International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

Advancing Thyroid Nodule Detection: A Deep Learning Driven Survey

M. Jenifer[1], K. P. Malarkodi[2]

[1]Research Scholar, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore-8, Tamil Nadu, India
Corresponding Author Email: *malarkodiphd2017[at]gmail.com*

[2]Assistant Professor, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore-8, Tamil Nadu, India

Abstract: ***Thyroid nodules are a common endocrine abnormality, and early identification is essential for accurate diagnosis and timely intervention. Ultrasound imaging remains the preferred clinical tool for thyroid nodule assessment, yet manual interpretation is highly dependent on radiologist expertise and often results in variability. Deep learning techniques, particularly convolutional and transformer-based architectures, offer enhanced capabilities for automated detection, segmentation, and classification of thyroid nodules. This review provides a structured and concise survey of recent deep learning developments in thyroid ultrasound analysis, outlining key models, datasets, strengths, and limitations. The study also highlights current trends in computer-aided diagnosis and identifies future research directions aimed at improving generalizability, interpretability, and clinical adaptability for real-world deployment.***

Keywords: Thyroid nodule detection, Ultrasound imaging, Deep learning, Convolutional neural networks, Computer-aided diagnosis, Medical image analysis.

1. Introduction

Thyroid nodules are common abnormalities in adults, and a small proportion may develop into malignancies, with women being almost three times more likely than males to get a diagnosis of thyroid cancer. Over the past thirty years, the prevalence of carcinoma of the thyroid has tripled or more in several high-income countries. Thyroid nodules are the early manifestations of thyroid cancer. Accurate screening of benign and malignant nodules is of great significance in improving the survival rate of thyroid cancer patients [1]. Thyroid nodules are discrete lesions that develop within the thyroid gland and are increasingly detected in the general population. Their prevalence has significantly risen owing to the extensive utilization of high-resolution ultrasound imaging. Although most thyroid nodules are harmless, a minority may represent dangerous tumors, necessitating precise detection and characterization clinically important [2]. Early identification of malignant nodules enables timely intervention and improves patient outcomes.

Thyroid nodules can be classified into solitary, multiple, cystic, or solid based on their composition [3]. Solid nodules carry a higher risk of malignancy, whereas cystic nodules are usually benign. Most nodules are asymptomatic and discovered incidentally. However, large or malignant nodules may cause symptoms such as neck swelling, hoarseness, difficulty swallowing, breathing discomfort, or hormonal imbalances [4]. The development of thyroid nodules is influenced by several factors, including iodine deficiency, radiation exposure, genetic predisposition, hormonal disorders, autoimmune thyroid diseases, and aging.

The prompt identification and diagnosis of thyroid nodules is crucial for decreasing the incidence and mortality of thyroid cancer. Ultrasound imaging is a widely used diagnostic tool because of its portability, affordability, and real-time imaging capability of pathological alterations in the thyroid, including margin, size, structure, and quantity [5]. Nonetheless, issues such as low image resolution, high noise interference, and artifacts pose significant challenges to the accurate delineation and diagnosis of thyroid nodules, particularly in cases where the nodules have irregular shapes and blurred boundaries.

Conventional early prediction of thyroid predominantly depends on imaging modalities including, Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and Fine-Needle Aspiration Biopsy (FNAB) [6]. Ultrasound is the most prevalent modality because to its simple traits, low cost, and immediate imaging capability. However, traditional diagnostic approaches are highly dependent on operator expertise, suffer from inter-observer variability, and face challenges due to ultrasound noise, low contrast, blurred boundaries, and overlapping visual aspects of healthy and cancerous lumps.

Artificial Intelligence (AI) has surfaced as a viable remedy to address these constraints through enabling automated evaluation of medical images. ML approach initially focused on handcrafted feature extraction followed by classification using algorithms such as support vector machines and random forests [7]. Although ML techniques improved diagnostic consistency, their performance is limited by manual feature design, poor scalability, and reduced robustness across diverse datasets.

DL, especially CNNs and Transformer-based architectures, structures, showed exceptional efficacy in thyroid nodule diagnosis, segmentation, and categorization. DL models automatically learn hierarchical feature representations directly increasing accuracy, lowering individuality, and facilitating preliminary identification [8]. As a result, DL-based computer-aided diagnosis systems offer enhanced accuracy, robustness, and clinical reliability, making them increasingly valuable tools for early thyroid nodule detection and diagnosis.

