than the VGG-16 model (81.34%).

The CNN, which was first proposed by Fukushima (26), is a supervised learning model that captures high-dimensional non-linear mappings between input and output. InceptionNet, VGGNet, and AlexNet are examples of

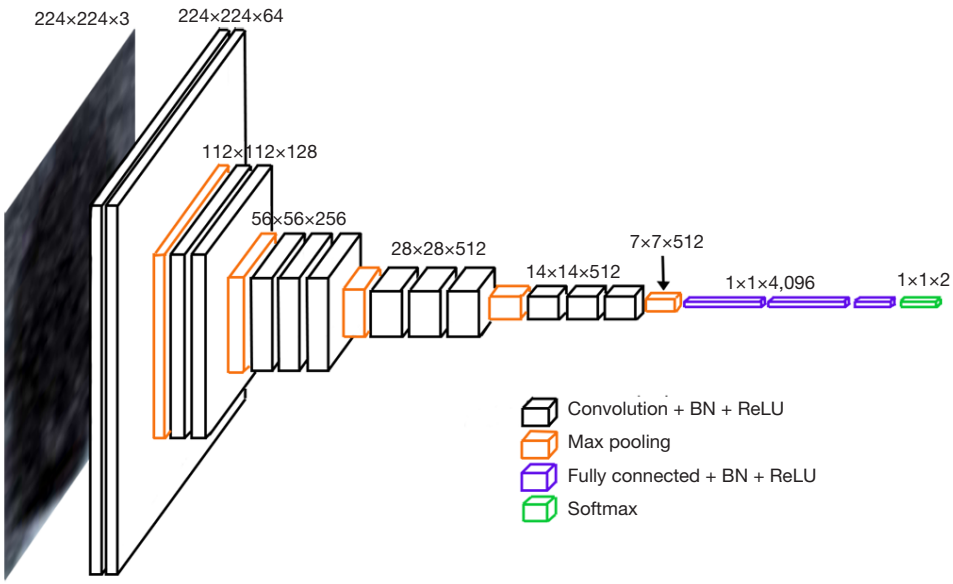

Figure 3 VGG-16T architecture. VGG-16T, Visual Geometry Group-16T; BN, batch normalization; ReLU, rectified linear unit.

Table 2 Baseline characteristics of the study population

Characteristics	Training		Testing	
	Malignant	Benign	Malignant	Benign
Patients (years old)	52.3±8.9	58.1±11.3	63.6±17.2	58.0±7.3
Planes of US images				
Longitudinal	163	179	40	39
Transverse	137	121	60	61
US machine types				
Siemens	49	65	25	9
GE	82	59	34	25
Mindray	18	60	18	40
Philips	122	15	13	6
Toshiba	29	101	10	20

US, ultrasound; GE, GE Healthcare.

deep-learning models that have been introduced into image classification. The VGG-16 model is a CNN model that has been said to perform better than the Inception v3 model in classifying medical images (27,28). Ye *et al.* (29) proposed a non-segmentation algorithm, VGG-16 model, to classify 740 and 861 benign TNs. Their model had a sensitivity of 87%, a specificity of 85.32%, and an accuracy of 86.12%. Compared to their study, the boundary of TNs in this study were cropped by a radiologist, which increase the accuracy of the TN boundaries. This may have contributed to the slightly higher sensitivity (87.43%) and specificity (85.43%) of our results. Zhou *et al.* (30) collected 1750 TNs and 3852 TNs from 2 sources to build online transfer learning (OTL) models to classify malignant and benign TNs, one of which was a VGG-16-based transfer learning model. That model achieved a sensitivity of 74.6% (validation: 76.7%) and a

Table 3 10-fold cross-validation results of the internal data set

Internal data set	Sensitivity (%)	Specificity (%)	Accuracy (%)
Model 1	88.00	95.00	91.50
Model 2	89.00	83.33	84.83
Model 3	87.00	89.67	88.33
Model 4	88.33	87.00	87.67
Model 5	85.67	80.33	80.33
Model 6	81.00	83.33	82.17
Model 7	92.33	90.33	91.33
Model 8	86.67	81.67	82.00
Model 9	86.33	79.67	83.00
Model 10	90.00	84.00	87.00

