European Journal of Radiology Open 13 (2024) 100609

Contents lists available at ScienceDirect

European Journal of Radiology Open

journal homepage: www.elsevier.com/locate/ejro

Deep learning model for diagnosis of thyroid nodules with size less than 1 cm: A multicenter, retrospective study

Na Feng [a,b,1], Shanshan Zhao [c,1], Kai Wang [d,1], Peizhe Chen [e], Yunpeng Wang [e], Yuan Gao [c], Zhengping Wang [d], Yidan Lu [a,b], Chen Chen [a,b], Jincao Yao [a,b,f,*], Zhikai Lei [g,*], Dong Xu [a,b,f,h,**]

[a] Department of Ultrasound, Zhejiang Cancer Hospital, Hangzhou 310022, China
[b] Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou 310000, China
[c] Department of Ultrasound, Shaoxing People's Hospital (Zhejiang University Shaoxing Hospital), Shaoxing 312300, China
[d] Department of Ultrasound, The Affiliated Dongyang Hospital of Wenzhou Medical University, Dongyang 322100, China
[e] College of Optical Science and Engineering, Zhejiang University, Hangzhou, Zhejiang Province, China
[f] Wenling Medical Big Data and Artificial Intelligence Research Institute, Taizhou 310061, China
[g] The First Affiliated Hospital of Zhejiang Chinese Medical University (Zhejiang Provincial Hospital of Chinese Medicine), Hangzhou 310003, China
[h] Department of Ultrasound, Taizhou Cancer Hospital, Taizhou 310022, China

HIGHLIGHTS

- Introduces TNT-Net, a novel AI model for improved diagnosis of small thyroid nodules (<1 cm).
- Provides radiologists with a standardized tool for more consistent nodule assessments.
- Utilizes AI technology to help avoid unnecessary thyroid nodule biopsies and overtreatment.
- Offers more objective risk evaluation to optimize patient management.

ARTICLE INFO

Keywords:
Thyroid nodules
Deep Learning
Transformer
Early diagnosis
Ultrasound image

ABSTRACT

Objective: To develop a ultrasound images based dual-channel deep learning model to achieve accurate early diagnosis of thyroid nodules less than 1 cm.
Methods: A dual-channel deep learning model called thyroid nodule transformer network (TNT-Net) was proposed. The model has two input channels for transverse and longitudinal ultrasound images of thyroid nodules, respectively. A total of 9649 nodules from 8455 patients across five hospitals were retrospectively collected. The data were divided into a training set (8453 nodules, 7369 patients), an internal test set (565 nodules, 512 patients), and an external test set (631 nodules, 574 patients).
Results: TNT-Net achieved an area under the curve (AUC) of 0.953 (95 % confidence interval (CI): 0.934, 0.969) on the internal test set and 0.941 (95 % CI: 0.921, 0.957) on the external test set, significantly outperforming traditional deep convolutional neural network models and single-channel swin transformer model, whose AUCs ranged from 0.800 (95 % CI: 0.759, 0.837) to 0.856 (95 % CI: 0.819, 0.881). Furthermore, feature heatmap visualization showed that TNT-Net could extract richer and more energetic malignant nodule patterns.
Conclusion: The proposed TNT-Net model significantly improved the recognition capability for thyroid nodules with size less than 1 cm. This model has the potential to reduce overdiagnosis and overtreatment of such nodules, providing essential support for precise management of thyroid nodules while complementing fine-needle aspiration biopsy.

* Correspondence to: Department of Ultrasound, Zhejiang Cancer Hospital, No.1 Banshan East Road, Hangzhou 310022, China.
** Correspondence to: Zhejiang Provincial Hospital of Chinese Medicine, 54 Youdian Road, Hangzhou 310003, China.
E-mail addresses: yaojc@zjcc.org.cn (J. Yao), lzk110@126.com (Z. Lei), xudong@zjcc.org.cn (D. Xu).
[1] #the authors share the first authorship

https://doi.org/10.1016/j.ejro.2024.100609
Received 12 August 2024; Received in revised form 20 October 2024; Accepted 29 October 2024

1. Introduction

Thyroid nodules are a common endocrine disorder, with the latest epidemiological studies showing that the prevalence of thyroid nodules in adults can reach up to 68 %, and the incidence rate in females is three times higher than in males [1–4]. Most of these nodules are benign, with only approximately 7–15 % being malignant [3,4]. Therefore, early and accurate diagnosis of the nature of thyroid nodules is crucial for clinical decision-making and management of this disease.

Currently, the diagnosis of thyroid nodules, especially those smaller than 1 cm, poses significant challenges. Conventional ultrasound examination relies heavily on the experience of radiologists, leading to potential inconsistencies and subjectivity in diagnosis[5–9]. Moreover, fine-needle aspiration biopsy (FNAB), while considered the gold standard, is invasive and not suitable for large-scale screening [7,10–12]. This situation highlights the need for more objective, non-invasive, and efficient diagnostic methods.

To address this issue, many studies have begun to explore the integration of AI technologies to develop computer-aided diagnosis (CAD) models based on ultrasound images and deep learning techniques, such as ThyNet, RedImageNet, and DeepThyNet [13–19]. In general, these models are typically trained on a large amount of ultrasound image data to extract latent image features that are difficult for the human eye to perceive, enabling the identification of malignant patterns in thyroid nodules. These ultrasound image-based CAD models have improved the diagnostic accuracy of thyroid nodules to a certain extent.

Although previous studies have explored various deep convolutional neural network (DCNN) models for the intelligent diagnosis of thyroid nodules [20–24], their classification performance often proved unsatisfactory when dealing with smaller thyroid nodules, particularly those with a size less than 1 cm [21,23]. As smaller thyroid nodules occupy a smaller area in ultrasound images, the limited ability of traditional DCNN convolutional kernels to handle smaller targets often leads to an insufficient capture of effective features, resulting in a relatively low accuracy in assessing the malignancy risk of such nodules. How to construct a model that can effectively extract subtle features from ultrasound images that characterize the properties of small thyroid nodules and accurately assess the malignancy risk remains a challenging problem.

