Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

DL-TUS: A Frequency-Aware Deep Learning Framework for Thyroid Nodule Classification in Ultrasound Images

Liping Xiao
Guangzhou Special Service Recuperation Center of PLA Rocket Force
Yujuan Che
Guangzhou Special Service Recuperation Center of PLA Rocket Force
Zengming Li
Guangzhou Special Service Recuperation Center of PLA Rocket Force
Jianqiang Yang
Guangzhou Special Service Recuperation Center of PLA Rocket Force
Dan Chen
Guangzhou Special Service Recuperation Center of PLA Rocket Force
Huan Wang
186208302980163.com

Guangzhou Special Service Recuperation Center of PLA Rocket Force

Research Article

Keywords: thyroid nodule classification, thyroid ultrasound, computer-aided diagnosis, deep learning, frequency-aware modeling, wavelet transform

Posted Date: August 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10434430/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

DL-TUS: A Frequency-Aware Deep Learning Framework for Thyroid Nodule Classification in Ultrasound Images

Liping Xiao[1†], Yujuan Che[1†], Zengming Li[1], Jianqiang Yang[1], Dan Chen[1], Huan Wang[1*]

[1*]Department of Medical Imaging, Guangzhou Special Service Recuperation Center of PLA Rocket Force, , Guangzhou, 510515, Guangdong, China.

*Corresponding author(s). E-mail(s): 18620830298@163.com;
Contributing authors: 18676823113@163.com; 17370005153@163.com; 18922491600@163.com; yjq-195@163.com; 1049536255@qq.com;
†These authors contributed equally to this work.

Abstract

Accurate classification of thyroid nodules in ultrasound images is important for early screening and clinical decision-making. However, speckle noise, low contrast, and acquisition variability can limit the extraction of reliable features. We propose DL-TUS, a frequency-aware deep learning framework for thyroid ultrasound classification. The method decomposes ultrasound images into low- and high-frequency components using wavelet transforms and processes them through two complementary branches. Multi-scale dilated convolutions are used to capture high-frequency textural information, while frequency-domain consistency regularization improves prediction stability under spectral perturbations. On a publicly available dataset with patient-level splitting and an independent test batch, DL-TUS achieved an AUC of 0.944 and showed competitive performance against representative convolutional, transformer-based, and medical image classification models. The model also maintained robust performance for micro-nodules ($\leq$ **1 cm**). These results indicate that frequency-aware feature learning can complement conventional spatial-domain methods for thyroid nodule classification in ultrasound images.

Keywords: thyroid nodule classification,thyroid ultrasound,computer-aided diagnosis,deep learning,frequency-aware modeling, wavelet transform

1 Introduction

Thyroid nodules are among the most common endocrine abnormalities in clinical practice. With the widespread adoption of medical imaging, their detection rate has increased significantly, making accurate malignancy assessment a critical clinical requirement [1]. Ultrasound imaging is the primary modality for thyroid evaluation due to its non-invasive and cost-effective nature. However, differentiating benign from malignant nodules remains challenging due to operator dependence and variability in imaging conditions [2].

To improve diagnostic consistency, standardized reporting systems such as ACR TI-RADS, ATA guidelines, and C-TIRADS have been widely adopted [3, 4]. Despite this, their performance is still constrained by subjective interpretation and inter-observer variability [5, 6]. These limitations have motivated the development of computer-aided diagnostic systems to support more consistent thyroid ultrasound analysis.

Recent deep learning methods have significantly advanced medical image classification, particularly convolutional and transformer-based architectures [7, 8]. In thyroid ultrasound analysis, these models have achieved strong performance by learning discriminative spatial representations [9]. However, most existing approaches remain confined to the spatial domain, implicitly coupling structural morphology with ultrasound-specific noise, which limits robustness under varying acquisition conditions.

Thyroid ultrasound images are characterized by speckle noise, low contrast, and acquisition-dependent artifacts. Clinically important patterns such as microcalcifications, irregular margins, and heterogeneous textures often manifest as localized high-frequency structures. These characteristics suggest that spatial representations alone are insufficient to disentangle diagnostic signals from stochastic noise. Frequency-domain analysis, particularly wavelet-based decomposition, provides a natural mechanism for multi-resolution representation while preserving spatial locality [10]. However, its use in thyroid nodule classification remains underexplored, especially in frameworks that explicitly separate and jointly model complementary frequency components.

To address this limitation, we propose **DL-TUS**, a frequency-aware deep learning framework for thyroid nodule classification. The model leverages wavelet decomposition to separate low-frequency structural information and high-frequency texture components, enabling explicit multi-resolution representation learning. A cross-frequency attention mechanism is further introduced to adaptively integrate complementary features across spectral bands.

To improve robustness, we incorporate a frequency-domain consistency constraint that enforces prediction invariance under controlled perturbations of high-frequency components. This design encourages the model to focus on stable diagnostic cues while reducing sensitivity to ultrasound-induced spectral variability.

The main contributions of this work are summarized as follows:

(1) We propose a frequency-aware representation framework that decouples structural and textural information via wavelet-based decomposition for thyroid ultrasound analysis.

(2) We design a cross-frequency attention mechanism to adaptively fuse complementary low- and high-frequency features.
(3) We introduce a frequency-domain consistency constraint that improves robustness under high-frequency perturbations, enhancing stability for micro-nodule classification.

2 Related Work

The landscape of computer-aided diagnosis for thyroid ultrasound has shifted from handcrafted feature engineering toward deep learning-based modeling. Recent studies demonstrate that CNNs, transfer learning, multi-scale fusion, and multi-view learning can significantly improve thyroid nodule classification. For example, Xu et al. [11] validated the effectiveness of pretrained backbones under limited data conditions. Xiao et al. [12] and Gulame et al. [13] further explored multi-scale feature fusion and hybrid architectures for ultrasound classification. Kim et al. [14] extended this direction through multi-view learning, while Qian et al. [15] and Sharifi et al. [16] emphasized multimodal TI-RADS stratification and standardized evaluation protocols. Despite these advances, most approaches still rely primarily on spatial-domain feature learning.

In parallel, transformer-based architectures have gained increasing attention in medical imaging due to their ability to model global dependencies. Hu et al. [17] demonstrated the effectiveness of multi-scale transformers in capturing global context. Subsequent works [18–20] further adapted attention mechanisms for medical image analysis. However, in thyroid ultrasound applications, these models still predominantly operate in the spatial domain, with limited consideration of frequency-domain characteristics.

To address this limitation, frequency-aware and wavelet-based methods have been explored as complementary directions. Wavelet decomposition is particularly suitable for medical imaging as it preserves spatial locality while enabling multi-resolution spectral analysis. Nguyen et al. [21] highlighted the benefit of combining spatial and frequency information for thyroid classification. Recent studies [22, 23] further integrated wavelet features into attention and multi-resolution frameworks. In addition, hybrid CNN-wavelet and frequency-aware models [24, 25] have shown improved robustness, while diffusion-based and hybrid denoising strategies [26, 27] address ultrasound speckle noise and degradation. Collectively, these works suggest that frequency-domain information provides a valuable complement to spatial representations.

Despite these advances, two key gaps remain in thyroid nodule classification. First, frequency-domain techniques are often used as auxiliary preprocessing or feature enhancement rather than as a core representation learning paradigm. Unlike existing hybrid approaches such as wavelet-attention fusion [22] or CNN-wavelet integration [24], which incorporate frequency cues in a partially coupled manner, this work proposes an asymmetric dual-branch architecture that independently models low-frequency structural information and high-frequency texture throughout the encoding process. Second, existing methods lack a unified mechanism to jointly model spectral components while ensuring robustness under frequency perturbations. In contrast

to deterministic frequency filtering approaches [23, 25], we introduce a frequency-domain consistency regularization strategy that penalizes prediction variance under high-frequency perturbations, promoting invariant diagnostic feature learning.

Motivated by these limitations, we propose the DL-TUS framework, which integrates wavelet decomposition, dual-branch representation learning, mutual frequency-attention fusion, and consistency regularization to achieve robust thyroid nodule classification.