Table 4 Diagnostic performance of VGG-16T and radiologists

	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC	95% CI
Radiologist (5 y)	71.00	61.00	66.00	0.660	0.584–0.736
Radiologist (10 y)	76.00	69.00	72.50	0.725	0.653–0.797
Radiologist (15 y)	75.00	71.00	73.00	0.730	0.659–0.801
Radiologist (Avg)	74.00	67.00	70.50	0.705	0.632–0.778
VGG-16T	85.00	79.00	82.00	0.829	0.770–0.879

VGG-16T, Visual Geometry Group-16T. AUC, the areas under the curves.

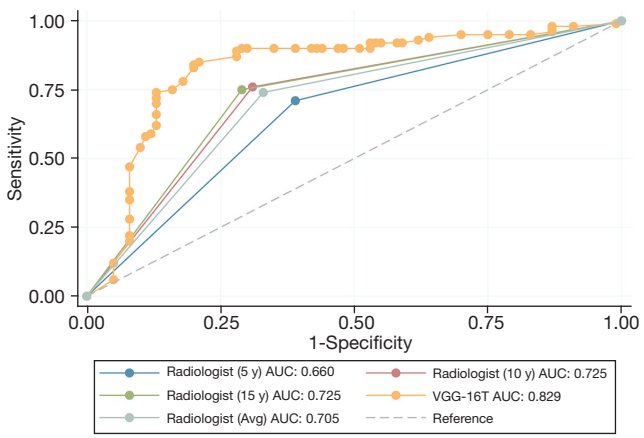

Figure 4 ROCs for the VGG-16T model and radiologists. VGG-16T, Visual Geometry Group-16T; ROCs, receiver operating characteristics; AUC, the areas under the curves.

specificity of 79.8% (validation: 79.5%). Our model achieved a higher sensitivity (training: 87.43%, external testing: 85.00%) and a comparable specificity (training: 85.43%, external testing: 79.00%); however, this may be due to the more balanced training samples in the present study. To the best of our knowledge, this is the first study to develop an improved VGG-16-based algorithm to classify thyroid cancer.

The introduction of a deep-learning method in thyroid cancer prediction would provide number of benefits in clinical practices. First, our CNN-based solution produced consistent predictions for the same images. Thus, substantial inter-observer variance among radiologists could be eliminated. Second, the application of a deep-learning algorithm could accelerate the diagnosis of TNs, which could increase the daily diagnostic capacity of US departments in China. Finally, our small-scale

results showed that the VGG-16T model outperformed radiologists, but a larger study needs to be conducted to confirm these results.

Our study had some limitations. First, we did not investigate correlations between the accuracy of the proposed deep-learning algorithm and TN size or cancer subtypes. Second, other ultrasound techniques also are applied to classify thyroid cancer, such as elastography. It will be our further work to add different ultrasound techniques combined with deep learning algorithms. Second, the number of samples used in this study was relatively low. A larger number of sample are required for future studies. Finally, the accuracy of the developed algorithm needs to be verified and enhanced.

Conclusions

Our in-house proposed VGG-16T model showed high specificity, sensitivity, and accuracy in differentiating between malignant and benign TNs. Additionally, its diagnostic performance was superior to that of experienced radiologists. The results of this pilot study are promising and indicate that deep-learning methods could assist clinicians in the classification of thyroid cancer.

Acknowledgments

Funding: This work was supported by the AI innovation and development project of Shanghai Municipal Commission of Economy and Informatization (grant number: 2020-RGZN-02044).