Recently, with the development of multimodal large models like ChatGPT, a new class of deep learning models based on the Transformer architecture has emerged [25–29]. Transformer significantly improve the ability of deep learning models to capture image features, especially those of small targets, using the self-attention mechanism [29]. Unlike traditional DCNN networks, the Transformer architecture overcomes the limitations of convolutional kernels and can directly establish mappings between arbitrary pixels or pixel blocks [30–32]. Compared to DCNN, this feature of the Transformer architecture offers a significant advantage in handling small nodules and capturing subtle features. This characteristic also presents new opportunities for addressing the computer-aided diagnosis problem of small thyroid nodules.

Therefore, based on the aforementioned techniques, this study proposes a dual-channel swin transformer (ST) model called the Thyroid Nodule Transformer Network (TNT-Net). The newly proposed model has two input interfaces for inputting transverse and longitudinal ultrasound images of the nodule. The model first thoroughly exploits the detailed information in the transverse and longitudinal ultrasound images of the nodule through two independent self-attention mechanism branch networks. It then fuses the features from the branch networks using sparse constraints and sends the fused features to the fully connected network at the backend of the model to perform recognition. Experimental results show that compared to traditional DCNN models, our model demonstrates better performance in evaluating nodules with size less than 1 cm.

This study is an extension of our previous work [21,22], and to the best of our knowledge, it is also the first exploration of utilizing a Transformer to simultaneously extract features from transverse and longitudinal ultrasound images to assess the risk of thyroid nodules. We not only constructed a classification model based on the visual Transformer but also performed comparative experiments with mainstream DCNN models to comprehensively evaluate the performance of both types of models on this task. This study aims to develop and evaluate a novel dual-channel deep learning model, TNT-Net, for improved diagnosis of thyroid nodules less than 1 cm. The goal is to assess whether this new architecture can enhance classification accuracy and robustness compared to traditional DCNN models, potentially offering a valuable reference for future applications of Transformer-based architectures in medical imaging tasks.

2. Materials and methods

2.1. Patients

This was a retrospective, multicenter diagnostic study that utilized thyroid nodule ultrasound imaging data from October 2019 to October 2022, collected from five hospitals in China. The study was approved by the ethics committees of all participating hospitals, and due to the retrospective nature of the study, informed consent was waived. All data were anonymized, and the model was trained and tested solely within the hospitals' artificial intelligence infrastructure. The hospitals participating in this multicenter study and their ethics committee approval numbers are as follows: Zhejiang Cancer Hospital (IRB-2020–287), Taizhou Cancer Hospital (IRB-2023001), Zhejiang Provincial Hospital of Traditional Chinese Medicine (IRB-2023-QS-010–02), Shaoxing People's Hospital (IRB-2022–083-Y-01), and Dongyang People's Hospital (IRB-2023-YX-355). Table 1 provides an overview of all the data. The study included a total of 9649 nodules from 8455 individuals.

2.2. Inclusion and exclusion criteria

Fig. 1 illustrates the overall inclusion and exclusion of data, as well as the division of the training and test sets. The specific data inclusion criteria were: (1) Patient age greater than or equal to 18 years. (2) Patients underwent preoperative thyroid ultrasound examination, and nodule images were preserved. (3) Patients had a definitive pathological result (underwent total or partial thyroidectomy, or the FNAB result of the affected side was Bethesda class II or VI). (4) Nodule size was less than 1 cm. The

specific data exclusion criteria were: (1) Preoperative thyroid ultrasound examination lacked clear ultrasound images, such as missing either the transverse or longitudinal ultrasound image of the nodule. (2) The patient received other treatments (such as radiation therapy) before surgery. (3) Patients had incomplete clinical information, such as unknown prior treatment records. The biopsy criteria adopted by multicenter hospitals are based on guidelines from the American Thyroid Association (ATA) [33]. For the majority of nodules smaller than 1 cm, particularly those without suspicious characteristics, regular ultrasound follow-up is generally recommended rather than immediate FNAB. However, radiologists also conduct comprehensive assessments considering patient-specific factors. For instance, FNAB may be recommended even for nodules smaller than 1 cm in cases where: abnormal cervical lymph nodes are detected during follow-up ultrasonography; there is a history of childhood neck radiation exposure or radiation contamination; a family history of thyroid cancer or thyroid cancer syndrome is present; 18F-FDG PET imaging yields positive results; or there is an abnormal elevation in serum calcitonin levels.

2.3. Model design

The overall architecture of the deep learning model is shown in Fig. 2, and the model code has been made publicly available (see

Table 1
Statistics of multicenter data. SD: Standard Deviation; Center 1: Zhejiang Cancer Hospital; Center 2: Taizhou Cancer Hospital; Center 3: Zhejiang Provincial Hospital of Traditional Chinese Medicine; Center 4: Shaoxing People's Hospital; Center 5: Dongyang People's Hospital.

Categories	Center 1	Center 2	Center 3	Center 4	Center 5
Number of patients	5650	1216	1015	263	311
Number of nodules	6402	1367	1249	308	323
Number of images	12804	2734	2498	616	646
Age, mean (SD)	49 (13)	50 (11)	47 (12)	52 (13)	49 (11)
Sex					
Female	4184	920	770	195	232
Male	1466	296	245	68	79
Size, mm (SD)	6.5 (1.7)	6.3 (2.1)	6.6 (1.9)	6.2 (1.6)	6.3 (1.9)
Pathological results					
Benign	3227	766	655	162	164
Malignant	3175	601	594	146	159