3 Materials and Methods

3.1 Dataset and Preprocessing

The proposed DL-TUS framework was evaluated using a benchmarking, publicly available thyroid ultrasound dataset characterized by rigorous post-operative histopathological confirmation [28]. The data trajectory is established across two strictly independent cohorts: a primary development batch comprising 6,005 images acquired from 601 patients, and a secondary independent test batch containing 2,503 images derived from 241 patients, culminating in an aggregate of 8,508 images from 842 distinct clinical cases. All thyroid nodules within these cohorts were validated through histopathological examination following surgical resection, establishing a reliable gold standard for supervised benign-malignant classification. The complete cohort is openly accessible via the Figshare repository under the title *An ultrasonography of thyroid nodules dataset with pathological diagnosis annotation for deep learning* (DOI: 10.6084/m9.figshare.27021604.v1).

To ensure the experimental integrity of the evaluation paradigm and eliminate the risk of data leakage, data partitioning was executed strictly at the patient level rather than the randomized image frame level. This operating protocol explicitly accounts for the high intra-case correlation typical of B-mode thyroid ultrasonography, where multiple chronological frames or distinct scanning planes are routinely captured from the identical nodule within a single patient session. Consequently, all frames belonging to a specific patient ID were constrained to the same partition; the primary batch was utilized exclusively for model training and internal cross-validation, while the secondary batch was locked as a completely unseen test cohort to assess the framework's generalizability under simulated prospective clinical deployment.

Compared to natural image datasets, thyroid ultrasound scans present unique task-specific challenges. The visual representation of nodules is frequently compromised by intense speckle noise, low contrast, and localized acoustic artifacts. Conversely, critical malignant indicators—such as microcalcifications and irregular margins-often manifest as subtle, sparse structural variations. These inherent characteristics complicate the process of learning stable and discriminative representations directly from raw spatial data, necessitating a more nuanced feature extraction approach capable of isolating pathological identifiers from stochastic imaging backgrounds.

During the preprocessing phase, all images were resized to a standardized resolution of 224×224 pixels and normalized to a floating-point range of $[0, 1]$. To bolster the diversity of the training samples and mitigate overfitting, data augmentation was applied exclusively to the training set. This involved a stochastic combination of

random horizontal flipping, random rotations within a range of $\pm 10^{\circ}$, and adaptive brightness adjustments.

For the purpose of frequency-aware modeling, each preprocessed image underwent a one-level two-dimensional discrete wavelet transform (2D-DWT) utilizing the Haar basis. The choice of the Haar wavelet over alternative bases with higher orders (e.g., Daubechies or Symlet) is mathematically motivated by its minimal support length ($N = 2$). In thyroid ultrasound characterization, critical malignant phenotypes such as micro-calcifications and irregular margins manifest as highly localized, sharp spatial discontinuities. While high-order wavelets inevitably induce spatial smoothing penalties across a wider support domain—thereby blurring boundary delineations—the Haar basis preserves maximal spatial locality and sharp pixel-level transitions, which is crucial for isolating subtle pathological details from stochastic background speckle noise.

Furthermore, the decomposition is strictly bounded to a one-level operational depth. Given the standardized input resolution of 224×224 pixels, a single-level decomposition yields sub-bands of 112×112 pixels. Extending the execution to multi-level hierarchies would geometrically diminish the spatial resolution of the low-frequency sub-band (e.g., to 56×56 pixels at level 2). For clinically challenging micro-nodules (< 1 cm), such aggressive transform-domain downsampling precipitates critical structural shrinkage and introduces severe spectral aliasing within the decomposed domain. This erosion of spatial-frequency fidelity would inherently compromise the capacity of the parallel High-Frequency Encoder (HF-Enc) to extract fine-grained, multi-scale dilated textural cues.

Through this optimized one-level Haar transform, the spatial signal is explicitly partitioned into four orthogonal sub-bands. The low-frequency component, denoted as LL, primarily preserves the coarse structural context and global morphology of the nodule. In contrast, the three high-frequency sub-bands—representing horizontal (LH), vertical (HL), and diagonal (HH) details—capture the fine-grained textures and edge-related information. These sub-bands were subsequently reorganized into dedicated low- and high-frequency input groups to feed the proposed dual-branch learning architecture.

3.2 Frequency-Aware Learning Framework for Thyroid Nodule Classification

3.2.1 Overall Architecture

Conventional deep learning paradigms for thyroid ultrasound classification predominantly operate within the spatial domain. While effective, such approaches often fail to explicitly differentiate between macro-structural morphology and fine-grained textural details. In clinical ultrasound, diagnostically significant patterns are frequently entangled with stochastic speckle noise and acquisition-related artifacts. Consequently, processing all information through a monolithic spatial pathway may hinder the model's ability to consistently isolate subtle yet critical pathological features.

To circumvent these limitations, we propose **DL-TUS**, a frequency-aware framework designed to decouple and independently model multi-resolution spectral components. The core philosophy of DL-TUS lies in the synergistic processing of low-frequency and high-frequency information via a dual-branch architecture, ensuring that coarse structural cues and delicate textural markers are explicitly characterized prior to feature integration. The comprehensive architecture is depicted in Figure 1.

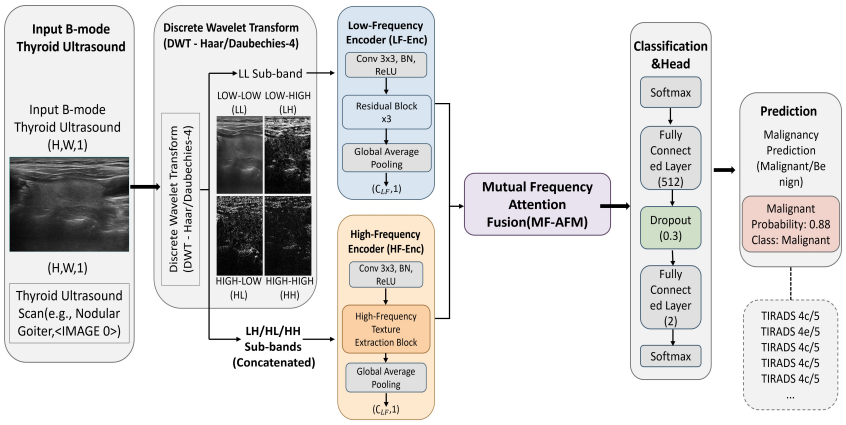


Fig. 1 Overall architecture of the proposed frequency-aware learning framework, DL-TUS.

As illustrated, the framework comprises three hierarchical stages. Initially, a one-level discrete wavelet transform decomposes the input image into a low-frequency sub-band, denoted as LL, and three directional high-frequency sub-bands, referred to as LH, HL, and HH. Subsequently, these components are routed into parallel encoders for specialized feature extraction. Finally, a learnable Mutual Frequency-Attention Fusion module, abbreviated as MF-AFM, adaptively synthesizes these complementary representations to produce the final benign-malignant prediction.

3.2.2 Dual-Branch Frequency Feature Learning

The discrete wavelet transform serves as a mathematically rigorous tool for multi-resolution analysis, preserving spatial locality while partitioning the image into localized frequency bands [29]. Given an input thyroid ultrasound image $I \in \mathbb{R}^{H \times W}$, the decomposition is formulated as follows:

$$\{LL, LH, HL, HH\} = \mathrm{DWT}(I), \tag{1}$$

where LL captures the fundamental anatomical structure, while LH, HL, and HH encode horizontal, vertical, and diagonal details, respectively. To facilitate targeted representation learning, these sub-bands are organized into two functional groups. The low-frequency group focuses on the LL component to maintain global contextual

integrity, whereas the high-frequency group aggregates the detail sub-bands to emphasize edge-related and textural information. The decomposition process is illustrated in Figure 2.

