EnM
ENDOCRINOLOGY
AND METABOLISM

Original Article

Endocrinol Metab 2025;40:216-224
https://doi.org/10.3803/EnM.2024.2058
pISSN 2093-596X · eISSN 2093-5978

Deep Learning Technology for Classification of Thyroid Nodules Using Multi-View Ultrasound Images: Potential Benefits and Challenges in Clinical Application

Jinyoung Kim[1], Min-Hee Kim[1], Dong-Jun Lim[1], Hankyeol Lee[2], Jae Jun Lee[3], Hyuk-Sang Kwon[1], Mee Kyoung Kim[1], Ki-Ho Song[1], Tae-Jung Kim[4], So Lyung Jung[5], Yong Oh Lee[3], Ki-Hyun Baek[1]

[1]Division of Endocrinology and Metabolism, Department of Internal Medicine, College of Medicine, The Catholic University of Korea; Departments of [2]Computer Engineering, [3]Industrial and Data Engineering, Hongik University; Departments of [4]Hospital Pathology, [5]Radiology, College of Medicine, The Catholic University of Korea, Seoul, Korea

Background: This study aimed to evaluate the applicability of deep learning technology to thyroid ultrasound images for classification of thyroid nodules.

Methods: This retrospective analysis included ultrasound images of patients with thyroid nodules investigated by fine-needle aspiration at the thyroid clinic of a single center from April 2010 to September 2012. Thyroid nodules with cytopathologic results of Bethesda category V (suspicious for malignancy) or VI (malignant) were defined as thyroid cancer. Multiple deep learning algorithms based on convolutional neural networks (CNNs) —ResNet, DenseNet, and EfficientNet—were utilized, and Siamese neural networks facilitated multi-view analysis of paired transverse and longitudinal ultrasound images.

Results: Among 1,048 analyzed thyroid nodules from 943 patients, 306 (29%) were identified as thyroid cancer. In a subgroup analysis of transverse and longitudinal images, longitudinal images showed superior prediction ability. Multi-view modeling, based on paired transverse and longitudinal images, significantly improved the model performance; with an accuracy of 0.82 (95% confidence intervals [CI], 0.80 to 0.86) with ResNet50, 0.83 (95% CI, 0.83 to 0.88) with DenseNet201, and 0.81 (95% CI, 0.79 to 0.84) with EfficientNetv2_s. Training with high-resolution images obtained using the latest equipment tended to improve model performance in association with increased sensitivity.

Conclusion: CNN algorithms applied to ultrasound images demonstrated substantial accuracy in thyroid nodule classification, indicating their potential as valuable tools for diagnosing thyroid cancer. However, in real-world clinical settings, it is important to aware that model performance may vary depending on the quality of images acquired by different physicians and imaging devices.

Keywords: Artificial intelligence; Deep learning; Thyroid nodule; Thyroid neoplasms; Ultrasonography

Received: 3 June 2024, **Revised:** 31 August 2024, **Accepted:** 23 September 2024

Corresponding authors: Ki-Hyun Baek
Division of Endocrinology and Metabolism, Department of Internal Medicine, Yeouido St. Mary's Hospital, College of Medicine, The Catholic University of Korea, 10 63-ro, Yeongdeungpo-gu, Seoul 07345, Korea
Tel: +82-2-3779-1400, **Fax:** +82-2-708-3132, **E-mail:** drbkh@catholic.ac.kr

Yong Oh Lee
Department of Industrial and Data Engineering, Hongik University, 94 Wausan-ro, Mapo-gu, Seoul 04066, Korea
Tel: +82-2-320-1795, **Fax:** +82-2-708-3132, **E-mail:** yongoh.lee@hongik.ac.kr

This manuscript was first published as a preprint: Kim J, Kim MH, Lim DJ, Lee H, Lee JJ, Kwon HS, et al. (2024). [Deep learning technology for classification of thyroid nodules using multi-view ultrasound images: potential benefits and challenges in clinical application]. Research Square. https://doi.org/10.21203/rs.3.rs-5527813/v1.

INTRODUCTION

Thyroid cancer develops at a younger age than other malignant tumors, and is the most common endocrine malignancy [1]. The prevalence of nodular lesions of the thyroid gland, including benign lesions, has been reported to vary from 20% to 70% worldwide [2], and approximately 10% of thyroid nodules are diagnosed as thyroid cancer [3]. Due to the high prevalence of thyroid nodules and the high probability of detecting thyroid cancer, accurately distinguishing and managing thyroid nodules is of critical importance.