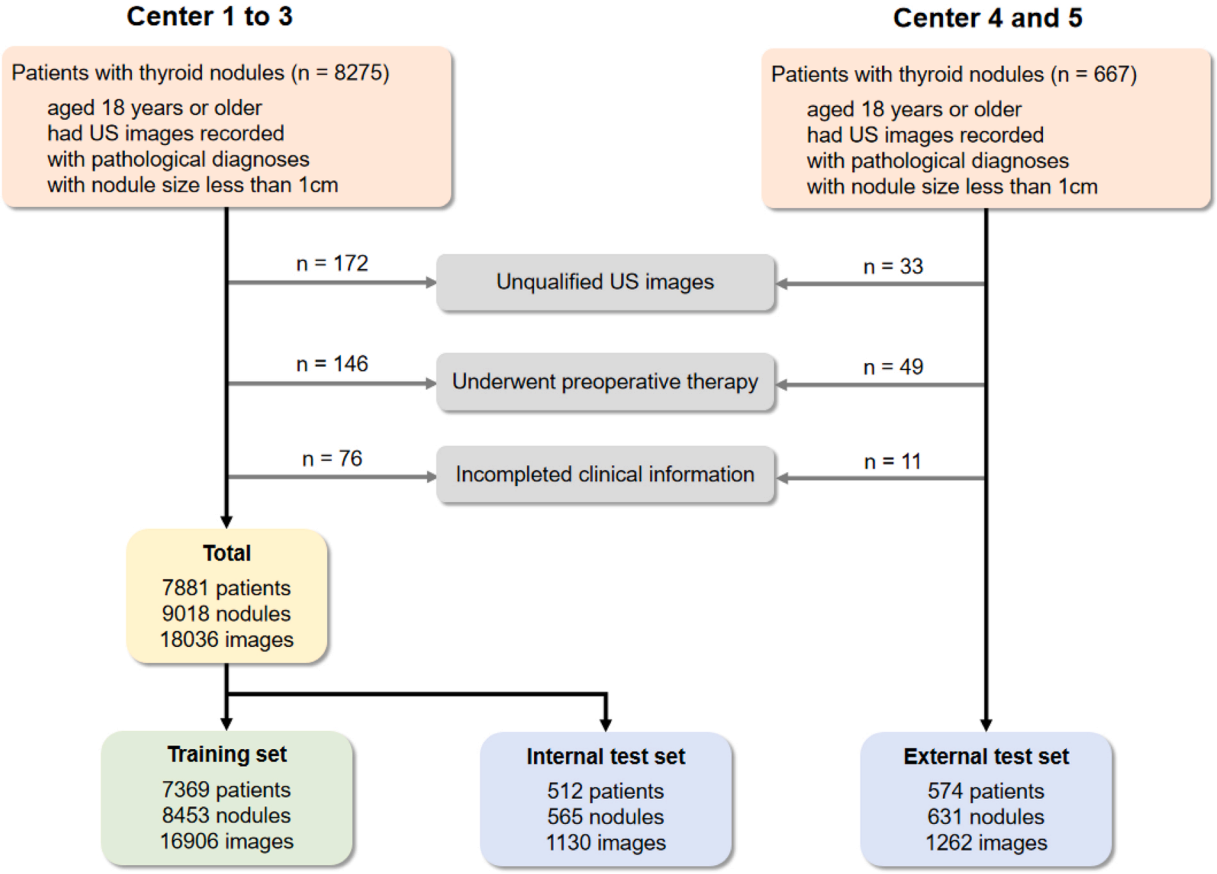

Fig. 1. Multicenter data collection and establishment of training and test sets.

Supplementary Material Section 1). As illustrated in Fig. 2, we collected transverse and longitudinal images of thyroid nodules and annotated the nodule regions in the images. To address the challenge of capturing features from the relatively small targets of thyroid nodules, we designed a dual-channel network model called TNT-Net.

The model consists of three main parts: image patching, backbone module, and feature fusion module. In the backbone part, we designed two independent feature extraction channels. The first channel takes the transverse ultrasound image of the nodule as input, while the other channel takes the longitudinal ultrasound image of the nodule as input. Considering the outstanding ability of Transformers in extracting features from small targets, we chose the Swin Transformer as the backbone network for each channel. To better fuse the features from different channels, we proposed a dual-channel joint sparse attention mechanism in the fully connected (FC) layer, thereby enhancing the fusion of transverse and longitudinal features extracted by the two channels of the model. The specific feature constraint method and formulas

are provided in Supplementary Material Section 2. For better training and testing of the model, we established a training set, an internal test set, and an external test set, with none of the test data being involved in model training. The training set included 8453 nodules from 7369 patients, the internal test set included 565 nodules from 512 patients, and the external test set included 631 nodules from 574 patients.

2.4. Data acquisition

During ultrasound image acquisition, patients were placed in a supine position and asked to extend their neck and temporarily stop swallowing to fully expose the neck area. Ultrasound imaging equipment included GE, Siemens, Toshiba, and Mindray (see Supplementary Material Section 3 for details). During image acquisition, the nodule was positioned at the center of the image as much as possible to ensure clear