A. Process involved in Thyroid Nodules Detection

The thyroid nodule detection pipeline begins with image acquisition, where ultrasound imaging is used to capture thyroid gland images. This is followed by pre-processing, which enhances image quality by reducing noise and improving contrast, and normalization using techniques such as median filtering, histogram equalization, and anisotropic diffusion. In the segmentation stage, the nodule region is isolated from surrounding tissues using DL models such as ThyNet [9], pix2pix GAN [10], and other CNN- or Transformer-based architectures. Next, feature extraction is performed to obtain discriminative characteristics like texture, shape, and intensity patterns, either implicitly using DL such as CNNs (ResNet, VGG) or explicitly using handcrafted features combined with ML methods. Finally, in the classification stage, the extracted features are used to categorize nodules as benign or malignant using models such as CNN classifiers, ensemble models, and hybrid ML–DL frameworks, enabling accurate and automated thyroid nodule diagnosis [11].

The remainder of the paper is organized as follows: Section II presents a comprehensive survey of existing DL–based approaches for thyroid nodule detection, segmentation, and categorization. Section III provides a comparative analysis of these approaches, highlighting key architectures, datasets, and methodological trends. Section IV discusses the performance evaluation of selected DL models using quantitative metrics to identify effective strategies for accurate thyroid nodule diagnosis. Ultimately, Section V summarizes the work by encapsulating the major outcomes, outlining current limitations, and recommending avenues for future research in AI-assisted thyroid nodule detection. Fig.1 depicts the process of thyroid nodule detection.

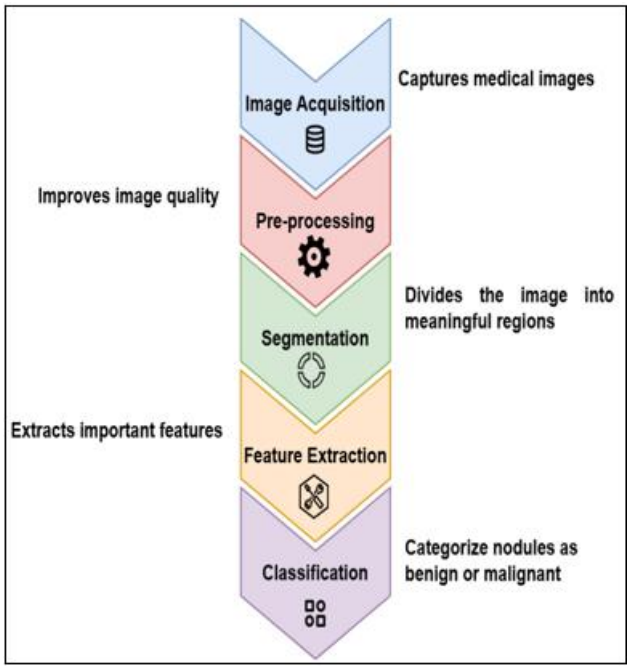

Figure 1: Process of thyroid nodule detection

2. Survey On Thyroid Nodule Detection Using Deep Learning

Cao et al. (2024) [12] developed a Tnc-Net for the automatic detection of thyroid nodule diseases with CNN. The network backbone can adapt to the problem of small data volume, capture global features with the help of simple channel attention, and effectively extract image information. Fusion of backbone and branch features, along with tailored training strategies, improves classification accuracy and clinical relevance. However, the method is limited by its dependence on 2D ultrasound images.

Swathi et al. (2024) [13] suggested QuCNet, a quantum-inspired CNN for optimized thyroid nodule classification from ultrasound images. This model integrates a quantum filter modification for enhanced feature extraction alongside a classical neural network for analysis. This framework performed dual-level analysis by first classifying nodules as benign or cancerous and then determining their corresponding worrisome (TI-RADS) categories, enabling comprehensive diagnostic characterization. However, the model is evaluated on a limited dataset, and its computational complexity.

Lee et al. (2024) [14] introduced a multihead neural network for diagnosing the signs of thyroid-associated orbitopathy utilizing multiple segmented three-dimensional CT images. This framework employed multiple convolutional embedding heads to process different segmented anatomical regions, followed by a group squeeze-and-excitation block to emphasize relevant features and a classifier for disease activity prediction. By leveraging information from multiple segmented regions, the model improved diagnostic performance compared to conventional single-image approaches. However, the method is limited to CT imaging and thyroid-associated orbitopathy cases, and its relatively low F1-score indicates challenges in class imbalance.