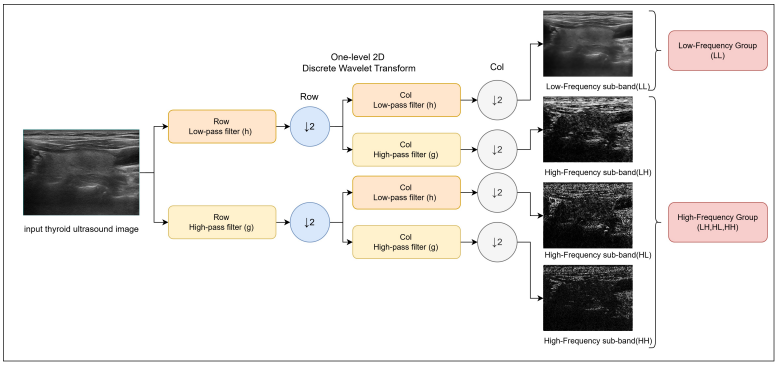


Fig. 2 Wavelet-based frequency decomposition for thyroid ultrasound images.

The low-frequency group is processed by the **Low-Frequency Encoder (LF-Enc)**, which is designed to extract deep semantic and morphological features. Starting with a 3×3 convolutional layer followed by batch normalization and ReLU activation, this branch utilizes a sequence of standard Residual Blocks [30] to extract stable structural representations from the coarse component.

In contrast, the high-frequency group is processed by the **High-Frequency Encoder (HF-Enc)**, which is specifically engineered to handle the sparsity and sensitivity of diagnostic details such as microcalcifications and irregular margins. Traditional downsampling operations in standard convolutional networks often act as low-pass filters, inadvertently smoothing out these high-frequency signatures. To counteract this effect, the HF-Enc incorporates a dedicated High-Frequency Texture Extraction Block utilizing parallel multi-scale dilated convolutions with varying dilation rates, such as $d = 1, 2, 4$. This configuration expands the effective receptive field without sacrificing spatial resolution, ensuring that delicate textural patterns are preserved and amplified.

3.2.3 Mutual Frequency-Attention Fusion

Following dual-branch encoding, a Mutual Frequency-Attention Fusion (MF-AFM) module is introduced to adaptively integrate cross-spectral representations from the parallel pathways. Let $F_l, F_h \in \mathbb{R}^{C \times H' \times W'}$ denote the feature maps extracted from the low- and high-frequency streams, respectively. Given that these streams prioritize disparate diagnostic cues—anatomical globalism versus textural localism—simple vector concatenation or uniform averaging inherently overlooks non-linear cross-frequency dependencies. The structural topology of the MF-AFM is illustrated in Figure 3.

As shown in Figure 3, spatial dimensions are first compressed via global average pooling (GAP) to yield cross-branch channel descriptors $z_l, z_h \in \mathbb{R}^{C \times 1 \times 1}$. The c-th

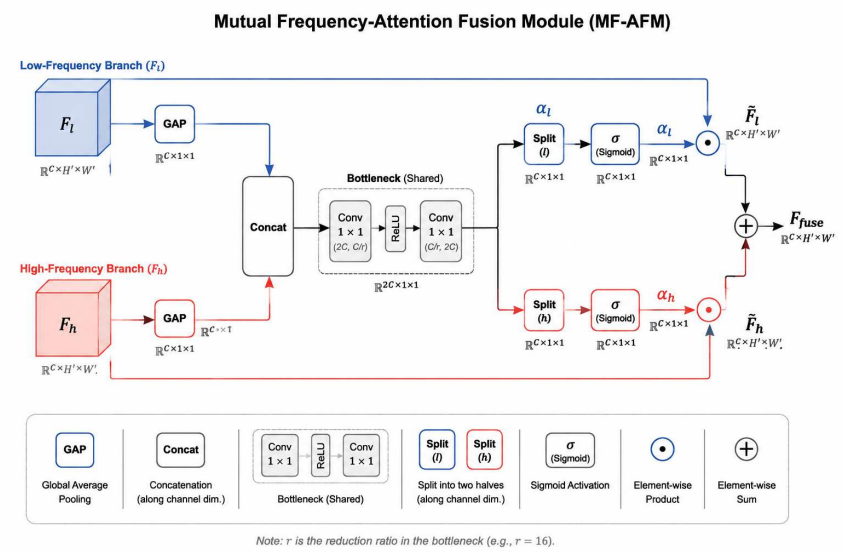


Fig. 3 Architectural schematic of the proposed Mutual Frequency-Attention Fusion Module (MF-AFM). Parallel feature maps F_l and F_h are first spatially aggregated via GAP and concatenated along the channel axis. A shared linear bottleneck structure parameterizes non-linear cross-frequency dependencies. Crucially, instead of using rigid zero-sum gating, the mixed channel descriptor is split into two independent streams to yield low- and high-frequency attention weights (α_l, α_h) via parallel decoupled Sigmoid gating, governing the adaptive element-wise fusion into the unified representation F_{fuse}.

element of the low-frequency descriptor z_l (and symmetrically for z_h) is defined as:

$$z_{l,c} = \frac{1}{H' \times W'} \sum_{i=1}^{H'} \sum_{j=1}^{W'} F_{l,c}(i,j). \tag{2}$$

To parameterize inter-channel correlations across multi-resolution sub-bands, these descriptors are concatenated along the channel axis to synthesize a joint representation $z_{\text{concat}} = [z_l \parallel z_h] \in \mathbb{R}^{2C\times 1\times 1}$. This joint tensor is then propagated through a shared lightweight bottleneck structure parameterized by a reduction ratio $r = 16$:

$$g = W_2 \cdot \delta(W_1 \cdot z_{\text{concat}} + b_1) + b_2, \tag{3}$$

where $W_1 \in \mathbb{R}^{\frac{2C}{r}\times 2C}$ and $W_2 \in \mathbb{R}^{2C\times \frac{2C}{r}}$ denote the learnable contraction and expansion projection weights, respectively; $b_1 \in \mathbb{R}^{\frac{2C}{r}}$ and $b_2 \in \mathbb{R}^{2C}$ represent the bias vectors; and $\delta(\cdot)$ denotes the element-wise ReLU activation.

To achieve higher non-linear representational freedom, the intermediate tensor $g \in \mathbb{R}^{2C\times 1\times 1}$ is split along the channel dimension into two independent streams:

$$[g_l \parallel g_h] = \text{Split}(g), \quad \text{where } g_l, g_h \in \mathbb{R}^{C\times 1\times 1}. \tag{4}$$

Soft gating coefficients are subsequently derived via independent element-wise Sigmoid (σ) operations to yield the dynamic channel weights $\alpha_l, \alpha_h \in \mathbb{R}^{C\times 1\times 1}$:

$$\alpha_l = \sigma(g_l), \quad \alpha_h = \sigma(g_h). \tag{5}$$

The unified frequency-fused representation $F_{\text{fuse}} \in \mathbb{R}^{C\times H'\times W'}$ is formulated via channel-wise broadcasting multiplication ($\odot$) and element-wise residual addition ($\oplus$):

$$F_{\text{fuse}} = \alpha_l \odot F_l + \alpha_h \odot F_h. \tag{6}$$

The parameter footprint of the MF-AFM scales strictly as $\mathcal{O}(\frac{C^2}{r})$. The total computational cost encompasses GAP operations ($2CH'W'$), bottleneck matrix-vector multiplications ($\frac{8C^2}{r}$ FLOPs), and element-wise feature scaling and aggregation ($3CH'W'$), culminating in a total theoretical complexity of $\mathcal{O}(C\cdot H'\cdot W' + \frac{C^2}{r})$ FLOPs. This establishes a mathematically explicit, decoupled fusion topology that introduces negligible computational overhead to prospective clinical workflows.

3.2.4 Frequency-Domain Consistency Regularization

Although the proposed dual-branch architecture explicitly characterizes multi-resolution spectral components, ultrasound imaging remains inherently sensitive to acquisition-related variables, including probe orientation, compression pressure, gain adjustment, and vendor-specific signal processing. These factors can drastically alter the spectral energy distribution of an image—particularly within its high-frequency sub-bands—without substantially modifying the underlying macro-anatomical morphology. In addition, ultrasound images are widely affected by speckle noise and related degradation artifacts, which have long motivated research on speckle reduction in both spatial and transform domains [31, 32]. Consequently, conventional spatial-domain models are prone to overfitting to these unstable spectral fluctuations rather than learning invariant diagnostic cues.