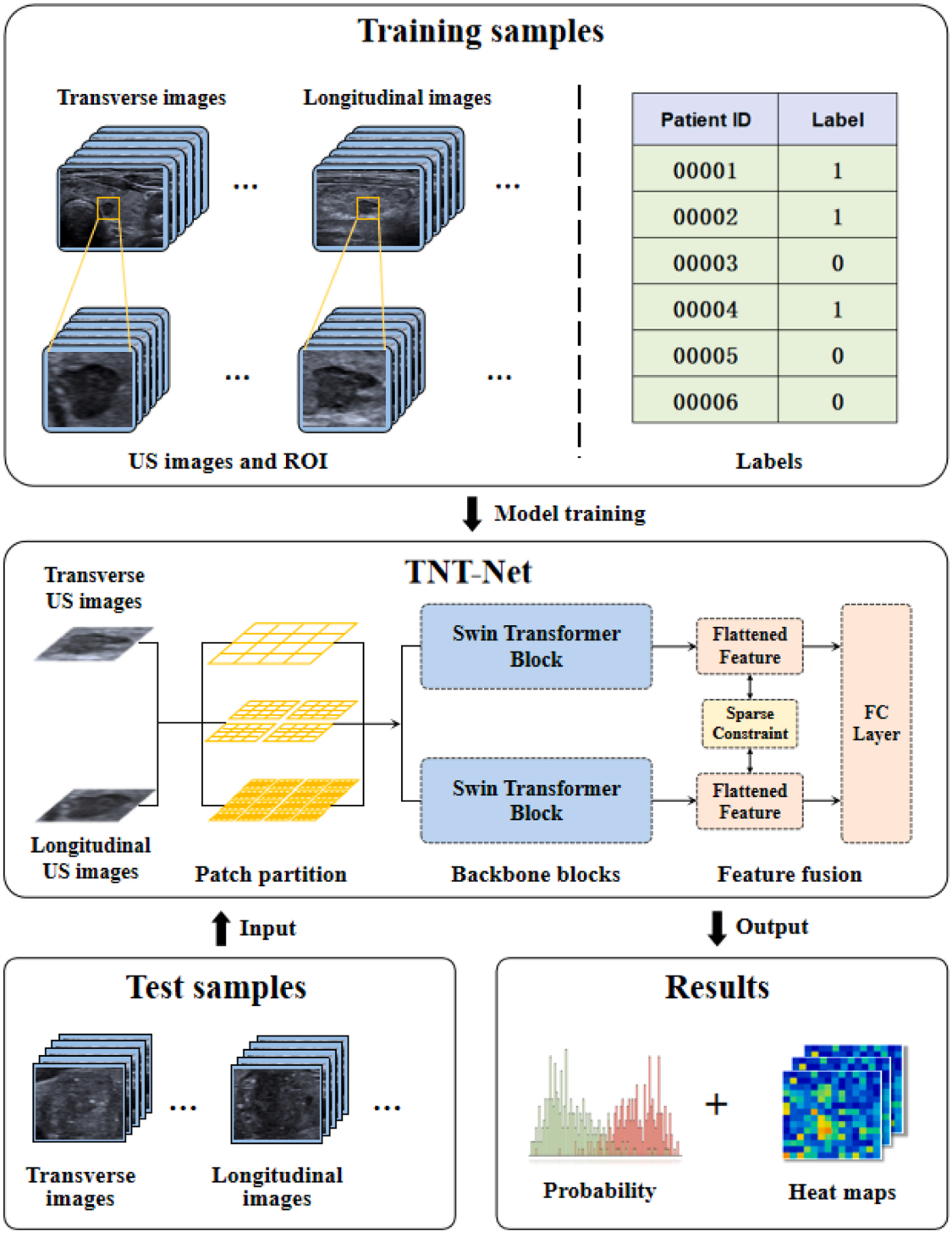

Fig. 2. Model framework and training and testing pipeline of TNT-Net.

visibility and optimal image quality. The acquired images were stored in DICOM or BMP format. In addition to the raw image data, relevant clinical information for each nodule, such as patient age, gender, and nodule size, was recorded. All included data underwent quality control by at least one radiologist.

2.5. Data statistics

We used various metrics to evaluate the model's performance, including receiver operating characteristic (ROC), area under the curve (AUC), true positive rate (TPR), true negative rate (TNR), positive predictive value (PPV), negative predictive value (NPV), and F1 score. Furthermore, we calculated the p-values between our proposed TNT-Net and traditional methods. The DeLong method was used to calculate the 95 % confidence intervals (CIs) for the metrics [34]. All model and statistical computations were performed using the Python programming language, utilizing libraries such as PyTorch, Matplotlib, NumPy, and Scikit-learn.

3. Results

We evaluated and compared the performance of ResNet50, DenseNet121, Inception V3, ST and TNT-Net. Fig. 3 shows the ROC curves and corresponding AUC values of different models on the internal test set. As can be seen from Fig. 3(a), our TNT-Net achieved the best diagnostic performance. Table 2 provides a detailed comparison of the diagnostic abilities of different CAD models on the internal test set. As shown in Table 2, the AUCs of the DCNN network models ResNet50, DenseNet121, and Inception V3 ranged from 0.800 (95 % CI: 0.759, 0.837) to 0.833 (95 % CI: 0.793, 0.868). The single-channel ST model had an AUC of 0.856 (95 % CI: 0.819, 0.881), slightly outperforming the DCNN networks. Our proposed TNT-Net model achieved the best recognition performance, with an AUC of 0.953 (95 % CI: 0.934, 0.969) on the internal test set, significantly better than the best DCNN ($P < 0.001$). Similarly, on the external test set, TNT-Net achieved an AUC of 0.941 (95 % CI: 0.921, 0.957), outperforming both the ST model and DCNN models (Fig. 3(b)). The similar ROC performance on the internal and external test sets also cross-validated the reliability of the model. Table 1 further provides the TNR, TPR, PPV, NPV, ACC, and F1 scores, along with their 95 % CIs, for the different models.

As shown in Tables 2 and 3, TNT-Net achieved a sensitivity of 0.912 (95 % CI: 0.885, 0.932) and a specificity of 0.925 (95 % CI: 0.887, 0.951) on the internal test set. On the external independent test set, the classification sensitivity and specificity were slightly lower than those on the internal test set, at 0.905 (95 % CI: 0.867, 0.933) and 0.911 (95 % CI: 0.875, 0.937), respectively. The comparisons in Tables 2 and 3 show that the overall performance of TNT-Net was significantly better than that of traditional DCNN models and the ST model ($p < 0.001$).