For differential diagnosis of thyroid nodules, the standard diagnostic approach is to examine the cytology from the suspected lesion using fine-needle aspiration (FNA) followed by review of pathological findings. However, because FNA is an invasive procedure, interpretation of ultrasound images is necessary for determining whether to biopsy [4]. Risk stratification based on the characteristics of ultrasound images is highly accurate [5-7], suggesting that the presence of cancer can be predicted by analyzing ultrasound images of thyroid nodules.

Meanwhile, artificial intelligence (AI) has emerged as a pivotal tool in diagnostic decision across various medical fields [8,9]. Deep learning models mimic the intricate processes of the human brain, and have been shown to be particularly promising in deciphering the complexities inherent in medical datasets [10]. Convolutional neural networks (CNNs), a type of deep learning, are well recognized in the realm of medical imaging for classification tasks [11]. By transforming images into digitized formats through convolution and pooling layers and refining the data with activation functions and fully-connected layer, CNNs are adept at extracting meaningful patterns from medical images.

In this study, we investigated the clinical applicability of AI to classify thyroid nodules by using several CNN algorithms to ultrasound images of thyroid nodules.

METHODS

Study subjects and diagnosis

Patients who underwent biopsy at the Thyroid Nodule Clinic, Department of Endocrinology, Yeouido St. Mary's Hospital from April 2010 to September 2012 were screened for this study. We included cases in which cross-sectional and longitudinal images of the nodule were clearly identifiable and showed no overlap with other nodules. Patients with pathologic confirmation based on cytology from FNA or surgical specimen during follow-up longer than 3 years were screened. If the FNA result was Bethesda category II (benign), the nodule was classified as benign. Cases diagnosed as Bethesda category V (suspicious for malignancy) or VI (malignant) on FNA were classified as papillary thyroid cancer (PTC). Cases with Bethesda category III (atypia of unknown significance) confirmed on initial FNA were analyzed only if the diagnosis was confirmed by repeated FNA or surgery. Cases with unsatisfactory pathologic results (Bethesda category I), those suspected of being a follicular neoplasm (Bethesda category IV), or those confirmed to be malignancies other than PTC (follicular thyroid cancer, medullary thyroid cancer, or lymphoma) were excluded.

The study protocol was approved by the Institutional Review Board of Yeouido St.Mary's Hospital (Study number SC22RISI0110). Written informed consent was waived because the medical records of patients were retrospectively analyzed.

Data preparation

Board-certified radiologists performed all imaging examinations and FNA. Until April 2011, ultrasound was performed at the study institution with a low-resolution ultrasound device (HDI 5000, Philips Healthcare, Best, Netherlands). After that, the equipment was changed to a high-resolution ultrasound device (iU22, Philips Healthcare). The resolution of the low-resolution images was 640×476 pixels, while that of high-resolution images was 1,024×768 pixels. Before and after the device change, 524 nodules were extracted, and a total of 1,048 nodules were analyzed.

For human reference, an endocrinologist with 5-year of experience (Jinyoung Kim) reviewed the images and stratified the nodules in line with traditional Korean Thyroid Imaging Reporting and Data System (K-TIRADS) scoring. To train deep learning algorithms with the thyroid ultrasound images, nodular lesions were annotated in the form of a bounding box. After obtaining the coordinates of the center of the bounding box, an image of 150 pixels on the top, bottom, left, and right sides was cropped, and a box-shaped image of 300×300 pixels was analyzed (Fig. 1).

For the AI models, augmentation techniques were applied considering the possibility of image deformation under various conditions that may occur during the acquisition of ultrasound images [12]. This standard deep learning image processing was performed in the present study by applying various random transformation methods to images, including; cropping to 256×256 pixels, rotating within 15°, adjusting the brightness or contrast, and flipping horizontally or vertically.