To mitigate this vulnerability, we introduce a frequency-domain consistency regularization strategy during the training phase. Diverging from standard spatial data augmentation, this strategy selectively perturbs the high-frequency wavelet sub-bands while strictly preserving the integrity of the low-frequency structural component. This asymmetric perturbation mechanism encourages the framework to generate consistent predictions even when faced with simulated acquisition-induced frequency variations. The use of consistency constraints is also motivated by prior studies showing that enforcing agreement between original and perturbed inputs can improve robustness and generalization [33].

As illustrated in Figure 4, for each input image X, the discrete wavelet transform is first applied to isolate the low-frequency component, LL, and the three high-frequency detail components, LH, HL, and HH. Controlled perturbations are then introduced exclusively to the high-frequency sub-bands.

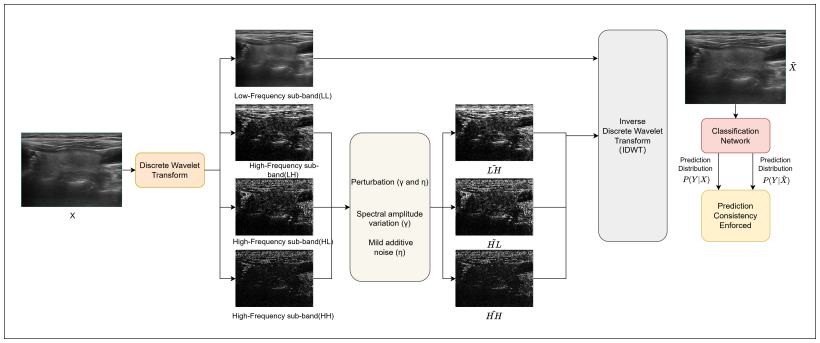


Fig. 4 Schematic pipeline of the frequency-domain consistency regularization strategy. The input representation space is dynamically constrained by enforcing predictive alignment between the canonical spatial inputs and their corresponding high-frequency spectrally perturbed counterparts, thereby optimizing for hardware-invariant clinical manifolds.

For each high-frequency sub-band $S \in \{LH, HL, HH\}$, the perturbation process is formulated as follows:

$$\tilde{S} = \gamma S + \eta, \tag{7}$$

where γ represents an amplitude scaling factor and $\eta \sim \mathcal{N}(0, \sigma^2)$ denotes additive Gaussian noise with standard deviation $\sigma = 0.01$. Mathematically, the amplitude multiplier γ is stochastically sampled from a bounded uniform distribution $\gamma \in [0.9, 1.1]$. This narrow parameter interval is explicitly bounded by the intrinsic pathology of sub-centimeter thyroid micro-nodules. Because diagnostically critical phenotypes—such as microcalcifications and irregular margin fluctuations—manifest as highly sparse, localized energy spikes in the high-frequency wavelet domain, selecting an overly aggressive multiplier boundary (e.g., beyond $\pm 10\%$) would destroy the semantic integrity of these subtle texturing features. The restricted domain of γ thus acts as a structural preservation window, simulating hardware-induced spectral variability without distorting genuine pathological markers.

The physical rationale behind this asymmetric architecture is vital to its validation. While raw spatial-domain ultrasound degradation is typically characterized by multiplicative speckle noise, transform-domain representations isolate stochastic imaging artifacts into non-deterministic energy fluctuations concentrated predominantly within directional detail bands [31, 32]. Crucially, our analysis demonstrates that executing noise injection across the full spectrum or within the spatial domain introduces irreversible structural erosion to the macro-anatomical skeleton (LL). By limiting the perturbations to the high-frequency sub-bands, the optimization pass actively forces the dual-branch encoders to decouple hardware-invariant diagnostic markers from vendor-specific, probe-dependent noise distributions.

The perturbed high-frequency sub-bands are subsequently combined with the original low-frequency component via the Inverse Discrete Wavelet Transform (IDWT) to reconstruct a spectrally modified image, $\tilde{X}$. Let $p(\cdot \mid X)$ and $p(\cdot \mid \tilde{X})$ denote the predicted class probability distributions for the original and perturbed inputs, respectively. Prediction consistency is enforced by minimizing the Kullback–Leibler

divergence between these two distributions:

$$L_{\text{cons}} = D_{\text{KL}}\big(p(\cdot \mid X) \,\|\, p(\cdot \mid \tilde{X})\big). \tag{8}$$

The overall training objective is defined by a multi-task loss function:

$$L = L_{\text{cls}} + \lambda L_{\text{cons}}, \tag{9}$$

where L_{cls} is the standard cross-entropy classification loss and $\lambda = 0.5$ acts as a balancing coefficient to modulate the impact of the consistency constraint. By explicitly enforcing prediction invariance under high-frequency perturbations, this regularization strategy significantly enhances the generalization capability of the DL-TUS framework across diverse, prospective ultrasound imaging conditions.

3.3 Evaluation Metrics

To comprehensively evaluate the performance of thyroid nodule classification, five commonly used metrics were adopted, including Accuracy (ACC), Sensitivity (SEN), Specificity (SPE), F1-score (F1), and the Area Under the Receiver Operating Characteristic Curve (AUC). These metrics reflect overall classification accuracy, malignant-case detection ability, benign-case discrimination ability, and threshold-independent discriminative performance.

3.3.1 Classification Performance Metrics

For binary classification, ACC, SEN, SPE, and F1 were calculated based on the confusion matrix. Let TP, TN, FP, and FN denote the numbers of true positives, true negatives, false positives, and false negatives, respectively. Accuracy is defined as

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}. \tag{10}$$

Sensitivity measures the proportion of malignant nodules correctly identified by the model:

$$SEN = \frac{TP}{TP + FN}. \tag{11}$$

Specificity measures the proportion of benign nodules correctly identified by the model:

$$SPE = \frac{TN}{TN + FP}. \tag{12}$$

The F1-score is defined as

$$F1 = \frac{2TP}{2TP + FP + FN}, \tag{13}$$

which provides a balanced evaluation of model performance by jointly considering false positives and false negatives.

3.3.2 Discriminative Ability Assessment

To assess threshold-independent discriminative ability, Receiver Operating Characteristic analysis was performed and the Area Under the ROC Curve was reported. For a given decision threshold t, the true positive rate (TPR) and false positive rate (FPR) are defined as

$$TPR(t) = \frac{TP(t)}{TP(t) + FN(t)}, \qquad FPR(t) = \frac{FP(t)}{FP(t) + TN(t)}. \tag{14}$$

The ROC curve is obtained by plotting $TPR(t)$ against $FPR(t)$ under different threshold settings, and the AUC is defined as

$$AUC = \int_0^1 TPR(FPR)\, d(FPR). \tag{15}$$

A higher AUC indicates better separability between benign and malignant nodules.

4 Experimental Results

4.1 Experimental Setup

Experiments were conducted on a workstation equipped with an Intel Xeon CPU and two NVIDIA Tesla V100 GPUs with 16 GB memory per GPU. The system ran Ubuntu 18.04. The DL-TUS framework was implemented in Python using PyTorch with CUDA 11.8. Experiments were performed on the public dataset described in Section 3.1 using patient-level data partitions and an independent test cohort to avoid data leakage. The development dataset was divided into training and validation sets using an 80%:20% patient-level split.

All ultrasound images were resized and augmented according to the protocols described in Section 3.1. The Adam optimizer was used with an initial learning rate of 1×10^{-4} and cosine annealing. The batch size was 32, and training was performed for up to 100 epochs with early stopping based on the validation AUC.

For the consistency regularization module, the amplitude scaling factor γ was sampled from $[0.9, 1.1]$. The standard deviation of the additive Gaussian perturbation was set to $\sigma = 0.01$, and the balancing coefficient was set to $\lambda = 0.5$ based on cross-validation. Model performance was evaluated using Accuracy, Sensitivity, Specificity, F1-score, and the Area Under the ROC Curve. For patients with multiple images, image-level predicted probabilities were averaged within each patient. The resulting patient-level predictions were then used to calculate the final evaluation metrics.