To better illustrate the model's discrimination results, Fig. 4 also compares the feature heatmaps of TNT-Net, ST, and a representative DCNN model (Inception V3). The high-heat regions of the feature heatmaps intuitively show the malignancy risk-related feature patterns extracted by the deep learning models. In Fig. 4, columns 1–4 represent malignant nodule images, and columns 5–8 represent benign nodule samples. Each sample includes the transverse (column 1 for each sample) and longitudinal (column 2 for each sample) images of the nodule. From the comparison of the feature heatmaps, we can observe that TNT-Net exhibited stronger feature discrimination ability compared to the ST model and DCNN networks. For malignant samples, our proposed TNT-Net extracted more energetic and richer features, while for benign nodules, our model yielded more consistent low-energy feature heatmaps. In contrast, the malignant features extracted by the ST model and Inception V3 had relatively weaker energy and discrimination, and the feature heatmaps for benign nodules exhibited noticeable interference. For example, the feature heatmap of benign sample 1 extracted by the ST model and Inception V3 model, and the longitudinal feature heatmap of benign sample 2 extracted by the Inception V3 model showed interference. These comparisons suggest that TNT-Net does not simply accumulate the energy of the nodule feature heatmaps but rather enhances the overall ability of the model to extract subtle nodule features through joint training on the transverse and longitudinal ultrasound images of the nodules.

4. Discussion

This study proposed a new Transformer-based dual-channel deep learning model, TNT-Net, for diagnosing the malignancy of thyroid nodules with size less than 1 cm. The model can simultaneously utilize the transverse and longitudinal ultrasound images of the nodule, capturing its features through the self-attention mechanism, significantly improving the recognition capability. On multicenter test datasets, TNT-Net consistently outperformed traditional DCNN models in terms of diagnostic performance, achieving better results in AUC, classification accuracy, and feature extraction ability.

The superior performance of the TNT-Net model is mainly attributed to the introduction of the dual-channel structure. By simultaneously inputting the transverse and longitudinal ultrasound images of the nodule, TNT-Net can comprehensively acquire the nodule's multidimensional information, thereby better extracting the internal morphological and structural features of the nodule. In contrast, traditional DCNN models such as ThyNet, DeepThyNet, RedimageNet typically utilize images from a single viewpoint, potentially overlooking some effective information [13,14,16]. Moreover, the Transformer architecture can directly establish long-range dependencies between any two pixel blocks in the image, better capturing the detailed features of small nodules [29]. This advantage gives TNT-Net a distinct advantage in handling smaller nodules. Furthermore, through the design of the feature fusion module, TNT-Net can effectively integrate feature information from the two channels, further enhancing the model's discrimination ability. This module adopts a constraint optimization method, avoiding the redundancy caused by simple concatenation. The training and testing on a dataset containing 9649 nodules from five hospitals also provided reliable data support for evaluating the model's performance.

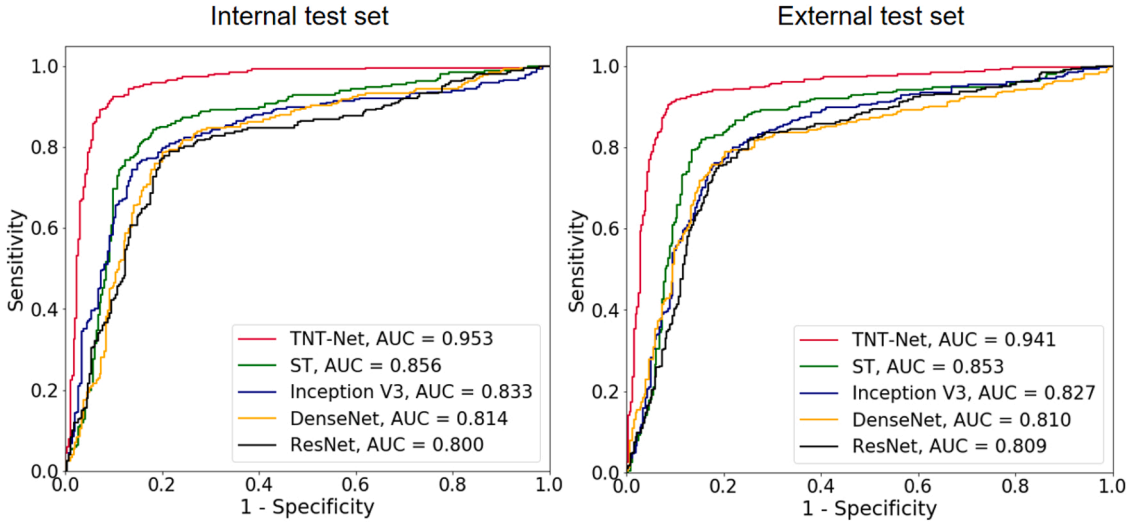

Fig. 3. Comparison of ROC curves and AUC values of TNT-Net, ST model, and traditional DCNN models on the internal test set and external test set. The left panel shows the results for the internal test set. The right panel shows the results for the external test set. TNT-Net achieved the best performance on both the internal and external test sets, with the red curve representing the ROC curve of TNT-Net.

Table 2
Comparison of the classification performance of TNT-Net, ST model, and traditional DCNN models on the internal test set.

Model	TPR (95 % CI)	TNR (95 % CI)	PPV (95 % CI)	NPV (95 % CI)	ACC (95 % CI)	F1
TNT-Net	0.912 (0.885,0.932)	0.925 (0.887,0.951)	0.899 (0.860,0.929)	0.881 (0.835,0.915)	0.937 (0.905,0.959)	0.908
ST	0.828 (0.795,0.857)	0.816 (0.766,0.858)	0.839 (0.793,0.876)	0.803 (0.749,0.847)	0.851 (0.808,0.885)	0.818
Inception V3	0.793 (0.758,0.824)	0.772 (0.718,0.818)	0.812 (0.764,0.852)	0.767 (0.711,0.816)	0.816 (0.770,0.854)	0.779
DenseNet121	0.789 (0.754,0.821)	0.779 (0.726,0.825)	0.799 (0.749,0.840)	0.756 (0.700,0.805)	0.818 (0.772,0.857)	0.778
ResNet50	0.786 (0.750,0.818)	0.775 (0.722,0.821)	0.795 (0.746,0.837)	0.753 (0.696,0.802)	0.815 (0.769,0.854)	0.774

Table 3
Comparison of the classification performance of TNT-Net, ST model, and traditional DCNN models on the external test set.

