Stratifying High-Risk Thyroid Nodules Using a Novel Deep Learning System

Authors
Chia-Po Fu[1, 2, 3, 4], Ming-Jen Yu[1], Yao-Sian Huang[5], Chiou-Shann Fuh[1, 6], Ruey-Feng Chang[1, 6]

Affiliations
1 Graduate Institute of Biomedical Electronics and Bioinformatics, College of Electrical Engineering and Computer Science, National Taiwan University, Taipei, Taiwan
2 Division of Endocrinology and Metabolism, Department of Medicine, Taichung Veterans General Hospital, Taichung, Taiwan
3 Department of Medicine, Chung Shan Medical University, Taichung, Taiwan
4 Department of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan
5 Department of Computer Science and Information Engineering, National Changhua University of Education, Changhua County, Taiwan
6 Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan

Key words
thyroid nodule, sonography, deep learning, ResNeSt50, Swin-T

received 26.03.2023
revised 15.06.2023
accepted 28.06.2023

Bibliography
Exp Clin Endocrinol Diabetes 2023; 131: 1–7
DOI 10.1055/a-2122-5585
ISSN 0947-7349

Georg Thieme Verlag, Rüdigerstraße 14,
70469 Stuttgart, Germany

Correspondence
Ruey-Feng Chang
Graduate Institute of Biomedical Electronics and Bioinformatics, College of Electrical Engineering and Computer Science
National Taiwan University
Taipei
Taiwan
Tel: + 886-2-33664888, Fax: + 886-2-23628167
rfchang@csie.ntu.edu.tw

ABSTRACT

Introduction The current ultrasound scan classification system for thyroid nodules is time-consuming, labor-intensive, and subjective. Artificial intelligence (AI) has been shown to increase the accuracy of predicting the malignancy rate of thyroid nodules. This study aims to demonstrate the state-of-the-art Swin Transformer to classify thyroid nodules.

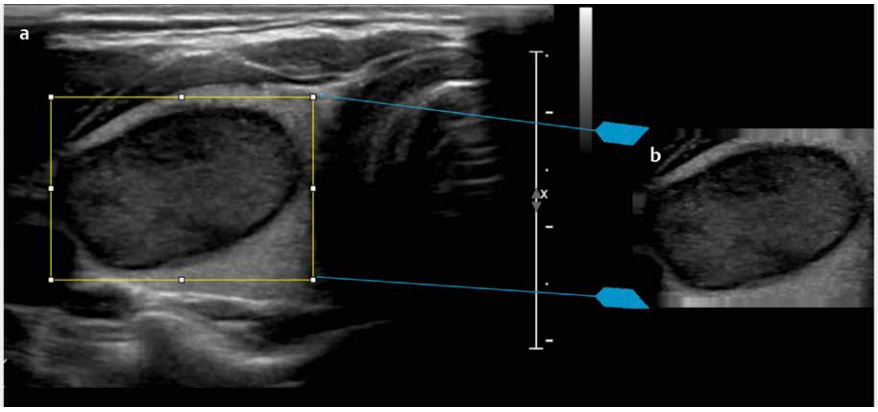
Materials and Methods Ultrasound images were collected prospectively from patients who received fine needle aspiration biopsy for thyroid nodules from January 2016 to June 2021. One hundred thirty-nine patients with malignant thyroid nodules were enrolled, while 235 patients with benign nodules served as controls. Images were fed to Swin-T and ResNeSt50 models to classify the thyroid nodules.

Results Patients with malignant nodules were younger and more likely male compared to those with benign nodules. The average sensitivity and specificity of Swin-T were 82.46 % and 84.29 %, respectively. The average sensitivity and specificity of ResNeSt50 were 72.51 % and 77.14 %, respectively. Receiver operating characteristics analysis revealed that the area under the curve of Swin-T was higher (AUC = 0.91) than that of ResNeSt50 (AUC = 0.82). The McNemar test evaluating the performance of these models showed that Swin-T had significantly better performance than ResNeSt50.

Swin-T classifier can be a useful tool in helping shared decision-making between physicians and patients with thyroid nodules, particularly in those with high-risk characteristics of sonographic patterns.