Yang et al. (2024) [15] suggested an improved YOLOv5-based computer-assisted evaluation framework for automatic thyroid nodule identification and categorization from ultrasound images. This model integrated a Coordinate Attention (CA) component to improve spatial and temporal feature representation and a Label Smoothing Regularization (LSR) strategy to improve robustness against noisy labels. This design enables accurate localization of lesion regions and effective differentiation between benign and malignant nodules while maintaining real-time inference speed. However, the study was conducted on a limited single-center dataset and primarily focused on detection and classification without incorporating detailed segmentation.

Liu et al. (2024) [16] developed a BFG&MSF-Net, a boundary-aware and multiple-scale integration DL framework for accurate thyroid nodules separation from ultrasound images. This model introduced a Boundary Feature Guidance Module (BFGM) to enhance edge detail extraction, addressing the common challenge of blurred and irregular nodule boundaries in ultrasound data. To capture rich contextual and scale-varying information, a Multi-Scale Perception Fusion Module (MSPFM) combines Positional Blended Attention (PBA) with Pyramid Squeeze Attention (PSA), while a Depthwise Separable Atrous Spatial Pyramid

Pooling Module (DSASPPM) further strengthens global contextual understanding. However, the study is limited by its validation being confined exclusively to ultrasound thyroid images.

Kumar et al. (2025) [17] introduced a DL–based framework for enhancing thyroid nodule assessment using ultrasound imaging. This study employed three pretrained CNN-ResNet-18, VGG-19, and AlexNet- which were fine-tuned to extract discriminative features from thyroid ultrasound images. This research work introduced a technique for thyroid nodule identification in ultrasonography (USG), employing DL to extract relevant features. However, without incorporating explicit segmentation or interpretability mechanisms, which may limit its clinical transparency and generalizability across diverse ultrasound acquisition settings.

Sharifi et al. (2025) [18] developed a DL–based Computer-Aided Diagnosis (CAD) framework for thyroid nodule risk stratification using ultrasound images in accordance with the ACR-TIRADS recommendations. This framework integrated a Faster R-CNN with a ResNet-101 backbone for accurate nodule recognition and a fine-tuned Xception model for nodule classification and risk-level stratification. This method demonstrated superior diagnostic performance compared to experienced radiologists, achieving high accuracy and agreement with the gold standard. However, evaluation on single-institution data may limit generalizability across diverse clinical settings and ultrasound devices.

Li et al. (2025) [19] suggested an Atlas-based Semantic Thyroid Nodule (ASTN) segmentation framework to tackle the limited adaptation performance of DL algorithms throughout ultrasound images acquired from different devices and imaging protocols. This method introduced a latent semantic feature co-registration strategy that focuses on lesion regions by extracting and aligning high-level semantic features between target images and an atlas composed of multiple anatomical templates. This framework improved a segmentation accuracy and robustness across heterogeneous datasets. However, atlas fusion processes, which may limit its efficiency in real-time clinical application.

Abdelrazik et al. (2025) [20] introduced a fully automated and interpretable DL approach for the separation of thyroid nodules and classifying carcinoma using ultrasound images. This framework adopts a two-stage pipeline in which a TransUNet model first performs precise nodule segmentation, and the resulting region of interest is subsequently classified using a ResNet-18 network. By constraining the classifier to focus on clinically relevant nodule regions, the method enhances interpretability and diagnostic reliability. However, the framework is evaluated on a relatively small single-center dataset, and the added segmentation stage increases computational complexity.

Al-Shahad et al. (2025) [21] suggested an enhanced pix2pix Generative Adversarial Network (GAN) framework to boost thyroid nodule segmentation via ultrasound images. The method integrates a U-Net–inspired encoder–decoder generator with a supervised loss strategy to stabilize GAN training and mitigate mode collapse. A multilayer CNN is employed as the discriminator to distinguish real and generated segmentation masks. However, the model requires carefully tuned loss functions and sufficient labeled data.

Yetginler et al. (2025) [22] introduced an enhanced V-Net–based DL simulation for automatic segmentation of thyroid nodule from ultrasound images. This framework integrated a squeeze-and-excitation (SE) mechanism within V-Net paradigm to emphasize informative feature channels while suppressing irrelevant background features. This model was evaluated on the DDTI and TN3K public datasets and illustrated superior Dice and IoU scores compared to conventional V-Net variants and recent benchmark models. However, the approach focuses solely on segmentation without incorporating malignancy classification.