Model	TPR (95 % CI)	TNR (95 % CI)	PPV (95 % CI)	NPV (95 % CI)	ACC (95 % CI)	F1
TNT-Net	0.905 (0.867,0.933)	0.911 (0.875,0.937)	0.899 (0.859,0.929)	0.916 (0.882,0.941)	0.908 (0.883,0.928)	0.905
ST	0.826 (0.780,0.865)	0.819 (0.774,0.857)	0.800 (0.751,0.841)	0.843 (0.801,0.878)	0.823 (0.791,0.850)	0.819
Inception V3	0.777 (0.727,0.820)	0.791 (0.744,0.832)	0.765 (0.714,0.810)	0.802 (0.757,0.841)	0.784 (0.751,0.815)	0.777
DenseNet121	0.784 (0.734,0.826)	0.798 (0.751,0.838)	0.772 (0.721,0.816)	0.808 (0.763,0.846)	0.791 (0.757,0.821)	0.784
ResNet50	0.764 (0.713,0.808)	0.788 (0.741,0.829)	0.760 (0.707,0.805)	0.792 (0.746,0.832)	0.777 (0.742,0.807)	0.768

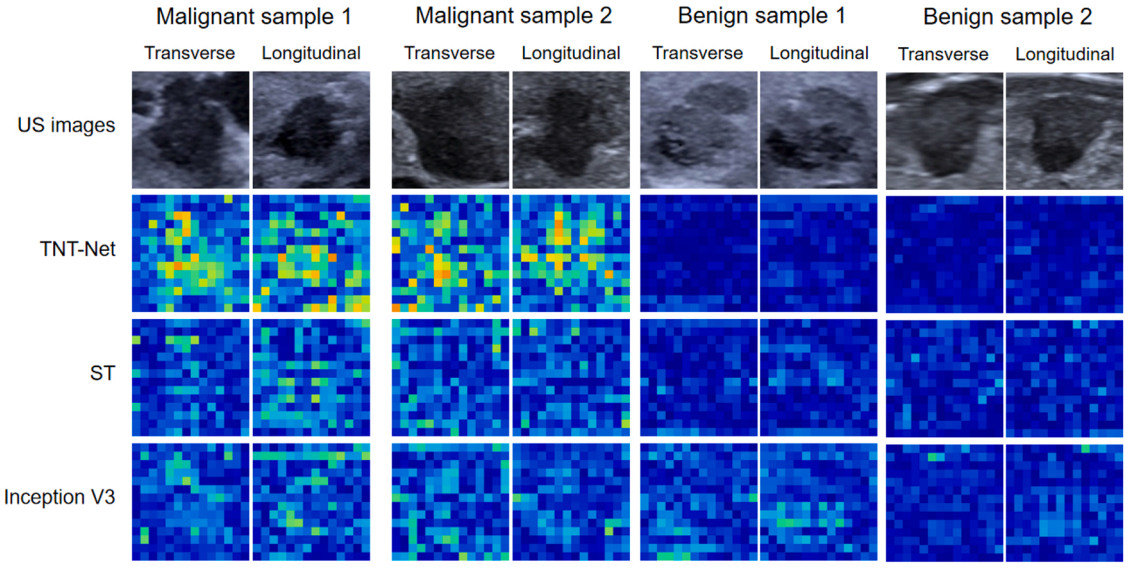

Fig. 4. Comparison of feature heatmaps between TNT-Net, ST model, and traditional Inception V3 model. High-heat regions correspond to features with higher malignancy risk. For each sample, the left side shows the transverse image and feature heatmap, while the right side shows the longitudinal image and feature heatmap. For TNT-Net, the calculation involves inputting both transverse and longitudinal images, resulting in two feature heatmaps. For other models, a single image is inputted to calculate the corresponding feature heatmap.

TNT-Net can be easily integrated into clinical practice and early screening for thyroid cancer. Since many thyroid nodules detected during early screening are small in size, and studies have shown that radiologists have relatively lower diagnostic concordance rates for smaller thyroid nodules [2,9,10]. As a standardized AI-assisted diagnostic tool, TNT-Net can provide more objective and consistent references for clinical decision-making, helping to avoid overdiagnosis and overtreatment caused by relying solely on subjective judgments, particularly for junior radiologists. For instance, in an actual screening environment, when the diagnosis of a junior radiologist conflicts with the model-assisted diagnosis, they could consider referring to the model's prediction probability and the feature heatmaps for transverse and longitudinal images. If a significant discrepancy is observed between the model's prediction probability, heatmaps, and the radiologist's assessment, a re-examination of the nodule or consultation with a more experienced radiologist could be considered to mitigate the risk of false positives and false negatives.

Furthermore, our analysis showed that the TNT-Net model performed consistently across both male and female patients. The model achieved an AUC of 0.942 (95 % CI: 0.919, 0.966) for female patients and 0.939 (95 % CI: 0.917, 0.962) for male patients on the external test set. This suggests that the model's performance is not significantly affected by gender, which is crucial for ensuring equitable diagnostic capabilities across patient demographics.

The study has some limitations. First, although we attempted to provide some comparisons and interpretability analysis through feature heatmaps, the "black box" nature of deep learning models and the difficulty in explaining features remain pressing issues that need to be addressed. Second, to obtain the pathological gold standard, we had to select cases that underwent surgery or biopsy and had definitive pathological results, which introduces a certain selection bias. Third, in clinical practice, radiologists can obtain dynamic information when interpreting images, whereas our study only utilized two static images, which to some extent results in the loss of some nodule information.

5. Conclusions

In conclusion, our study demonstrates that the new generation of Transformer deep learning models based on the self-attention mechanism shows promising application prospects in assisting the diagnosis of

thyroid nodules smaller than 1 cm. These models have the potential to reduce unnecessary biopsies and overtreatment while supporting precise disease management. In future research, we plan to incorporate dynamic ultrasound data and integrate clinical information with imaging features to provide a more comprehensive risk assessment. Additionally, we aim to conduct prospective studies to validate the model's performance in real-time clinical settings and investigate its interpretability to offer deeper insights into the decision-making process.