Introduction

Thyroid nodules are distinct mass lesions within the thyroid gland; some are palpable on physical examination, while some are not. With the widespread use of diagnostic imaging, improved access to health care, and an increase in disease insight of the affected individuals, more thyroid “incidentalomas” are being discovered. The



▶ **Fig. 1** ■■■

prevalence of thyroid incidentaloma detected through ultrasound examination ranges from 19 % to 67 % [1]. Furthermore, the overall incidence of thyroid cancer is rising [2]. Additionally, non-palpable nodules or incidentalomas carry the same risk of malignancy as palpable nodules larger than 1 cm [3]. According to one systematic review, ultrasound examination is not accurate enough to predict thyroid cancer [4]. Fine-needle aspiration biopsy (FNAB) remains the best method to differentiate the nodules between benign and malignant, with high sensitivity and specificity of 87 % and 100 %, respectively [5]. Nevertheless, among all clinically identified thyroid nodules, only 5 % are malignant [6], and individuals require FNAB semiannually or annually when ultrasound findings are ambiguous.

The Thyroid Imaging Reporting and Data System (TIRADS), developed by the American Thyroid Association and the American College of Radiology, is the most commonly used system to classify thyroid nodules [7, 8]. Nevertheless, the TIRADS classification can be labor-intensive as the performer needs to obtain and enter all the necessary profile characteristics before the classification. The interpretation can also be subjective and vary across different observers [9]. The prediction rates for malignancy for the highest TR score (TR5) is at least 20 % using TIRADs [10] and 26–87 % using EU-TIRADS [11]. However, both classification systems are limited by high variability. Therefore, this study aimed to increase the sensitivity and specificity of ultrasound scans in the evaluation of nodular goiters using artificial intelligence (AI) and machine learning tools to help physicians with additional support in making decisions for managing patients with thyroid. At the same time, the utilization of these tools can help patients avoid the unnecessary annual FNAB for thyroid nodules, especially nodules displaying high-risk characteristics on ultrasound scans.

Materials and Methods

Subjects

Consecutive patients aged 18 years or older who visited the Endocrinology Clinic for treatment of nodular goiters were recruited for this retrospective study. Blood samples from patients were drawn three months before or after ultrasound scans. Ultrasound images of thyroid nodules were collected from patients who underwent FNAB with or without receiving a blood test. All procedures were performed in accordance with the guidelines and regulations of the American Thyroid Association. Patients were excluded from the study if they had any abnormal thyroid function test results or had a prior diagnosis of Hashimoto's thyroiditis or Graves' disease. Real-time ultrasonography-guided FNAB was performed by some experienced endocrinologists. If patients had more than one nodule, the nodule with high-risk characteristics on ultrasound scans, such as hypoechoic appearance, irregular margins, microcalcifications, a taller-than-wide shape, or extrathyroidal extension, etc., was selected for FNAB regardless of nodule size. All FNAB specimens were processed using the ThinPrep method (Hologic Inc., Waltham, MA, USA) with Papanicolaou stain [12]. Cytological examination was performed and the results were reported using the Bethesda system [13] by a group of pathologists. FNAB results indicating suspicion of malignancy or definitive diagnosis of malignancy were defined as malignancy according to the Bethesda system, while nodules with benign diagnosis were defined as benign in the present study.

Ethical considerations

The study protocol was approved by the Institutional Review Board.

Image database

Thyroid ultrasound images were obtained using frequencies from 5 MHz to 12 MHz, obtained from an ultrasound machine (Affiniti 30, Philips, Netherlands). The images were downloaded in Digital Imaging and Communications in Medicine (DICOM) format. Image of a transverse B-mode thyroid nodule with maximal diameters was taken, and then the image of the nodule was cropped by an endocrinologist (▶ **Fig. 1**). Thyroid ultrasound images were excluded if the definitive diagnosis of FNAB was not benign or suspicious for malignancy or malignancy, as defined by the Bethesda system.