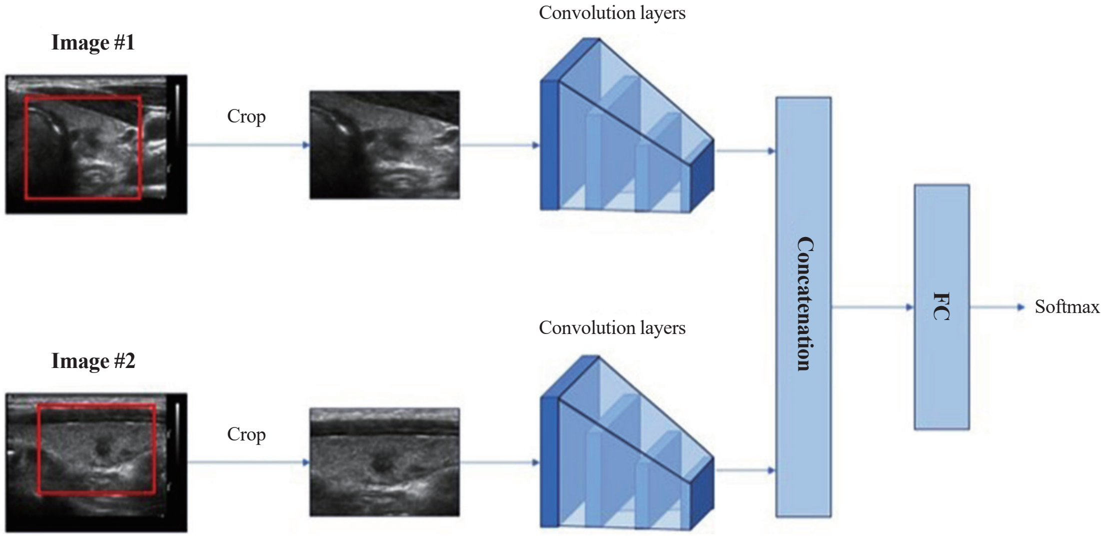

Fig. 1. Study flowchart. FC, fully-connected layers.

Statistical analysis

Continuous variables are reported as using means and standard deviations, while categorical variables are presented as numbers and percentages. Because this study was conducted without an external validation set, five-fold cross-validation was used to evaluate model performance. Model performance was described in terms of sensitivity, specificity, positive predictive value, negative predictive value, and accuracy. This value was calculated based on K-TIRADS category 5 for human reference, and in the AI models, by combining the binary prediction results of five test sets in cross-validation. Area under the curve (AUC) values were calculated based on the four classes (K-TIRADS categories 2–5) for human, and continuous malignancy probability values of the soft-max function for AI models. To compare model performance, DeLong's test was conducted for the AUC values.

Each CNN model was implemented using PyTorch (https://pytorch.org) with the runtime environment configured on an RTX 4,080 graphics processing unit (GPU). The batch size was set to 64 for single-view analysis, and the batch size was set to 32 for multi-view analysis. Epochs were set to 100, and we employed a learning rate scheduler that starts at 0.005 and reduces the learning rate by half if no performance improvement is observed within 20 epochs. Optimizer used was stochastic gradient descent (SGD) with a momentum value of 0.9 to enhance the stability and speed of model training. Weight decay was set to 0.0005 to prevent overfitting by limiting the increase in model weight and controlling the complexity of the model structure.

Statistical analyses were performed with R version 4.3.3 program (R Foundation for Statistical Computing, Vienna, Austria) and Python version 3.13.0 (https://www.python.org).

RESULTS

Baseline characteristics

Among the 943 patients in the study cohort, the average age was 48 years and 74% were female. The number of thyroid nodules analyzed was 1,048, and the average maximal diameter was 1.2 cm. The malignancy rate among all nodules was 29% with rates of 28% in the low-resolution group and 30% in the high-resolution image groups, which was a non-significant difference (Table 1).

Comparisons of cross-sectional images, longitudinal section images and multi-view analysis

Subgroup analyses of transverse and longitudinal images from each nodule showed that the longitudinal section was superior (Fig. 2, Supplemental Tables S1-S6). Multi-view models that analyzed pairs of transverse and longitudinal images had statistically significantly higher performance than when the transverse section was used alone ($P<0.05$).

Comparisons of deep learning models with the traditional K-TIRADS

When examining the classification performance of the traditional K-TIRADS in our study cohort, the sensitivity was 90% for the category 5. We compared the performance of AI-models with the traditional scoring by AUC values, and there was no

Table 1. Baseline Characteristics of the Study Cohort

Variable	1,048 Thyroid nodules in 943 patients
Age, yr	48±12
Sex	
Men	249 (26)
Women	694 (74)
Size, cm	1.2±0.8
K-TIRADS	
2	60 (6)
3	323 (31)
4	370 (35)
5	295 (28)
Initial cytopathology	
Bethesda I	21 (2)
Bethesda II	702 (67)
Bethesda III	67 (6)
Bethesda V, VI	258 (25)
Final diagnosis	
Total images (n=1,048)	
Benign	742 (71)
Malignancy	306 (29)
Low-resolution images (n=524)	
Benign	377 (72)
Malignancy	147 (28)
High-resolution images (n=524)	
Benign	365 (70)
Malignancy	159 (30)

Values are expressed as mean±standard deviation or number (%). K-TIRADS, Korean Thyroid Imaging Reporting and Data System.

significant difference result, especially for DenseNet201. The specificity of the all three AI models was excellent, with AUC greater than 90% (Table 2, Supplemental Tables S7-S9).