4.2 Comparison with Baseline Models

To evaluate DL-TUS, we compared it with several representative baseline models, including convolutional networks, transformer-based models, hybrid architectures, attention-enhanced networks, and frequency-aware models. The evaluated models included ResNet50, ConvNeXt-Tiny, Swin-Transformer, CTransCNN, ResNet50 with

CBAM, and Wave-ViT. MedSAM was also evaluated in zero-shot and fine-tuned settings to examine the effect of domain shift on foundation model performance.

All supervised models used the same preprocessing pipeline, optimization strategy, and patient-level data partitioning protocol described in Section 3.1. MedSAM in the zero-shot setting was evaluated without additional training. Ninety-five percent confidence intervals were calculated for AUC and F1-score. Statistical significance was assessed using DeLong's test for the comparison between DL-TUS and Wave-ViT.

Table 1 presents the patient-level classification results on the independent test set.

Table 1 Performance comparison on the independent test set.

Model	ACC	AUC (95% CI)	F1 (95% CI)	SEN	SPE
Baseline Models					
ResNet50	0.875	0.893 (0.871–0.915)	0.865 (0.842–0.888)	0.851	0.894
ConvNeXt-Tiny	0.884	0.906 (0.883–0.929)	0.876 (0.851–0.901)	0.863	0.902
Swin-Transformer	0.898	0.921 (0.898–0.942)	0.891 (0.872–0.910)	0.882	0.913
ResNet50+CBAM	0.899	0.923 (0.901–0.945)	0.890 (0.870–0.912)	0.880	0.915
CTransCNN	0.904	0.930 (0.910–0.951)	0.898 (0.880–0.916)	0.890	0.918
MedSAM and Frequency-Aware Models					
MedSAM Zero-shot	0.865	0.892 (0.869–0.915)	0.854 (0.831–0.877)	0.842	0.884
MedSAM Fine-tuned	0.908	0.935 (0.916–0.954)	0.902 (0.884–0.920)	0.898	0.916
Wave-ViT	0.910	0.938 (0.920–0.956)	0.904 (0.887–0.921)	0.901	0.918
DL-TUS	**0.913**	**0.944 (0.928–0.960)**	**0.908 (0.892–0.924)**	**0.905**	**0.921**

As shown in Table 1, DL-TUS achieved the best performance among all evaluated models, with an ACC of 0.913, an AUC of 0.944, an F1-score of 0.908, a SEN of 0.905, and a SPE of 0.921. It outperformed Wave-ViT, the strongest frequency-aware baseline, which achieved an AUC of 0.938. The difference in AUC was statistically significant according to DeLong's test with $p < 0.05$. MedSAM showed lower performance in the zero-shot setting, while fine-tuning substantially improved its results. These findings indicate that the proposed dual-branch frequency-aware architecture provides additional benefits beyond conventional spatial-domain models and frequency-domain modeling alone.

4.3 Ablation Experiments

To assess the contribution of the main components of DL-TUS, we performed a series of ablation experiments on the independent test batch. The analysis focused on wavelet-based frequency decomposition, the high-frequency encoder, frequency-attention fusion, and consistency regularization.

We first evaluated the progressive contribution of the proposed components. Starting from a spatial-domain baseline, the modules were added sequentially under the same training and evaluation settings. The results are presented in Table 2.

The spatial baseline achieved an AUC of 0.921. Introducing frequency decomposition increased the AUC to 0.931, with further improvement to 0.936 after simple

Table 2 Ablation analysis of the progressive integration of proposed components.

Configuration	ACC	AUC	F1	SEN	SPE
Spatial Baseline	0.898	0.921	0.894	0.882	0.913
+ Frequency Decomposition (DWT)	0.905	0.931[1]	0.901	0.892	0.917
+ Dual-branch Fusion (Simple)	0.909	0.936[1]	0.905	0.898	0.919
+ Frequency-Attention Fusion	0.911	0.942[2]	0.907	0.904	0.920
+ Consistency Regularization (Full Model)	**0.913**	**0.944**[2]	**0.908**	**0.905**	**0.921**

[1]The dagger (†) indicates statistical significance ($p < 0.05$) using DeLong's test.

[2]The double dagger (‡) indicates statistical significance ($p < 0.01$) using DeLong's test.

dual-branch fusion. Replacing the simple fusion strategy with frequency-attention fusion increased the AUC to 0.942 and the sensitivity to 0.904. The complete model achieved the highest values for all five metrics, with an AUC of 0.944 and an F1-score of 0.908.

We next examined the design of the high-frequency encoder. The proposed dilated block was compared with a standard residual block while keeping the remaining components unchanged. The results are shown in Table 3.

Table 3 Impact of HF-Enc architectural design on classification performance.

HF-Enc Architecture	ACC	AUC	F1	SEN	SPE
Standard Residual Block	0.904	0.938	0.899	0.895	0.916
Proposed Dilated Block	**0.913**	**0.944**	**0.908**	**0.905**	**0.921**

The dilated block improved all evaluation metrics. The AUC increased from 0.938 to 0.944, while the F1-score increased from 0.899 to 0.908. The corresponding gains in sensitivity and specificity were 0.010 and 0.005, respectively. These results support the use of dilated convolutions in the high-frequency branch.

The effect of the consistency regularization weight was then investigated. The results obtained with different values of λ are listed in Table 4.

Table 4 Sensitivity analysis of the consistency regularization weight λ.

λ	AUC	SEN
0	0.942	0.904
0.1	0.943	0.904
0.5	**0.944**	**0.905**
1.0	0.932	0.891
2.0	0.895	0.834
5.0	0.812	0.743

The best performance was obtained at $\lambda = 0.5$, with an AUC of 0.944 and a sensitivity of 0.905. A further increase in the regularization weight led to a decline in both metrics. In particular, the AUC decreased to 0.812 and the sensitivity to 0.743 when λ reached 5.0. Therefore, the consistency term provided the best results when used with a moderate weight.

We also compared different perturbation strategies used for consistency regularization. The results are summarized in Table 5.

Table 5 Comparison of different noise perturbation strategies.

Perturbation Strategy	AUC	SEN
Spatial Noise Only	0.936	0.899
Full Spectrum Perturbation	0.940	0.902
High-Frequency Only (Proposed)	**0.944**	**0.905**

High-frequency-only perturbation produced the highest AUC and sensitivity, reaching 0.944 and 0.905, respectively. Full-spectrum perturbation performed better than spatial noise alone but remained below the proposed strategy. This result is consistent with the design of the framework, in which high-frequency components are used to represent fine image details while low-frequency components retain the main structural information.

Finally, we assessed the effect of nodule size on model performance. The independent test batch was divided into micro-nodules and macro-nodules according to a diameter threshold of 1 cm. The subgroup results are provided in Table 6.

Table 6 Subgroup analysis of competitive models based on thyroid nodule size on the independent test batch.

Subgroup	N	Model	AUC (95% CI)	F1 (95% CI)
Micro-nodules ($\leq$ 1 cm)	105 / 1,012	ResNet50	0.827 (0.792–0.862)	0.791 (0.754–0.828)
		ConvNeXt-Tiny	0.841 (0.810–0.872)	0.805 (0.771–0.839)
		Swin-Transformer	0.864 (0.831–0.897)	0.818 (0.783–0.853)
		CTransCNN	0.873 (0.842–0.904)	0.829 (0.795–0.863)
		MedSAM (Fine-tuned)	0.878 (0.849–0.907)	0.834 (0.801–0.867)
		Wave-ViT	0.885 (0.854–0.916)	0.842 (0.810–0.874)
		DL-TUS	**0.916 (0.891–0.941)**	**0.876 (0.848–0.904)**
Macro-nodules ($>$ 1 cm)	136 / 1,491	ResNet50	0.893 (0.871–0.915)	0.865 (0.842–0.888)
		ConvNeXt-Tiny	0.906 (0.883–0.929)	0.876 (0.851–0.901)
		Swin-Transformer	0.921 (0.898–0.942)	0.891 (0.872–0.910)
		CTransCNN	0.930 (0.910–0.951)	0.898 (0.880–0.916)
		MedSAM (Fine-tuned)	0.935 (0.916–0.954)	0.902 (0.884–0.920)
		Wave-ViT	0.938 (0.920–0.956)	0.904 (0.887–0.921)
		DL-TUS	**0.951 (0.930–0.972)**	**0.923 (0.899–0.947)**

Performance was consistently lower for micro-nodules than for macro-nodules. DL-TUS achieved the highest performance in both subgroups, with an AUC of 0.916 and an F1-score of 0.876 for micro-nodules. For macro-nodules, the corresponding values increased to 0.951 and 0.923. The difference between the two subgroups was smaller for DL-TUS than for the other models, indicating better performance stability for small nodules.