Swin-Transformer and ResNeSt50 model

We used the Swin-T (Tiny) version of the Swin Transformer [14] as our model, which is about a quarter of the computation complexity size of Swin-B (Base). Then, we leveraged the features learned by the Swin-T model pretrained on ImageNet-1K and fine-tuned it

on our dataset. All input images were resized to 224 × 224 pixels to align with the original image size of ImageNet-1K. The maximum epoch we trained for a model was set to 50 epochs, and the best model for each fold was selected based on the highest validation accuracy score achieved. The learning rate of the model was set to 5×10^{-6}, applied with cosine learning rate decay and a warm-up for 100 steps. The AdamW optimizer employed in the earlier study on Swin Transformer was used; the weight decay was set to 5×10^{-4}, and the Binary Cross Entropy loss function was used. Random horizontal flipping was the only data augmentation applied, with a 50 % probability of horizontal flipping the input images. Since the original Swin-T model was used to predict 1000 classes of images in the ImageNet-1K dataset, the last linear layer was modified so that it could predict cases of two classes, either benign or malignant; thus, the total number of parameters was slightly smaller than original Swin-T. To maintain the similarity in the number of parameters, we used ResNeSt50 and modified its last linear layer, like the above-mentioned explanation. Consequently, the modified ResNeSt50 resulted in a smaller number of parameters than the original ResNeSt50 model. The training parameters for ResNeSt50 were nearly identical to those used for the Swin-T model, except for the learning rate. The learning rate of the ResNeSt50 was set at a fixed value 1×10^{-4} to reach high performance because a small learning rate and cosine warm-up do not yield satisfactory results.

Statistical analysis

Continuous variables are presented as mean ± SD. The student's t-test was used for comparison between groups. Both Swin-T and ResNeSt50 models were programmed using Python language. The receiver operating characteristics (ROC) curve was performed to illustrate the area under the curve (AUC) between Swin-T and ResNeSt50. The McNemar test was used to compare the performance of Swin-T and ResNeSt50. All data were considered statistically significant at $P < 0.05$ using Statistical Package for the Social Sciences, version 12.0 (SPSS, Chicago, IL).

Results

Baseline characteristics, nodule size, and serum biochemistries of patients with benign or malignant nodules

A total of 16,380 patients received thyroid ultrasound examinations and FNAB at our institution from January 2016 to June 2021. A total of 139 malignant nodules were detected and the rate of malignancy was 0.84 %. In addition, 235 patients with the FNAB with benign reports were randomly collected using the function of Rand in Microsoft Excel. Finally, 374 patients who met the criteria of undergoing thyroid ultrasound examinations and FNAB were enrolled in this study. Compared with patients with benign nodules, those with malignant nodules were younger and more likely to be male ($P < 0.001$; ▶ **Table 1**). The size of the thyroid gland (including the right lobe and left lobe) in patients with malignant nodules was significantly smaller than that in patients with benign nodules. The size of the isthmus was insignificantly different. The mean nodule size (nodule width 1.15 ± 0.76 cm and nodule height 0.98 ± 0.67 cm) was smaller in patients with malignant nodules than those with benign ones (nodule width 1.83 ± 0.85 cm, nodule height 1.15 ± 0.76 cm, both $P < 0.001$; ▶ **Table 1**). There was no significant difference between patients with benign or malignant nodules in white blood cell count, hemoglobin platelet, neutrophil, lymphocyte, plasma sugar, creatinine, alanine aminotransferase, thyroid stimulating hormone, or free thyroxin levels (▶ **Table 1**).

Image database analysis

The ultrasonography images were distributed equally into five folds during the validation, and some images may have been duplicated and used. The performance of the Swin-T model was superior to that of ResNeSt50, except that Swin-T had lower sensitivity in fold 4 (▶ **Table 2**). The average sensitivity and specificity of Swin-T were 82.46 and 84.29 %, respectively. The average sensitivity and specificity of ResNeSt50 were 72.51 and 77.14 %, respectively (▶ **Table 2**). In the ROC analysis, the AUC of Swin-T was higher (AUC = 0.91) than ResNeSt50 (AUC = 0.82) (▶ **Fig. 2**). In the McNemar test, used to compare the superiority between Swin-T and ResNeSt50, the chi-square statistics was 12.99 ($P < 0.001$, ▶ **Table 3**). Heatmap of ResNest50 and Swin-T showed a more accurate visualization of the Swin-T model (▶ **Fig. 3a, b, c**).