Comparisons based on the differences in image resolution

Subgroup analysis was performed on different datasets according to image resolution, reflecting the change in ultrasound imaging equipment over time. The model that learned from high-resolution images had a significantly better performance than the model learned from low-resolution images, and this pattern was consistent for all three CNN algorithms (Fig. 3, Supplemental Tables S10-S15).

DISCUSSION

We designed this study to examine the classification performance of CNN algorithms applied to two-dimensional ultrasound images of thyroid nodules. Improvement in image resolution increased the sensitivity of the model, while refinement of AI algorithms improved the accuracy by increasing specificity. Developing a sufficiently learned and refined deep learning algorithm is expected to be helpful in cancer diagnosis based on image findings.

Among the previous studies in which deep learning technology was applied for differentiating thyroid nodules by ultrasound, the largest-scale study was a multicenter study in China [13]. The researchers named the model ThyNet, and it was made available to the public through GitHub [14]. Considering the advantages and disadvantages of previously developed algorithms, ThyNet combines three algorithms, ResNet, ResNext, and DenseNet, through a weighted voting system that achieved

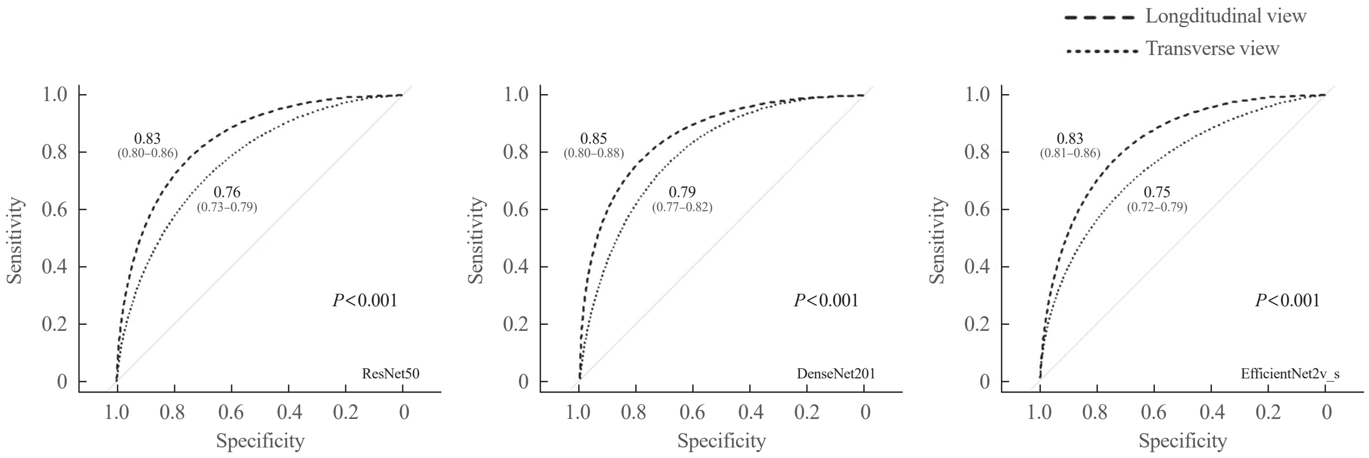

Fig. 2. Comparison of model performance according to image view.

Table 2. Comparison of Model Performance with Traditional Scoring by an Endocrinologist with 5 Years of Experience

	Sensitivity[a]	Specificity[a]	PPV[a]	NPV[a]	Accuracy[a] (95% CI)	AUC[a] (95% CI)	*P* value[a]
Traditional K-TIRADS							
K-TIRADS							
Category 5	0.90	0.73	0.89	0.76	0.85 (0.83–0.87)	0.88[b] (0.86–0.90)	Reference
Multi-view modeling							
ResNet50	0.65	0.90	0.72	0.86	0.82 (0.80–0.85)	0.83 (0.80–0.86)	<0.001[c]
DenseNet201	0.62	0.92	0.76	0.86	0.83 (0.81–0.86)	0.86 (0.83–0.88)	0.107[c]
EfficientNet2v_s	0.58	0.91	0.73	0.84	0.81 (0.79–0.84)	0.78 (0.75–0.81)	<0.001[c]

PPV, positive predictive value; NPV, negative predictive value; CI, confidence interval; AUC, area under the curve; K-TIRADS, Korean Thyroid Imaging Reporting and Data System.
[a]The results were calculated from the result combined with all folds of five-fold validation; [b]Calculated from the K-TIRADS scores, categorized from 2 to 5; [c]Calculated by DeLong's test based on soft-max values.