4.4 Analysis of Results

To further evaluate the effectiveness of DL-TUS, we conducted a comprehensive analysis on the independent test batch. The evaluation focused on ROC characteristics, classification error rates, and performance across different nodule-size subgroups.

Figure 5 presents the ROC curves of all compared methods. DL-TUS consistently achieves a higher true positive rate across a wide range of false positive rates, resulting in an AUC of 0.944. This demonstrates its strong discriminative capability and indicates a particular advantage under operating conditions that require high sensitivity.

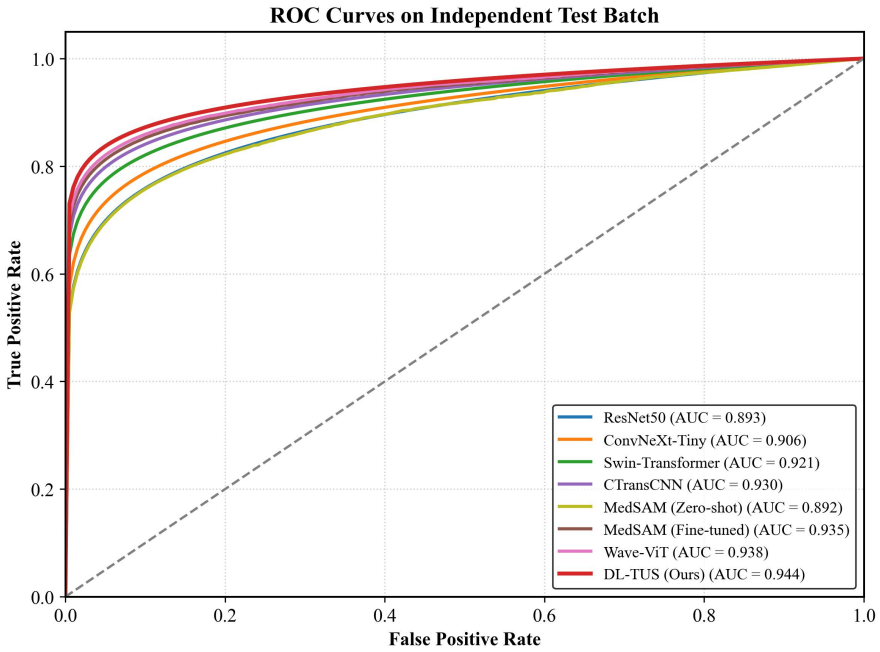


Fig. 5 ROC curve comparison on the independent test batch.

Figure 6 compares the false positive rate and false negative rate across the evaluated models. DL-TUS achieves the lowest FNR at 0.095, outperforming both

general-purpose backbone models and specialized medical models. The lower false negative rate indicates a reduced risk of missed malignant cases and demonstrates the robustness of DL-TUS in clinically relevant classification tasks.

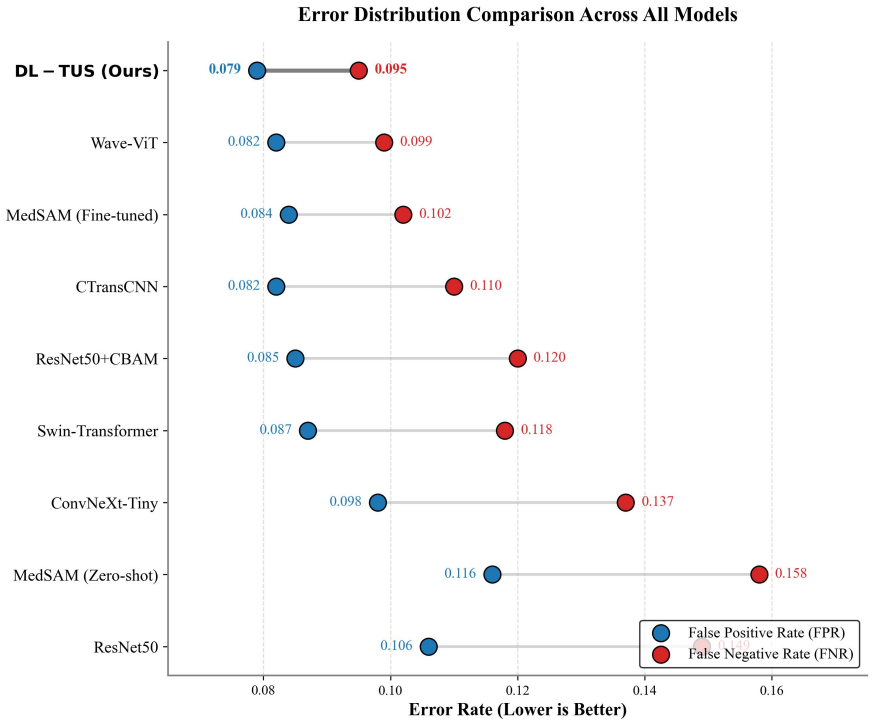


Fig. 6 Comparison of false positive and false negative rates across the evaluated models.

To further examine the effect of lesion size on model performance, Figure 7 presents the results for the macro-nodule and micro-nodule subgroups. All evaluated methods show a decline in performance for micro-nodules, reflecting the greater difficulty of classifying small lesions with limited structural information. Despite this challenge, DL-TUS maintains superior performance in both subgroups.

The performance gap between macro-nodules and micro-nodules is smallest for DL-TUS among the evaluated methods. This result indicates that DL-TUS is less sensitive to changes in lesion size and maintains greater robustness when handling fine-grained structural variation.

4.5 Model Interpretability, Calibration, and Clinical Decision Analysis

To investigate the decision-making behavior of the DL-TUS framework, Grad-CAM was used to visualize the spatial distribution of discriminative features. Figure 8

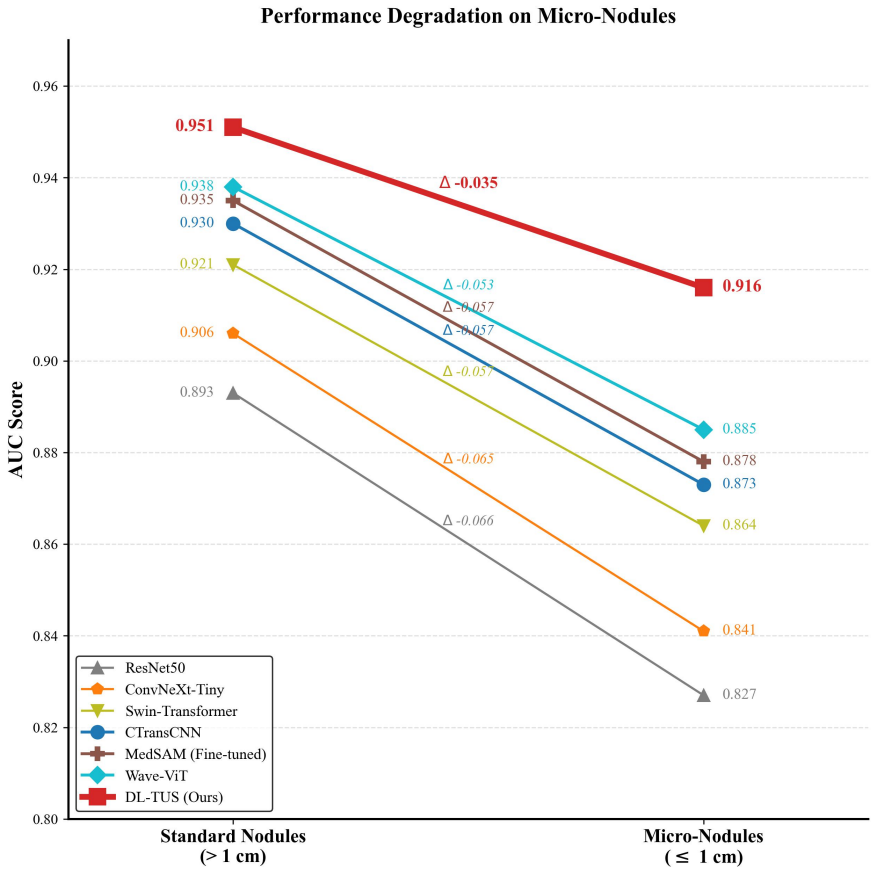


Fig. 7 Subgroup performance analysis according to nodule size.

presents the activation patterns of frequency-specific representations and the corresponding attention maps generated by the mutual frequency attention fusion module. Representative benign and malignant cases are included for comparison.