Discussion

The primary finding of this study was that the Swin-T of Swin Transformer is a good AI tool for predicting the malignancy of thyroid nodules with high-risk characteristics. An experienced endocrinologist can evaluate the characteristics of thyroid ultrasonography or utilize TIRADS to justify the possibility of malignancy before making a decision for further FNAB. Nevertheless, interobserver and intraobserver variations still remain [9]. This study presents solutions for consistency using a machine learning model to increase the confidence of the physician, reduce the number of FNABs performed on benign nodules, and increase the follow-up rates of the patients. The malignancy rate of FNAB was only 0.84 % in the present study, suggesting that unnecessary biopsy is a big unsolved problem in clinical practice.

FNAB should be performed in thyroid nodules with a size greater than 1 cm and highly suspicious sonographic patterns, e. g., hypoechoic, irregular margins, microcalcifications, taller than wide nodules, or extrathyroidal extension [7]. Nevertheless, in Taiwan, the performance of FNAB for detecting thyroid cancer is low for thyroid nodules larger than 3 cm [15]. Also, thyroid nodules of less than 1.2 cm in size carry a higher risk of malignancy than those greater than 1.2 cm [16]. Thus, we suggest that FNAB should be performed in nodules of any size with a highly suspicious sonographic pattern of malignancy.

In addition, the level of thyroid-stimulating hormone (TSH) did not predict the malignancy of the thyroid nodules in the present study. A recent case-control study conducted in Germany demonstrated an elevated risk of thyroid cancer in a J curve of TSH levels [17]. Two previous studies also reported the association of higher serum TSH levels with thyroid cancer, while some patients had autoimmune thyroid disease [18, 19]. A possible explanation is that autoimmune thyroid diseases, i. e., Graves' disease and Hashimoto's thyroiditis, are associated with thyroid cancer with either elevated or suppressed TSH levels [20–22]. In this study, autoimmune thyroid disease was excluded, and TSH levels were not associated

▶ **Table 1** Baseline characteristics of patients with benign or malignant thyroid nodules.

	Patients with benign nodules (N = 235)	Patients with malignant nodules (N = 139)	P-value
Age, y/o	55.6 ± 13.0	49.2 ± 13.1	< 0.001
Sex, M (%)	32 (13.6)	39 (28.1)	< 0.001
Isthmus, cm	0.45 ± 0.36	0.41 ± 0.82	0.46
Right lobe width, cm	2.42 ± 0.66	1.89 ± 0.56	< 0.001
Right lobe height, cm	2.04 ± 0.57	1.60 ± 0.41	< 0.001
Left lobe width, cm	2.30 ± 0.70	1.88 ± 0.50	< 0.001
Left lobe height, cm	1.86 ± 0.60	1.50 ± 0.46	< 0.001
Nodule width, cm	1.83 ± 0.85	1.15 ± 0.76	< 0.001
Nodule height, cm	1.42 ± 0.64	0.98 ± 0.67	< 0.001
White blood cell count,/μL	6878 ± 2420 (N = 74)	6601 ± 1943 (N = 110)	0.39
Hemoglobin, g/dL	13.3 ± 2.0 (N = 78)	13.8 ± 1.6 (N = 112)	0.09
Platelet, x1000/μL	277 ± 91(N = 71)	267 ± 71 (N = 112)	0.38
Neutrophil, %	61 ± 10 (N = 53)	61 ± 9 (N = 69)	0.99
Lymphocyte, %	29 ± 10 (N = 56)	31 ± 8 (N = 69)	0.42
Plasma sugar, mg/dL	104 ± 24 (N = 58)	106 ± 28 (N = 43)	0.72
Creatinine, mg/dL	0.89 ± 1.00 (N = 105)	0.85 ± 0.74 (N = 119)	0.75
ALT, mg/dL	23 ± 19 (N = 106)	25 ± 19 (N = 114)	0.49
TSH, μIU/mL	1.68 ± 4.30 (N = 144)	2.11 ± 4.01 (N = 101)	0.44
Free T4, ng/dL	1.06 ± 0.15 N = 123)	1.05 ± 0.13 (N = 103)	0.56

ALT: alanine aminotransferase; TSH: thyroid stimulating hormone.