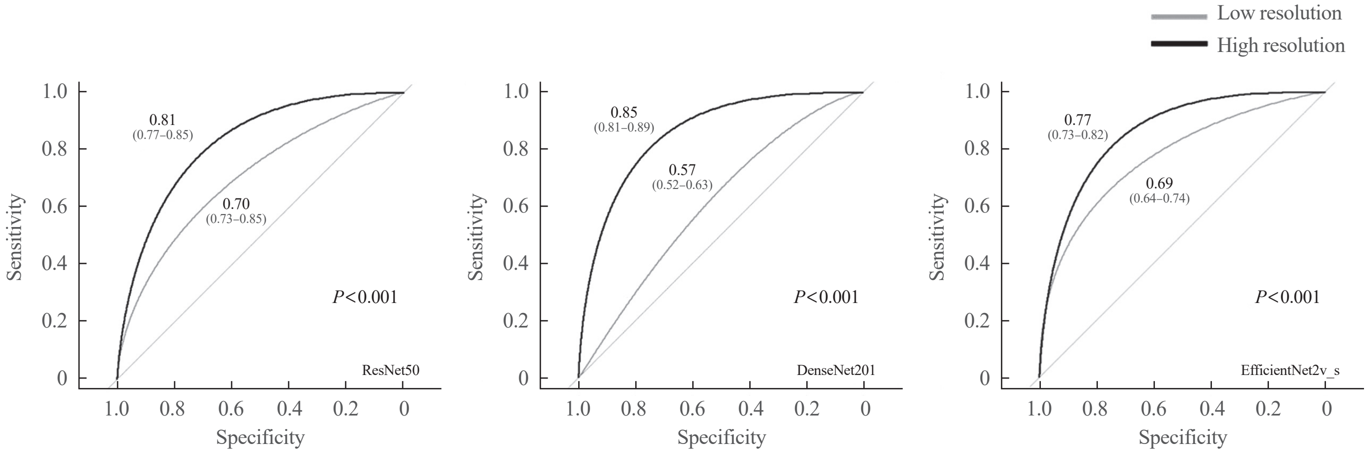

Fig. 3. Comparison of model performance according to image resolution.

an accuracy greater than 90% in diagnosing thyroid nodules. In Korea, radiologists have attempted to analyze thyroid nodule ultrasound images using deep learning. Alex-Net [15] and VGG-Net [16] were compared with ResNet-based algorithms and classification accuracies greater than 80% were reported, indicating their clinical applicability. In a recent large-scale study, AI-Thyroid developed using 17 different CNN algorithms was validated at 25 multi-centers. This well-trained model yielded an AUC of approximately 90% regardless of the malignancy rate of each institution [17].

CNN is basically composed of an architecture of convolution layers. Although performance is expected to improve as the neural network becomes larger and deeper, the complex structure increases the number of calculations, complicates parameter settings, and can cause problems associated with overfitting when reporting the results. Therefore, new network structures have been developed to improve performance while addressing these challenges. In this study, three recently developed models—ResNet, DenseNet, and EfficientNet—were applied to thyroid ultrasound images. The concept of residual connections to ease the training of deeper networks was introduced in ResNet [18]. DenseNet connects each layer using a channel-wise concatenation method rather than sequentially connecting layers to improve information flow and gradient propagation [19]. EfficientNet uses a compound scaling method to optimize the balance among network depth, width, and resolution. This can significantly increase model performance while simplifying the parameter setting method [20]. Consistent with the findings of a previous study [17], all three algorithms showed excellent performance in this study cohort (Table 2).

Since thyroid images are generally taken in both transverse and longitudinal views, we performed deep learning analysis on images from each view and compared the accuracy. In our results, longitudinal images showed superior accuracy (Fig. 2). We assume that this is related to the way these images are preprocessed. The disadvantage of the box-shaped labeling method is difficult to focus only on the nodule, especially in transverse images, which were included the numerous structures present around the thyroid gland. Therefore, we additionally used Siamese networks for pairing images from transverse and longitudinal views [21]. This multi-view analysis mirrors the multifaceted approach of radiologists when assessing images, aligning the automated process closely with clinical practice. Previous researchers have demonstrated that multi-view approaches can enhance model performance by leveraging complementary information from different perspectives [22]. However, depending on the characteristics of the image and the anatomical region of the target diseases, the optimal number of views to improve classification accuracy may vary, and further investigation is needed on thyroid images [23]. Computer-aided diagnosis systems (CADS) are currently used for thyroid nodules in a semi-automated manner [24], in other words, physicians performing ultrasound examination play an important role in acquisition of sonogram images. The multi-view analysis introduced in this study could resolve inter-observer variability by analyzing two or more images.