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at https://dx.doi.org/10.21037/atm-21-4328

Data Sharing Statement: Available at https://dx.doi.org/10.21037/atm-21-4328

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://dx.doi.org/10.21037/atm-21-4328). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This retrospective study was approved by the Institutional Review Board of Pudong New Area People's Hospital Affiliated to Shanghai University of Medicine and Health Sciences (the Hospital), and patients' informed consent was waived due to the nature of this retrospective study. All procedures performed in this study involving human participants were in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Haugen BR. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: What is new and what has changed? Cancer 2017;123:372-81.
2. 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.
3. Ha EJ, Baek JH, Na DG. Risk Stratification of Thyroid Nodules on Ultrasonography: Current Status and Perspectives. Thyroid 2017;27:1463-8.
4. Choi SH, Kim EK, Kwak JY, et al. Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules. Thyroid 2010;20:167-72.
5. Kim HG, Kwak JY, Kim EK, et al. Man to man training: can it help improve the diagnostic performances and interobserver variabilities of thyroid ultrasonography in residents? Eur J Radiol 2012;81:e352-6.
6. Park CS, Kim SH, Jung SL, et al. Observer variability in the sonographic evaluation of thyroid nodules. J Clin Ultrasound 2010;38:287-93.
7. Park SH, Kim SJ, Kim EK, et al. Interobserver agreement in assessing the sonographic and elastographic features of malignant thyroid nodules. AJR Am J Roentgenol 2009;193:W416-23.

8. 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.
9. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology 2016;278:563-77.
10. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48:441-6.
11. Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 2016;316:2402-10.
12. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115-8.
13. Titano JJ, Badgeley M, Schefflein J, et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. Nat Med 2018;24:1337-41.
14. Dargan S, Kumar M, Ayyagari MR, et al. A survey of deep learning and its applications: A new paradigm to machine learning. Arch Comput Methods Eng 2020;27:1071-92.
15. Shin HC, Roth HR, Gao M, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Trans Med Imaging 2016;35:1285-98.
16. Anwar SM, Majid M, Qayyum A, et al. Medical Image Analysis using Convolutional Neural Networks: A Review. J Med Syst 2018;42:226.
17. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. The 3rd International Conference on Learning Representations (ICLR2015). Available online: https://arxiv.org/abs/1409.1556
18. Zhao R, Song W, Zhang W, et al. Accelerating binarized convolutional neural networks with software-programmable FPGAs. Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays. February 2017:15-24. Available online: hhttps://doi.org/10.1145/3020078.3021741
19. Jiang L, Kim M, Wen W, et al. XNOR-POP: A processing in-memory architecture for binary convolutional neural networks in wide-IO2 DRAMS. Jul 24, 2017 in ISLPED (International Symposium on Low Power Electronics and Design). doi: 10.1109/ISLPED.2017.8009163.
20. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun 2016:770-8.
21. Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun 2017:1-4.
22. Howard AG, Zhu M, Chen B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications. (2017). Available online: https://arxiv.org/abs/1704.04861
23. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 2010:807-14.
24. 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;207:859-64.
25. Nam SJ, Yoo J, Lee HS, et al. Quantitative Evaluation for Differentiating Malignant and Benign Thyroid Nodules Using Histogram Analysis of Grayscale Sonograms. J Ultrasound Med 2016;35:775-82.
26. Fukushima K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern 1980;36:193-202.
27. Guan Q, Wang Y, Ping B, et al. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study. J Cancer 2019;10:4876-82.
28. Yadav SS, Jadhav SM. Deep convolutional neural network based medical image classification for disease diagnosis. Journal of Big Data 2019;6:1-18.
29. Ye H, Hang J, Chen X, et al. An intelligent platform for ultrasound diagnosis of thyroid nodules. Sci Rep 2020;10:13223.
30. Zhou H, Wang K, Tian J. Online Transfer Learning for Differential Diagnosis of Benign and Malignant Thyroid Nodules With Ultrasound Images. IEEE Trans Biomed Eng 2020;67:2773-80.

(English Language Editor: L. Huleatt)

Cite this article as: Zhu YC, Jin PF, Bao J, Jiang Q, Wang X. Thyroid ultrasound image classification using a convolutional neural network. Ann Transl Med 2021;9(20):1526. doi: 10.21037/atm-21-4328