▶ **Table 2** Comparison of Swin-T and ResNeSt50 classifiers.

	Swin-T			ResNeSt50		
Fold (Benign/Malignancy)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
Folder 1 (41/35)	82.86	82.93	82.89	80.00	65.85	72.37
Folder 2 (43/34)	82.35	88.37	85.71	73.53	74.42	74.03
Folder 3 (42/35)	80.00	80.95	80.52	68.57	80.95	75.32
Folder 4 (42/34)	79.41	88.10	84.21	85.29	80.95	82.89
Folder 5 (42/33)	87.88	80.95	84.00	54.55	83.33	70.67
Average (210/171)	82.46	84.29	83.46	72.51	77.14	75.07

▶ **Table 3** McNemar test of Swin-T and ResNeSt50 models.

		ResNeSt50	
		correct	wrong
Swin-T	correct	265	53
	wrong	21	42

Chi-square statistics = 12.99; *P*-value < 0.001.

with thyroid cancer, consistent with the results of a previous meta-analysis [23].

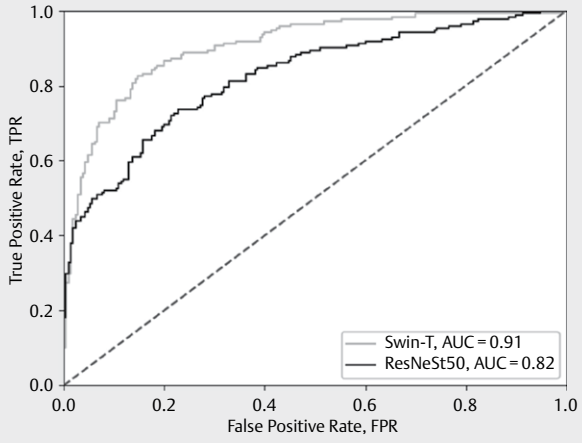
TIRADS is a labor-intensive and sometimes subjective classification method [9]. Machine learning ultrasonography image analysis is time-saving and can help to increase the confidence of physicians in sharing decision-making with patients. Real-time ultrasonography-guided FNAB can minimize the possibility of missed nodules. Machine-learning computer-aided diagnostic systems performed once can lower the necessity of repeat FNABs, and reduce the burden of mental/emotional pressure and anxiety of patients. While many AI models have been developed, here, we chose Swin Transformer [14] and ResNest [24] because both are state-of-the-art classifiers. Over the past few years, Swin Transformer has been used extensively in natural language processing, likely because of its ability to extract sequential information and establish long-range dependency [25].

Recently, Google has introduced a new transformer method, Vision Transformer (ViT) [26], that has been applied to computer vision tasks, initiating the era of transformer applications on computer vision. Though Vision Transformer seems to perform well on huge datasets such as ImageNet, it usually underperforms compared to other models when trained on tiny datasets. One explanation is that ViT lacks locality, such as convolutional layers. Instead of using ViT, we selected Swin Transformer as our model structure. A key element in Swin Transformer, called Window-based Multi-head Attention, computes attention in each window that contains several image patches and integrates the strengths of convoluted

neural network (CNN) and Transformer. It achieved computational complexity linear to image size, and introduced locality to Transformer.

The image AI similarity model developed by Thomas et al. has an accuracy of 81.5 %, which was better than that of ResNet34 (77.7 %) [27]; in the present study, the Swin-T model had an accuracy of 83 %, also better than that of ResNest50 (75 %). All of their testing nodules were at least one dimension greater than 1 cm, while the Swin-T model collected nodules of about 1 cm in size. Furthermore, our Swin-T model had superior performance than the previous deep CNN (AUC = 0.91 vs. 0.835–0.85) [28]. This model can be applied to clinical practice because there is no restriction on nodule size selection.