The overall accuracy of this study reported lower than that in previous studies due to its low sensitivity. This low sensitivity is likely due to the low resolution of images in the initial dataset. Therefore, subgroup analysis was performed on images taken before and after April 2011, as this is when our clinic started using a new ultrasound device with improved image resolution. The results confirmed that the better resolution of raw images improved the accuracy by increasing the sensitivity of diagnosis (Fig. 3). Based on these findings, we suggest that that the prediction performance of AI models tended to increase as the quality and quantity of the training images increased.

Various types of AI algorithms have been developed to analyze ultrasound images of thyroid nodules; however, application of the developed algorithms in the real world is a major challenge [25]. AI systems provided additional information for classification of thyroid nodules through a local computer or web-based service. S-detect, the most well-known CADS that has been validated in numerous studies, segments nodules and reads sonographic features with considerable accuracy [26]. This system is uniquely implemented in the ultrasound machine and has been commercialized. AI-Thyroid, a deep learning model developed by Korean radiologists, can report the malignancy rate of thyroid nodules and is publicly available on a web-based service [27]. Because the image characteristics of the device that takes the images and the accuracy of the algorithm are closely related, it may be more realistic to match appropriate algorithms to specific machines. When using a web-based algorithm, it is necessary to interpret the result carefully by considering the type of machine and method used to take image.

Multi-view algorithms that learned approximately 1,000 nodules had classification performance comparable to that of the K-TIRADS classification, which is based on the interpretation of an endocrinologist with 5 years of experience. Additionally, deep learning models in our study had superior specificity than the traditional method, which is an advantage in clinical application. Because thyroid cancer is generally indolent [28], high specificity is essential for clinical application to avoid over-diagnosis and treatment. In particular, indeterminate nodules, which are thought to benefit from additional diagnostic testing because the diagnosis is unclear even after biopsy, are of clinical unmet need [29]. Evaluation of the presence of BRAF mutations, which is recommended for additional diagnosis of indeterminate nodules [30], has a sensitivity of only 40% but is recognized for its clinical usefulness because its specificity reaches 100% [31]. Therefore, the application of deep learning models to indeterminate nodules in clinical practice is worthy of future research [32].

With deep learning algorithms, probability-based scoring can be augmented by additionally reviewing the soft-max function results in the last step of the cascade of digitization for computer analysis of images. Images quantified through a convolutional matrix are converted into regression form to predict the class level. Currently, the regression value for each class is normalized through a soft-max function and is output as a probability value between 0 and 1. Typical output is given as the class for which this probability value is highest [33]. Because cancer risk is stratified into five levels for thyroid imaging under the existing guidelines [34-36], continuous cancer probability values can provide additional information to help with thyroid nodule management [37].

Several limitations of this study also need to be discussed. First, the dataset size for model training was relatively small. Compared with a previous study in which the accuracy of the AI improved by learning more than 10,000 nodules [13,17], our model learned from only approximately 1,000 nodules. In addition, the dataset we used was from a single center, and external validation was not performed. The selection bias that commonly occurs in retrospective studies is called dataset shift in the field of AI [38], and external validation is essential for generalization of the model and clinical application because the accuracy of the algorithm is greatly dependent on the quality of training data [39]. Second, the dataset comprised patients who visited the hospital more than 10 years previously and may not reflect recent patient trends. However, the definitive diagnosis of thyroid nodules even in patients who did not receive surgical treatment was estimated using a sufficient follow-up period. In addition, data were collected before and after replacement of devices and showed clear differences in resolution for comparison between diagnostic devices. Last, because we only included binary classes of classic PTC or benign, this AI model cannot differentiate rare cancers of the thyroid gland (e.g., follicular neoplasm, medullary thyroid cancer, anaplastic thyroid cancer, and lymphoma). Therefore, clinicians should take rare pathologic categories into consideration when using current binary prediction type AI based algorithms.

Application of up-to-date deep learning technology to thyroid ultrasound images is expected to be helpful in differentiating of thyroid nodules. However, in real-world clinical settings, it is important to understand that model performance can greatly depend on the quality of the images acquired by different physicians and devices.