Funding statement

This work was supported by the National Key Research and Development Program of China (2022YFF0608403), the Pioneer and Leading Goose R&D Program of Zhejiang (2023C04039), the National Natural Science Foundation of China (82071946), the Research Program of Zhejiang Provincial Department of Health (2022KY110, 2023KY066, 2023XY066).

Ethical statement

The hospitals participating in this multicenter study and their ethics committee approval numbers are as follows: Zhejiang Cancer Hospital (IRB-2020–287), Taizhou Cancer Hospital (IRB-2023001), Zhejiang Provincial Hospital of Traditional Chinese Medicine (IRB-2023-QS-010–02), Shaoxing People's Hospital (IRB-2022–083-Y-01), and Dongyang People's Hospital (IRB-2023-YX-355).

CRediT authorship contribution statement

Peizhe Chen: Validation, Methodology, Formal analysis, Data curation. **Yunpeng Wang:** Validation, Methodology, Investigation, Formal analysis, Data curation. **Yuan Gao:** Validation, Resources, Formal analysis, Data curation. **Yidan Lu:** Validation, Resources, Methodology, Formal analysis. **Zhikai Lei:** Writing – review & editing, Validation, Resources, Methodology, Investigation, Data curation, Conceptualization. **Dong Xu:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Na Feng:** Writing – review & editing, Writing – original draft, Resources, Investigation, Funding acquisition, Data curation, Conceptualization. **Shanshan Zhao:** Writing – original draft, Resources, Methodology, Investigation, Data curation, Conceptualization. **Kai Wang:** Writing – original draft, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Zhengping Wang:** Validation, Resources, Methodology, Investigation, Data curation. **Jincao Yao:** Writing – review & editing, Visualization, Supervision, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Chen Chen:** Validation, Resources, Methodology, Formal analysis, Data curation.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejro.2024.100609.