The goal of the present study was not to replace the physicians' role or FNAB. Instead, with the emergence of shared decision-making, the role of physician-patient communication was emphasized [29]. AI is just a tool to help the physician communicate with patients. A previous study demonstrated that decreased decisional conflict and fewer nodule FNAB are possible with web-based thyroid nodule conversation aid [30]. FNAB should also account for anxiety, life expectancy, comorbidities of the patient, and other relevant considerations. Essentially, the image of the nodule was selected into the Swin-T and interpreted as benign or malignant with 83 % accuracy. Patients have three choices after a shared decision-making discussion, i. e., nodule monitor, nodule FNAB, or surgical intervention (▶ **Fig. 4**). To the best of our knowledge, this is the first study to use the Swin-T classifier for the classification of thyroid nodules. We also agree with Thomas et al. that physicians have an active role in the AI model. Besides, the better the accuracy, the better physicians can integrate all available information, i. e., age, gender, molecular markers, and history of received radiotherapy or lymphadenopathy. Results of the present study suggest that accuracy will continue to increase with the promising development of AI.



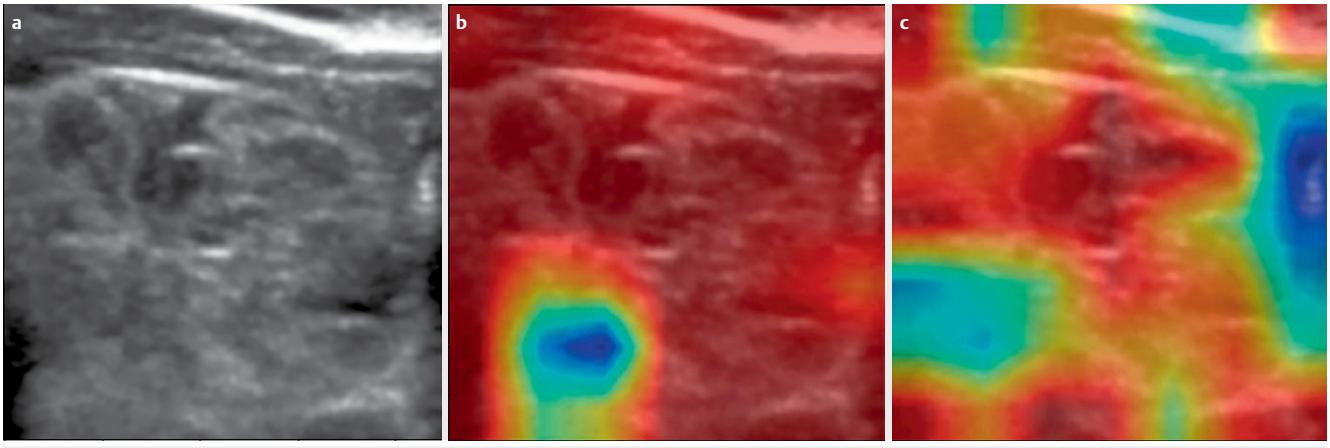
▶ **Fig. 2** ■■■

This study has several limitations. First, only Asian subjects were included and results may not be applicable to non-Asian populations. Second, the image database was relatively small and could benefit from more image input for analysis. Third, the benign or malignancy reports were based on cytology reports instead of pathology reports; thus, the definitive diagnosis may not be correct. However, this only had a minor impact on our results because we chose diagnostic categories II, V, and VI of the Bethesda system. Finally, measurements of TSH levels were not all checked on the same day as ultrasound scans, and this may interfere with results because serum TSH results typically have significant daily and diurnal variations.