CONFLICTS OF INTEREST

Mee Kyoung Kim is a deputy editor of the journal. But she was not involved in the peer reviewer selection, evaluation, or decision process of this article. No other potential conflicts of interest relevant to this article were reported.

ACKNOWLEDGMENTS

This research was supported by the Basic Science Research Program through a grant from the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2023-00245534).

AUTHOR CONTRIBUTIONS

Conception or design: J.K., K.H.B. Acquisition, analysis, or interpretation of data: J.K., H.L., J.J.L., Y.O.L. Drafting the work or revising: J.K., M.H.K., D.J.L., H.S.K., M.K.K., K.H.S., T.J.K., S.L.J., Y.O.L., K.H.B. Final approval of the manuscript: J.K., M.H.K., D.J.L., H.L., J.J.L., H.S.K., M.K.K., K.H.S., T. J.K., S.L.J., Y.O.L., K.H.B.

ORCID

Jinyoung Kim *https://orcid.org/0000-0003-4581-9493*
Yong Oh Lee *https://orcid.org/0000-0003-3817-3620*
Ki-Hyun Baek *https://orcid.org/0000-0002-4097-1077*

REFERENCES

1. Choi YM, Lee J, Kwak MK, Jeon MJ, Kim TY, Hong EG, et al. Recent changes in the incidence of thyroid cancer in Korea between 2005 and 2018: analysis of Korean national data. Endocrinol Metab (Seoul) 2022;37:791-9.
2. Wiltshire JJ, Drake TM, Uttley L, Balasubramanian SP. Systematic review of trends in the incidence rates of thyroid cancer. Thyroid 2016;26:1541-52.
3. Frates MC, Benson CB, Doubilet PM, Kunreuther E, Contreras M, Cibas ES, et al. Prevalence and distribution of carcinoma in patients with solitary and multiple thyroid nodules on sonography. J Clin Endocrinol Metab 2006;91:3411-7.
4. Durante C, Grani G, Lamartina L, Filetti S, Mandel SJ, Cooper DS. The diagnosis and management of thyroid nodules: a review. JAMA 2018;319:914-24.
5. Koc AM, Adıbelli ZH, Erkul Z, Sahin Y, Dilek I. Comparison of diagnostic accuracy of ACR-TIRADS, American Thyroid Association (ATA), and EU-TIRADS guidelines in detecting thyroid malignancy. Eur J Radiol 2020;133:109390.
6. Chung SR, Ahn HS, Choi YJ, Lee JY, Yoo RE, Lee YJ, et al. Diagnostic performance of the modified Korean thyroid imaging reporting and data system for thyroid malignancy: a multicenter validation study. Korean J Radiol 2021;22: 1579-86.
7. Kim DH, Kim SW, Basurrah MA, Lee J, Hwang SH. Diagnostic performance of six ultrasound risk stratification systems for thyroid nodules: a systematic review and network meta-analysis. AJR Am J Roentgenol 2023;220:791-803.
8. Rajpurkar P, Lungren MP. The current and future state of ai interpretation of medical images. N Engl J Med 2023;388: 1981-90.
9. Assie G, Allassonniere S. Artificial intelligence in endocrinology: on track toward great opportunities. J Clin Endocrinol Metab 2024;109:e1462-7.
10. Egger J, Gsaxner C, Pepe A, Pomykala KL, Jonske F, Kurz M, et al. Medical deep learning: a systematic meta-review. Comput Methods Programs Biomed 2022;221:106874.
11. Rawat W, Wang Z. Deep convolutional neural networks for image classification: a comprehensive review. Neural Com-

put 2017;29:2352-449.