Zhang et al. (2025) [23] developed an uneven encoder–decoder DL framework-based utilizing CNN as well as attention mechanisms for automatic thyroid nodule segmentation from ultrasound images. This model integrated an Efficient Convolutional Block (ECB) for extracting advanced linguistic attributes, a Convolutional Modulation Module (CMM) to improve depiction of characteristics, and a Spatial Semantic Enhancement Module (SSEM) to improve boundary detail reconstruction. In thyroid nodule segmentation tasks on ultra sound images, attaining an optimal equilibrium among inference speed and segmentation accuracy. However, the model may exhibit suboptimal performance when handling complex cases involving multiple nodules.

Zhao et al. (2025) [24] suggested a UTV-ST Swin Transformer, a multimodal DL framework that combines ultrasound video data with standardized clinical information to predict thyroid nodule invasiveness. This model used a Video Swin Transformer to extract spatiotemporal features from ultrasound videos and a KAN-based text analysis module to process clinical data, which are then fused for classification into non-invasive, central lymph node metastasis (CLNM), and central plus lateral lymph node metastasis (CLNM+LLNM) categories. However, the model's performance can still be improved, particularly in extremely complex or low-quality imaging conditions.

Lu et al. (2025) [25] developed MFSE-TransUNet, an enhanced Transformer-based segmentation structure for thyroid nodule ultrasound images which includes dynamic feature calibration and edge enhancement mechanisms. Building upon the hybrid CNN–Transformer architecture of TransUNet, the model introduces adaptive convolution to better handle morphological variability of thyroid nodules, hierarchical multi-scale feature fusion to account for unequal feature contributions. An edge enhancement strategy to preserve boundary information during up sampling. However, the use of transformer layers increases computational complexity.

Zhou et al. (2025) [26] suggested an automated and explainable ML framework for newborn screening of thyroid dysfunction using laboratory indicators rather than medical imaging. The two-phase system first applies a Random Forest (RF) classifier to distinguish healthy newborns from those with thyroid dysfunction, followed by a stacking ensemble model to further categorize affected cases into congenital

hypothyroidism or elevated TSH subtypes. SHAP-based feature selection enhances model interpretability by identifying key biomarkers such as TSH, FT4, and T4, while a large language model (LLM) is incorporated to assist in low-confidence diagnostic cases. However, it is limited to tabular screening data and does not address image-based thyroid analysis, which may restrict its applicability to imaging-driven CAD systems.

Xu et al. (2025) [27] introduced a thyroid nodule classification framework for ultrasound imaging by combining conventional ML with advanced transfer learning methodologies. This approach employed ITK-Snap for image preprocessing and feature extraction, followed by LASSO-based feature selection to reduce dimensionality. A Support Vector Machine (SVM) model and an Inception V3 transfer learning model was independently developed and subsequently fused using post-fusion strategies to enhance diagnostic performance. However, the study is limited by its retrospective, single-center design and reliance on manual preprocessing.

Zhang et al. (2025) [28] developed TN5000, a large-scale open-access ultrasound image dataset designed specifically for thyroid nodule detection and classification. TN5000 provides detailed statistical analysis of image characteristics and establishes benchmark results using baseline DL techniques for identifying and categorizing tasks. As the largest publicly available thyroid ultrasound dataset with professional labeling, TN5000 functions as an important asset for developing and fairly assessing AI-based computer-aided diagnosis systems. However, the dataset is limited to B-mode ultrasound images and does not incorporate multimodal clinical data, which may constrain its applicability to more comprehensive diagnostic frameworks.

Banerjee et al. (2025) [29] developed TATHA, a hybrid DL framework for thyroid ultrasound segmentation that integrates deep feature attention, statistical validation, and edge detection. This model enhanced a segmentation accuracy by reducing speckle noise, improving contrast, and combining predictions from multiple feature extraction units. It surpasses U-Net, PSPNet, and Vision Transformers in cross-validation tests. However, the framework is limited by a small dataset and may face generalization challenges on multi-center datasets.