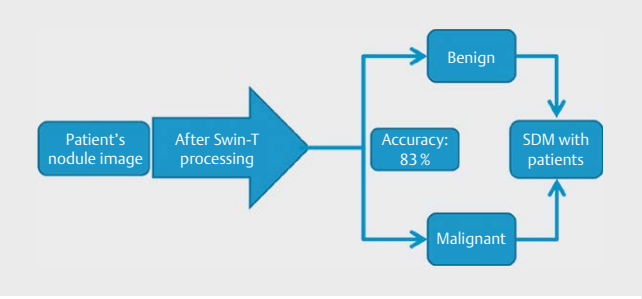
Swin-T classifier can be a useful tool in helping shared decision-making between physicians and patients with thyroid nodules of any size, particularly in those with high-risk characteristics of the sonographic pattern. The model consistently demonstrated better performance when compared with the current state-of-the-art image similarity algorithm.

Authors' Contributions

Chia-Po Fu and Ruey-Feng Chang performed the study design, image collection, statistical analysis, and draft writing. Min-Jen Yu, Yao-Sian Huang, Chiou-Shann Fuh, and Ruey-Feng Chang performed the coding, and image analysis. All authors reviewed and agreed on the final version of the article.



▶ **Fig. 3** ■■■



▶ **Fig. 4** ■■■

Funding

This work was supported by grants from the Ministry of Science and Technology of Taiwan (MOST 110–2221-E-002–122-MY3 and MOST 110–2221-E-002–123-MY3).

Conflict of Interest

The authors declare that they have no conflict of interest.

One sentence summary Stratifying High Risk Thyroid Nodules Using AI

References

[1] Welker MJ, Orlov D. Thyroid nodules. Am Fam Physician 2003; 67: 559–566

[2] Wiltshire JJ, Drake TM, Uttley L et al. Systematic review of trends in the incidence rates of thyroid cancer. Thyroid 2016; 26: 1541–1552. DOI: 10.1089/thy.2016.0100

[3] Hagag P, Strauss S, Weiss M. Role of ultrasound-guided fine-needle aspiration biopsy in evaluation of nonpalpable thyroid nodules. Thyroid 1998; 8: 989–995. DOI: 10.1089/thy.1998.8.989

[4] Brito JP, Gionfriddo MR, Al Nofal A et al. The accuracy of thyroid nodule ultrasound to predict thyroid cancer: Systematic review and meta-analysis. J Clin Endocrinol Metab 2014; 99: 1253–1263. DOI: 10.1210/jc.2013-2928

[5] Wu HH, Jones JN, Osman J. Fine-needle aspiration cytology of the thyroid: Ten years experience in a community teaching hospital. Diagn Cytopathol 2006; 34: 93–96. DOI: 10.1002/dc.20389

[6] Hegedus L. Clinical practice. The thyroid nodule. N Engl J Med 2004; 351: 1764–1771. DOI: 10.1056/NEJMcp031436

[7] 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: 1–133. DOI: 10.1089/thy.2015.0020

[8] Horvath E, Majlis S, Rossi R et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. J Clin Endocrinol Metab 2009; 94: 1748–1751. DOI: 10.1210/jc.2008-1724

[9] Choi SH, Kim EK, Kwak JY et al. Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules. Thyroid 2010; 20: 167–172. DOI: 10.1089/thy.2008.0354

[10] Tessler FN, Middleton WD, Grant EG et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. J Am Coll Radiol 2017; 14: 587–595. DOI: 10.1016/j.jacr.2017.01.046

[11] Russ G, Bonnema SJ, Erdogan MF et al. European Thyroid Association Guidelines for Ultrasound Malignancy Risk Stratification of Thyroid Nodules in Adults: The EU-TIRADS. Eur Thyroid J 2017; 6: 225–237. DOI: 10.1159/000478927

[12] Tulecke MA, Wang HH. ThinPrep for cytologic evaluation of follicular thyroid lesions: correlation with histologic findings. Diagn Cytopathol 2004; 30: 7–13. DOI: 10.1002/dc.10391

[13] NCI Thyroid FNA State of the Science ConferenceCibas ES, Ali SZ. The Bethesda System For Reporting Thyroid Cytopathology. Am J Clin Pathol 2009; 132: 658–665. DOI: 10.1309/AJCPPHLWMI3JV4LA

[14] Liu Z, Lin Y, Cao Y et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows Available at:https://ui.adsabs.harvard.edu/abs/2021arXiv210314030L; Date Accessed: March 01, 2021