12. Chlap P, Min H, Vandenberg N, Dowling J, Holloway L, Haworth A. A review of medical image data augmentation techniques for deep learning applications. J Med Imaging Radiat Oncol 2021;65:545-63.
13. Peng S, Liu Y, Lv W, Liu L, Zhou Q, Yang H, et al. Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study. Lancet Digit Health 2021;3:e250-9.
14. Guo GHY. ThyNet [Internet]. San Francisco: GitHub; 2021 [cited 2024 Oct 15]. Available from: https://github.com/sprint2200/ThyNet.
15. Park VY, Han K, Seong YK, Park MH, Kim EK, Moon HJ, et al. Diagnosis of thyroid nodules: performance of a deep learning convolutional neural network model vs. radiologists. Sci Rep 2019;9:17843.
16. Kim YJ, Choi Y, Hur SJ, Park KS, Kim HJ, Seo M, et al. Deep convolutional neural network for classification of thyroid nodules on ultrasound: comparison of the diagnostic performance with that of radiologists. Eur J Radiol 2022;152:110335.
17. Ha EJ, Lee JH, Lee DH, Moon J, Lee H, Kim YN, et al. Artificial intelligence model assisting thyroid nodule diagnosis and management: a multicenter diagnostic study. J Clin Endocrinol Metab 2024;109:527-35.
18. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 2016;2016:770-8.
19. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 2017;2017:2261-9.
20. Tan M, Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks. Proc Mach Learn Res 2019;97:6105-14.
21. Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition. Proceedings of the 32nd International Conference on Machine Learning; 2015 Jul 6-11; Lille, France. Available from: https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf
22. Liu K, Kang G. Multiview convolutional neural networks for lung nodule classification. Int J Imaging Syst Technol 2017;27:12-22.
23. Wang M, Ma Z, Wang Y, Liu J, Guo J. A multi-view convolutional neural network method combining attention mechanism for diagnosing autism spectrum disorder. PLoS One 2023;18:e0295621.
24. Tessler FN, Thomas J. Artificial intelligence for evaluation of thyroid nodules: a primer. Thyroid 2023;33:150-8.
25. Sant VR, Radhachandran A, Ivezic V, Lee DT, Livhits MJ, Wu JX, et al. From bench-to-bedside: how artificial intelligence is changing thyroid nodule diagnostics, a systematic review. J Clin Endocrinol Metab 2024;109:1684-93.
26. Choi YJ, Baek JH, Park HS, Shim WH, Kim TY, Shong YK, et al. A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of thyroid nodules on ultrasound: initial clinical assessment. Thyroid 2017;27:546-52.
27. Kim J. Cancer Classification from Ultrasound Image: Deep learning algorithm to detect thyroid cancer [Internet]. Clinical Decision Support System; 2023 [cited 2024 Oct 15]. Available from: http://us.cdss.co.kr.
28. Jensen CB, Saucke MC, Francis DO, Voils CI, Pitt SC. From overdiagnosis to overtreatment of low-risk thyroid cancer: a thematic analysis of attitudes and beliefs of endocrinologists, surgeons, and patients. Thyroid 2020;30:696-703.
29. van Kinschot CM, Soekhai VR, de Bekker-Grob EW, Visser WE, Peeters RP, van Noord C, et al. Preferences of patients, clinicians, and healthy controls for the management of a Bethesda III thyroid nodule. Head Neck 2023;45:1772-81.
30. Park YJ, Lee EK, Song YS, Kang SH, Koo BS, Kim SW, et al. 2023 Korean Thyroid Association management guidelines for patients with thyroid nodules. Int J Thyroidol 2023;16:1-31.
31. Jinih M, Foley N, Osho O, Houlihan L, Toor AA, Khan JZ, et al. BRAFV600E mutation as a predictor of thyroid malignancy in indeterminate nodules: a systematic review and meta-analysis. Eur J Surg Oncol 2017;43:1219-27.
32. Gild ML, Chan M, Gajera J, Lurie B, Gandomkar Z, Clifton-Bligh RJ. Risk stratification of indeterminate thyroid nodules using ultrasound and machine learning algorithms. Clin Endocrinol (Oxf) 2022;96:646-52.
33. Dong Q, Zhu X, Gong S. Single-label multi-class image classification by deep logistic regression. Proc AAAI Conf Artif Intell 2019;33:3486-93.
34. Ha EJ, Chung SR, Na DG, Ahn HS, Chung J, Lee JY, et al. 2021 Korean thyroid imaging reporting and data system and imaging-based management of thyroid nodules: Korean Society of Thyroid Radiology consensus statement and recommendations. Korean J Radiol 2021;22:2094-123.
35. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid Association guidelines for

ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. Eur Thyroid J 2017;6:225-37.

36. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. J Am Coll Radiol 2017;14:587-95.

37. Liang X, Huang Y, Cai Y, Liao J, Chen Z. A computer-aided diagnosis system and thyroid imaging reporting and data system for dual validation of ultrasound-guided fine-needle aspiration of indeterminate thyroid nodules. Front Oncol 2021; 11:611436.

38. Moreno-Torres JG, Raeder T, Alaiz-Rodríguez R, Chawla NV, Herrera F. A unifying view on dataset shift in classification. Pattern Recognit 2012;45:521-30.

39. Yu AC, Mohajer B, Eng J. External validation of deep learning algorithms for radiologic diagnosis: a systematic review. Radiol Artif Intell 2022;4:e210064.