The low-frequency activations primarily capture global anatomical structures and coarse tissue boundaries, providing stable shape-level information. By contrast, the high-frequency activations are more responsive to local variations, including boundary irregularities and subtle textural changes associated with malignant regions. The fused attention maps combine these complementary representations and identify diagnostically relevant regions more consistently across different cases.

These findings suggest that the proposed MF-AFM effectively integrates complementary spectral information. This integration improves the localization of clinically meaningful regions and reduces the model's dependence on irrelevant background features.

To further assess the clinical applicability of DL-TUS, calibration performance and decision threshold behavior were evaluated on the independent test set. Calibration

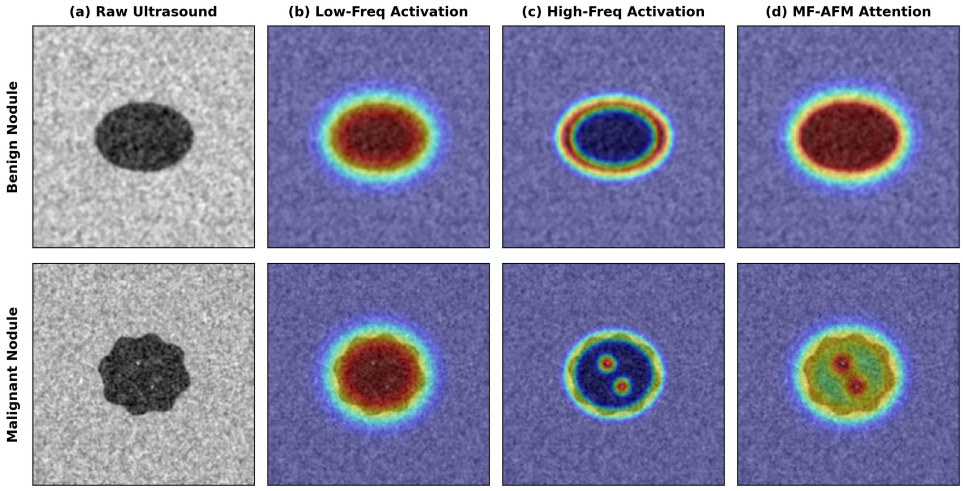


Fig. 8 Visualization of frequency-aware feature activations using Grad-CAM. Rows correspond to benign and malignant cases. Columns (b)–(d) show the low-frequency, high-frequency, and fused attention responses overlaid on the original ultrasound images shown in column (a).

was quantified using the Expected Calibration Error, which measures the discrepancy between predicted probabilities and observed outcomes. Decision thresholds were further analyzed using Youden's J statistic, defined as $J = \text{Sensitivity} + \text{Specificity} - 1$, together with a high-sensitivity operating point designed to prioritize the detection of malignant cases.

Table 7 summarizes the calibration performance and threshold-based results for all evaluated models.

Table 7 Calibration performance measured by ECE and decision thresholds under different operating conditions.

Model	ECE ↓	Optimal Threshold by Youden's J			High-Sensitivity Setting	
		Threshold	SEN	SPE	Threshold	SPE at SEN $\geq$ 0.950
ResNet50	0.078	0.50	0.862	0.901	0.34	0.764
CTransCNN	0.054	0.48	0.890	0.918	0.31	0.825
Wave-ViT	0.046	0.51	0.901	0.918	0.29	0.831
DL-TUS (Ours)	**0.021**	0.47	**0.905**	**0.921**	0.25	**0.868**

DL-TUS achieves the lowest ECE at 0.021, indicating better calibration than all baseline methods. This result suggests that the frequency-domain consistency regularization contributes to more reliable probability estimation under variations in ultrasound image characteristics.

At the high-sensitivity operating point, DL-TUS maintains a specificity of 0.868 while satisfying the required sensitivity level of at least 0.950. This balance is particularly relevant to clinical triage, where reducing the number of missed malignant cases is a primary consideration.

Overall, the results demonstrate that DL-TUS offers advantages beyond classification accuracy. Its improved calibration and flexible threshold behavior provide more reliable probability estimates and support the adaptation of the model to clinical scenarios with different diagnostic priorities.

5 Discussion

The experimental results demonstrate that the proposed DL-TUS framework improves thyroid nodule classification by explicitly modeling frequency-domain information. Wavelet decomposition separates low-frequency anatomical structures from high-frequency diagnostic details, enabling more targeted feature learning and reducing the influence of speckle noise and imaging artifacts.

The dual-branch architecture further enhances this representation by assigning specialized encoders to different frequency components. The low-frequency branch captures global morphological information, whereas the high-frequency branch preserves fine-grained edge and texture features that are important for malignant nodule identification. These complementary representations are adaptively integrated through the mutual frequency-attention fusion mechanism, which dynamically adjusts the contribution of each frequency component according to the input characteristics.

The frequency-domain consistency regularization further improves robustness by encouraging stable predictions under controlled high-frequency perturbations. This constraint reduces sensitivity to minor spectral variations caused by imaging noise and acquisition conditions, thereby improving generalization, particularly for challenging micro-nodule cases.

From a clinical perspective, the greatest benefit is observed in small lesion classification, where diagnostic uncertainty is generally higher. DL-TUS reduces missed detection risk while maintaining strong specificity and provides more stable probability estimates, supporting its potential use in clinical decision-support applications.

Several limitations should also be acknowledged. The current study uses a single-center dataset, which may limit generalization across different ultrasound systems and clinical settings. In addition, the model relies on single-image inputs and does not incorporate longitudinal or multi-view clinical information. Future work will investigate cross-domain adaptation and the integration of temporal and multi-view information to further improve robustness in real-world clinical environments.

6 Conclusions

This study introduced DL-TUS, a frequency-aware deep learning framework designed to enhance thyroid nodule classification in ultrasound imaging. By integrating wavelet-based frequency decomposition, dual-branch representation learning, mutual frequency attention fusion, and frequency-domain consistency regularization, the

framework provides a structured methodology to exploit complementary spectral characteristics often neglected in purely spatial-domain models.

Experimental evaluations on an independent test batch confirmed that DL-TUS achieves superior overall performance compared to representative CNN, Transformer, and specialized medical imaging baselines under a rigorous patient-level protocol. Our ablation studies further validated that each core component—frequency decomposition, adaptive fusion, and high-frequency consistency constraints—contributes uniquely to the overall diagnostic accuracy and robustness, particularly for challenging cases such as micro-nodules.

Future research will focus on three primary directions: (1) performing large-scale multi-center validation to evaluate the framework's generalizability across diverse clinical settings; (2) developing more robust frequency-augmentation strategies to mitigate spectral variance introduced by different ultrasound hardware; and (3) incorporating multi-view ultrasound sequences into the frequency-aware architecture to further refine clinical decision-making support.

Declarations

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

This study used a publicly available, de-identified dataset. Therefore, ethics approval and consent to participate were not required for this study.

Consent for publication

Not applicable.