[15] Kornelius E, Lo SC, Huang CN et al. The risk of thyroid cancer in patients with thyroid nodule 3 Cm or larger. Endocr Pract 2020; 26: 1286–1290. DOI: 10.4158/EP-2020-0136

[16] Shayganfar A, Hashemi P, Esfahani MM et al. Prediction of thyroid nodule malignancy using thyroid imaging reporting and data system (TIRADS) and nodule size. Clin Imaging 2020; 60: 222–227. DOI: 10.1016/j.clinimag.2019.10.004

[17] Schiffmann L, Kostev K, Kalder M. Association between various thyroid gland diseases, TSH values and thyroid cancer: A case-control study. J Cancer Res Clin Oncol 2020; 146: 2989–2994. DOI: 10.1007/s00432-020-03283-x

[18] Haymart MR, Repplinger DJ, Leverson GE et al. Higher serum thyroid stimulating hormone level in thyroid nodule patients is associated with greater risks of differentiated thyroid cancer and advanced tumor stage. J Clin Endocrinol Metab 2008; 93: 809–814. DOI: 10.1210/jc.2007-2215

[19] Boelaert K, Horacek J, Holder RL et al. Serum thyrotropin concentration as a novel predictor of malignancy in thyroid nodules investigated by fine-needle aspiration. J Clin Endocrinol Metab 2006; 91: 4295–4301. DOI: 10.1210/jc.2006-0527

[20] Mekraksakit P, Rattanawong P, Karnchanasorn R et al. Prognosis of differentiated thyroid carcinoma in patients with Graves disease: A systematic review and meta-analysis. Endocr Pract 2019; 25: 1323–1337. DOI: 10.4158/EP-2019-0201

[21] Staniforth JUL, Erdirimanne S, Eslick GD. Thyroid carcinoma in Graves' disease: A meta-analysis. Int J Surg 2016; 27: 118–125. DOI: 10.1016/j.ijsu.2015.11.027

[22] Feldt-Rasmussen U. Hashimoto's thyroiditis as a risk factor for thyroid cancer. Curr Opin Endocrinol Diabetes Obes 2020; 27: 364–371. DOI: 10.1097/MED.0000000000000570

[23] Negro R, Valcavi R, Toulis KA. Incidental thyroid cancer in toxic and nontoxic goiter: Is TSH associated with malignanCy rate? Results of a meta-analysis. Endocr Pract 2013; 19: 212–218. DOI: 10.4158/EP12234.OR

[24] Zhang H, Wu C, Zhang Z et al. ResNeSt: Split-attention networks. Available at:https://ui.adsabs.harvard.edu/abs/2020arXiv200408955Z; Date Accessed: April 01, 2020

[25] Dai Y, Gao Y, Liu F. TransMed: Transformers advance multi-modal medical image classification. Diagnostics (Basel) 2021; 11:. DOI: 10.3390/diagnostics11081384

[26] Dosovitskiy A, Beyer L, Kolesnikov A et al. An image is worth 16x16 words: Transformers for image recognition at scale. Available at:https://ui.adsabs.harvard.edu/abs/2020arXiv201011929D; Date Accessed: October 01, 2020

[27] Thomas J, Haertling T. AIBx, artificial intelligence model to risk stratify thyroid nodules. Thyroid 2020; 30: 878–884. DOI: 10.1089/thy.2019.0752

[28] Ko SY, Lee JH, Yoon JH et al. Deep convolutional neural network for the diagnosis of thyroid nodules on ultrasound. Head Neck 2019; 41: 885–891. DOI: 10.1002/hed.25415

[29] Bomhof-Roordink H, Gartner FR, Stiggelbout AM et al. Key components of shared decision making models: A systematic review. BMJ Open 2019; 9: e031763. DOI: 10.1136/bmjopen-2019-031763

[30] Singh ONM, Bagautdinova D, Hargraves I et al. Development and pilot testing of a conversation aid to support the evaluation of patients with thyroid nodules. Clin Endocrinol 2022; 96: 627–636. DOI: 10.1111/cen.14599