References

[1] D.W. Chen, B.H.H. Lang, D.S.A. McLeod, K. Newbold, M.R. Haymart, Thyroid cancer, Lancet 401 (10387) (2023) 1531–1544, https://doi.org/10.1016/S0140-6736(23)000, 20-X.
[2] L. Boucai, M. Zafereo, M.E. Cabanillas, Thyroid cancer: a review, JAMA 331 (5) (2024) 425–435, https://doi.org/10.1001/jama.2023.26348.
[3] R.L. Siegel, A.N. Giaquinto, A. Jemal, Cancer statistics, 2024, CA: a Cancer J. Clin. 74 (1) (2024) 12–49, https://doi.org/10.3322/caac.21820.
[4] E.D. Rossi, Z. Baloch, The impact of the 2022 WHO classification of thyroid neoplasms on everyday practice of cytopathology, Endocr. Pathol. 34 (1) (2023) 23–33, https://doi.org/10.1007/s12022-023-09756-2.
[5] J.A. Fagin, Y.E. Nikiforov, Progress in thyroid cancer genomics: a 40-year journey, Thyroid 33 (11) (2023) 1271–1286, https://doi.org/10.1089/thy.2023.0045.
[6] P. Sun, Y. Wei, C. Chang, J. Du, Y. Tong, Ultrasound-based nomogram for predicting the aggressiveness of papillary thyroid carcinoma in adolescents and young adults, Acad. Radiol. 31 (2) (2024) 523–535, https://doi.org/10.1016/j.acra.2023.05.009.
[7] F. Pitoia, P. Trimboli, New insights in thyroid diagnosis and treatment, Rev. Endocr. Metab. Disord. 25 (1) (2024) 1–3, https://doi.org/10.1007/s11154-023-09859-5.
[8] R.A. Levine, History of thyroid ultrasound, Thyroid 33 (8) (2023) 894–902, https://doi.org/10.1089/thy.2022.0346.
[9] Y. Lee, M.R. Alam, H. Park, K. Yim, K.J. Seo, G. Hwang, D. Kim, Y. Chung, G. Gong, N.H. Cho, C.W. Yoo, Y. Chong, H.J. Choi, Improved diagnostic accuracy of thyroid fine-needle aspiration cytology with artificial intelligence technology, Thyroid 34 (6) (2024) 723–734, https://doi.org/10.1089/thy.2023.0384.
[10] J. Bojunga, P. Trimboli, Thyroid ultrasound and its ancillary techniques, Rev. Endocr. Metab. Disord. 25 (1) (2024) 161–173, https://doi.org/10.1007/s11154-023-09841-1.
[11] J. Wang, N. Zheng, H. Wan, Q. Yao, S. Jia, X. Zhang, S. Fu, J. Ruan, G. He, X. Chen, S. Li, R. Chen, B. Lai, J. Wang, Q. Jiang, N. Ouyang, Y. Zhang, Deep learning models for thyroid nodules diagnosis of fine-needle aspiration biopsy: a retrospective, prospective, multicentre study in China, Lancet Digit. Health 6 (7) (2024) e458–e469, https://doi.org/10.1016/S2589-7500(24)00085-2.
[12] S.Z. Ali, Z.W. Baloch, B. Cochand-Priollet, F.C. Schmitt, P. Vielh, P.A. VanderLaan, The 2023 Bethesda System for reporting thyroid cytopathology, J. Am. Soc. Cytopathol. 12 (5) (2023) 319–325, https://doi.org/10.1016/j.jasc.2023.05.005.
[13] S. Peng, Y. Liu, W. Lv, L. Liu, Q. Zhou, H. Yang, J. Ren, G. Liu, X. Wang, X. Zhang, et al., Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study, Lancet Digit. Health 3 (2021) e250–e259, https://doi.org/10.1016/S2589-7500(21)00041-8.
[14] X. Mei, Z. Liu, P.M. Robson, B. Marinelli, M. Huang, A. Doshi, A. Jacobi, C. Cao, K. E. Link, T. Yang, et al., RadImageNet: an open radiologic deep learning research dataset for effective transfer learning, Radiol. Artif. Intell. 4 (2022) e210315, https://doi.org/10.1148/ryai.210315.
[15] S. Bhattacharya, R.K. Mahato, S. Singh, G.K. Bhatti, S.S. Mastana, J.S. Bhatti, Advances and challenges in thyroid cancer: The interplay of genetic modulators, targeted therapies, and AI-driven approaches, Life Sci. 332 (2023) 122110, https://doi.org/10.1016/j.lfs.2023.122110.
[16] Anonymous, DeepThy-Net: a multimodal deep learning method for predicting cervical lymph node metastasis in papillary thyroid cancer. Adv. Intell. Syst. 4 (10) (2022) 2200100 https://doi.org/10.1002/aisy.202200100.
[17] L.Q. Zhou, S.E. Zeng, J.W. Xu, W.Z. Lv, D. Mei, J.J. Tu, F. Jiang, X.W. Cui, C. F. Dietrich, Deep learning predicts cervical lymph node metastasis in clinically node-negative papillary thyroid carcinoma, Insights Imaging 14 (1) (2023) 222, https://doi.org/10.1186/s13244-023-01550-2.
[18] J. Wang, C. Dong, Y.Z. Zhang, L. Wang, X. Yuan, M. He, S. Xu, Q. Zhou, J. Jiang, A novel approach to quantify calcifications of thyroid nodules in US images based on deep learning: predicting the risk of cervical lymph node metastasis in papillary thyroid cancer patients, Eur. Radiol. 33 (12) (2023) 9347–9356, https://doi.org/10.1007/s00330-023-09909-1.
[19] Z. Wang, X. Wang, T. Wang, J. Qiu, W. Lu, Localization and risk stratification of thyroid nodules in ultrasound images through deep learning, Ultrasound Med. Biol. 50 (6) (2024) 882–887, https://doi.org/10.1016/j.ultrasmedbio.2024.02.013.
[20] C.P. Fu, M.J. Yu, Y.S. Huang, C.S. Fuh, R.F. Chang, Stratifying high-risk thyroid nodules using a novel deep learning system, Exp. Clin. Endocrinol. Diabetes 131 (10) (2023) 508–514, https://doi.org/10.1055/a-2122-5585.
[21] Anonymous, The clinical value of artificial intelligence in assisting junior radiologists in thyroid ultrasound: a multicenter prospective study from real clinical practice, BMC Med. 22 (1) (2024) 293, https://doi.org/10.1186/s12916-024-03510-z.
[22] Anonymous, Deep learning to assist composition classification and thyroid solid nodule diagnosis: a multicenter diagnostic study, Eur. Radiol. 34 (4) (2024) 2323–2333, https://doi.org/10.1007/s00330-023-10269-z.
[23] Anonymous, The auxiliary diagnosis of thyroid echogenic foci based on a deep learning segmentation model: A two-center study, Eur. J. Radiol. 167 (2023) 111033, https://doi.org/10.1016/j.ejrad.2023.111033.
[24] Anonymous, AI diagnosis of Bethesda category IV thyroid nodules, iScience 26 (11) (2023) 108114, https://doi.org/10.1016/j.isci.2023.108114.
[25] S. Kalidindi, J. Baradwaj, Advancing radiology with GPT-4: Innovations in clinical applications, patient engagement, research, and learning, Eur. J. Radiol. Open 13 (2024) 100589, https://doi.org/10.1016/j.ejro.2024.100589.
[26] A. Anaya-Isaza, L. Mera-Jiménez, L. Verdugo-Alejo, L. Sarasti, Optimizing MRI-based brain tumor classification and detection using AI: A comparative analysis of neural networks, transfer learning, data augmentation, and the cross-transformer network, Eur. J. Radiol. Open 10 (2023) 100484, https://doi.org/10.1016/j.ejro.2023.100484.
[27] S.C. Fanni, C. Romei, G. Ferrando, F. Volpi, C.A. D'Amore, C. Bedini, S. Ubbiali, S. Valentino, E. Neri, Natural language processing to convert unstructured COVID-19 chest-CT reports into structured reports, Eur. J. Radiol. Open 11 (2023) 100512, https://doi.org/10.1016/j.ejro.2023.100512.

[28] L. Meomartino, A. Greco, M. Di Giancamillo, A. Brunetti, G. Gnudi, Imaging techniques in Veterinary Medicine. Part I: Radiography and Ultrasonography, Eur. J. Radiol. Open 8 (2021) 100382, https://doi.org/10.1016/j.ejro.2021.100382.
[29] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: hierarchical vision transformer using shifted windows, Proc. IEEE/CVF Int. Conf. Comput. Vis. 1 (2021) 10012–10022, https://doi.org/10.48550/arXiv.2103.14030.
[30] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. 2016 IEEE Conference on computer vision and pattern recognition (CVPR), Las Vegas, NV, USA (2016) 770–778, https://doi.org/10.1109/CVPR.2016.90.
[31] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K.Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708). https://doi.org/10.1109/CVPR.2017.243.
[32] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (2016) 2818–2826, https://doi.org/10.1109/CVPR.2016.308.
[33] D.S. Ross, H.B. Burch, D.S. Cooper, M.C. Greenlee, P. Laurberg, A.L. Maia, S. A. Rivkees, M. Samuels, J.A. Sosa, M.N. Stan, M.A. Walter, 2016 American thyroid association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis, Thyroid.: Off. J. Am. Thyroid. Assoc. 26 (10) (2016) 1343–1421, https://doi.org/10.1089/thy.2016.0229.
[34] L. Zou, Y.H. Choi, L. Guizzetti, D. Shu, J. Zou, G. Zou, Extending the DeLong algorithm for comparing areas under correlated receiver operating characteristic curves with missing data, 10.1002/sim.10172. Statistics in medicine, Advance online publication, 2024, https://doi.org/10.1002/sim.10172.