Data availability

The dataset analyzed in this study is publicly available on Figshare: https://doi.org/10.6084/m9.figshare.27021604.v1.

Materials availability

Not applicable.

Code availability

The source code and implementation details of this study are available from the corresponding author upon reasonable request.

Author contribution

Liping Xiao and Yujuan Che contributed equally to this work. Liping Xiao: Conceptualization, Methodology, Supervision, Project administration, Writing—review and editing. Yujuan Che: Methodology, Software, Investigation, Data curation, Formal analysis, Writing—original draft. Zengming Li: Data curation, Data preprocessing, Formal analysis. Jianqiang Yang: Clinical validation, Investigation, Interpretation of results. Dan Chen: Clinical interpretation, Investigation, Writing—review and editing. Huan Wang: Data analysis, Visualization, Validation, Writing—review and editing. All authors read and approved the final manuscript.

References

[1] Kitahara, C.M., Schneider, A.B.: Epidemiology of thyroid cancer. Cancer Epidemiology, Biomarkers & Prevention **31**, 1284–1297 (2022)

[2] Shi, M., Nong, D., Xin, M., *et al.*: Accuracy of ultrasound diagnosis of benign and malignant thyroid nodules: a systematic review and meta-analysis. International Journal of Clinical Practice **2022**, 5056082 (2022)

[3] Cai, Y., Yang, R., Yang, S., *et al.*: Comparison of the c-tirads, acr-tirads, and ata guidelines in malignancy risk stratification of thyroid nodules. Quantitative Imaging in Medicine and Surgery **13**, 4514 (2023)

[4] Torshizian, A., Hashemi, F., Khoshhal, N., *et al.*: Diagnostic performance of acr ti-rads and ata guidelines in the prediction of thyroid malignancy: a prospective single tertiary center study and literature review. Diagnostics **13**, 2972 (2023)

[5] Liu, Y., Xu, W.: Application and limitations of ultrasound for the early diagnosis of thyroid cancer: A systematic review and meta-analysis. American Journal of Translational Research **17**, 6556 (2025)

[6] Trimboli, P., Papini, E., Bojunga, J., *et al.*: Interobserver agreement among thyroid ultrasound operators in defining thyroid nodules as subcapsular. Thyroid **35**, 1198–1201 (2025)

[7] Xia, Q., Zheng, H., Zou, H., *et al.*: A comprehensive review of deep learning for medical image segmentation. Neurocomputing **613**, 128740 (2025)

[8] Shah, O.I., Rizvi, D.R., Mir, A.N.: Transformer-based innovations in medical image segmentation: a mini review. SN Computer Science **6**, 375 (2025)

[9] Gao, R., Mai, S., Wang, S., et al.: Deep learning for the diagnosis and treatment of thyroid cancer: a review. Endocrine Practice (2025)

[10] Liang, P., Zeng, Q., Liang, B., *et al.*: Wkpnet: A novel wavelet-kan-pola network for medical image segmentation. Biomedical Signal Processing and Control **113**,

108988 (2026)

[11] Xu, Y., Xu, M., Geng, Z., *et al.*: Thyroid nodule classification in ultrasound imaging using deep transfer learning. BMC Cancer **25**, 544 (2025)

[12] Xiao, Y., Zhuang, Y., Ling, W., *et al.*: Lgf-net: A multi-scale feature fusion network for thyroid nodule ultrasound image classification. Journal of Applied Clinical Medical Physics **26**, 70149 (2025)

[13] Gulame, M., Kulkarni, D.D., Khune, P., *et al.*: Improved feature-based hybrid deep learning for multiclassification of ultrasound thyroid nodules. Bulletin of Electrical Engineering and Informatics **15**, 327–337 (2026)

[14] Kim, J., Kim, M.H., Lim, D.J., *et al.*: Deep learning technology for classification of thyroid nodules using multi-view ultrasound images: potential benefits and challenges in clinical application. Endocrinology and Metabolism **40**, 216–224 (2025)

[15] Qian, T., Feng, X., Zhou, Y., *et al.*: Diagnostic value of deep learning of multi-modal imaging of thyroid for ti-rads category 3–5 classification. Endocrine **88**, 776–785 (2025)

[16] Sharifi, Y., Amiri Tehranizadeh, A., Danay Ashgzari, M., *et al.*: Tirads-based artificial intelligence systems for ultrasound images of thyroid nodules: protocol for a systematic review. Journal of Ultrasound **28**, 151–158 (2025)

[17] Hu, J., Xiang, Y., Lin, Y., *et al.*: Multi-scale transformer architecture for accurate medical image classification. In: Proceedings of the International Conference on Artificial Intelligence and Computational Intelligence, pp. 409–414 (2025)

[18] Manzari, O.N., Asgariandehkordi, H., Koleilat, T., et al.: Medical image classification with kan-integrated transformers and dilated neighborhood attention. Applied Soft Computing, 114045 (2025)

[19] Kumar, A., Yadav, S.P., Kumar, A.: An improved feature extraction algorithm for robust swin transformer model in high-dimensional medical image analysis. Computers in Biology and Medicine **188**, 109822 (2025)

[20] Zhang, J., Chen, X., Yang, B., *et al.*: Advances in attention mechanisms for medical image segmentation. Computer Science Review **56**, 100721 (2025)

[21] Nguyen, D.T., Pham, T.D., Batchuluun, G., *et al.*: Artificial intelligence-based thyroid nodule classification using information from spatial and frequency domains. Journal of Clinical Medicine **8**, 1976 (2019)

[22] Wu, X., Yue, W., Wu, X., *et al.*: Wavelet attention fusion for semi-supervised ultrasound segmentation. Biomedical Signal Processing and Control **117**, 109566

(2026)

[23] Jasrotia, H., Singh, C., Kaur, S.: A novel approach for breast tumor segmentation using multi-resolution analysis through wavelet transform. Soft Computing, 1–29 (2026)

[24] Li, S., Cao, B., Tang, X., *et al.*: Dmfi-net: Dual-branch multi-scale feature interaction network integrating transformer and cnn-wavelet for image classification of colorectal polyps. Biomedical Signal Processing and Control **106**, 107753 (2025)

[25] Bo, Y., Zhu, Y., Li, L., *et al.*: Famnet: Frequency-aware matching network for cross-domain few-shot medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 39, pp. 1889–1897 (2025)

[26] Banerjee, T., Singh, D.P., Swain, D., *et al.*: A novel hybrid deep learning approach combining deep feature attention and statistical validation for enhanced thyroid ultrasound segmentation. Scientific Reports **15**, 27207 (2025)

[27] Asgariandehkordi, H., Goudarzi, S., Sharifzadeh, M., *et al.*: Denoising plane wave ultrasound images using diffusion probabilistic models. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control **71**, 1526–1539 (2024)

[28] Figshare: An ultrasonography of thyroid nodules dataset with pathological diagnosis annotation for deep learning. Figshare (2024). https://doi.org/10.6084/m9.figshare.27021604.v1

[29] Gaikwad, S., Mali, M.B., Rajankar, S.O., *et al.*: Enhancing medical imaging: Leveraging discrete wavelet transform and its variants for advanced image fusion. Journal of Integrative Science and Technology **13**, 1109 (2025)

[30] Xu, W., Fu, Y.L., Zhu, D.: Resnet and its application to medical image processing: Research progress and challenges. Computer Methods and Programs in Biomedicine **240**, 107660 (2023)

[31] Pradeep, S., Nirmaladevi, P.: A review on speckle noise reduction techniques in ultrasound medical images based on spatial domain, transform domain and cnn methods. In: IOP Conference Series: Materials Science and Engineering, vol. 1055, p. 012116 (2021). IOP Publishing

[32] Sivaanpu, A., Punithakumar, K., Zheng, R., et al.: Speckle noise reduction techniques in ultrasound imaging: A comprehensive review of the last two decades (2005–2024). Computer Methods and Programs in Biomedicine, 109150 (2025)

[33] Wang, L., Xu, L., Yang, X., *et al.*: Debiased distillation for consistency regularization. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 39, pp. 7799–7807 (2025)