Zhou et al. (2025) [30] developed TNVis, a diagnosis device for tumor of the thyroid using DL, leveraging dynamic ultrasound video and 3D visualization to enhance clinical assessment. The two-stage framework segments and visualizes nodules, improving radiologists' diagnostic performance and mimicking clinical workflows. The 3D visualization closely mimics radiologists clinical workflows, offering personalized and precise evaluation, particularly for complex nodules that may be difficult to assess on 2D images. However, the model's complexity and dependence on multi-center video data may limit deployment in resource-constrained settings.

Wang et al. (2025) [31] suggested YOLO-Thyroid, an improved YOLOv8 architecture for automatic recognition of thyroid nodules in ultrasound images. Incorporating a C2fA module with CA to enhance obtaining traits and optimized loss functions- class-weighted binary cross-entropy to address class disparities and SCYLLA-IoU (SIoU) for better boundary identification. This method was evaluated on a freely accessible thyroid ultrasound dataset. with data augmentation and format optimization. However, its performance is still moderate, and validation is limited to 2D ultrasound images.

3. Comparative Analysis

This section compares the above-mentioned models, highlighting its advantages, disadvantages, datasets employed and performance measures, as in Table 1.

Table I: Summary of recent deep learning models thyroid nodule detection

Author & Year	Techniques Used	Merits	Demerits	Dataset Used	Performance Metrics
Cao et al. (2024) [12]	CNN-Tnc-Net	Handles small dataset and captures global and local features.	Limited to 2D ultrasound images and has not been externally validated across multiple clinical centers.	Nanjiang TCM Hospital dataset. 386 training sets (including 302 benign nodules and 84 malignant nodules), 98 validation sets, and 122 test sets.	Accuracy = 90.2%, Precision = 74.2%, Sensitivity = 85.2%, Specificity = 91.6%, F1 Score = 79.3%, AUC = 0.867%
Swathi et al. (2024) [13]	CNN-QuCNet	Achieves high accuracy for both benign–malignant and TI-RADS suspicious level classification.	Validation confined to a limited dataset may affect the model's generalizability across diverse populations.	TN3K dataset of 127 images for tumor classification and 98 images for suspicious level classification	Accuracy = 95.74%, Precision = 96.04%, Recall = 95.74%, F1 Score = 95.83%
Lee et al. (2024) [14]	Squeeze-and-excitation (SE) block and classifier	Effectively leverages multiple segmented regions, improves diagnostic performance.	Limited to CT-based thyroid-associated orbitopathy images, relatively low F1-score	Chung-Ang hospital Thyroid-associated orbitopathy CT dataset from 1,068 patients. Each patient has 80 to 400 image slices.	AUC = 0.800%, Accuracy = 72.1%, F1-score = 41.6%, Sensitivity = 72.8%, Specificity = 72%
Yang et al. (2024) [15]	YOLOv5, CA, LSR	CA enhances spatial and positional feature extraction.	Focuses on detection and classification without explicit segmentation.	Siemens Medical Solutions dataset. 191thyroid ultrasound images from 171 patients.	Mean Average Precision (mAP) = 95.3%, Accuracy = 90%

Liu et al. (2024) [16]	BFG&MSF-Net, BFGM, PBA, PSA, DSASPPM, RM	Effectively captures boundary details, improves small and irregular nodule segmentation.	Validation is restricted to ultrasound thyroid images.	TN3K and DDTI. Images of thyroid nodules comprises 3,493 ultrasound images from 2,421 individuals.	For TN3K: DSC% = 86.82%, For DDTI: DSC% = 81.02%
Kumar et al. (2025) [17]	VGG-19, AlexNet, ResNet-18	Higher classification accuracy ResNet-18 shows superior performance and reduces unnecessary FNAB procedures.	The absence of explicit segmentation and interpretability limits clinical transparency.	DDTI dataset. An overall of 299 individuals were enrolled in the research, totally 427 images with 357 thyroid nodules(malignant) and 70 benign	For Accuracy: AlexNet = 83.59%, ResNet-18 = 97.13%, VGG-19 = 90.31%
Sharifi et al. (2025) [18]	CAD, Faster R-CNN, ResNet-101, ACR-TIRAD	Outperforms experienced radiologists and reduces interobserver variability.	Performance may vary across ultrasound machines with differing imaging characteristics.	PERSIAN dataset. 2450 ultrasound images with 3250 nodules from 1037 patients (single institution)	Accuracy = 98%, AUC = 0.99%, Precision = 96.7%; Recall = 91.2%, Mean Kappa = 0.85%
Li et al. (2025) [19]	ASTN segmentation framework	Robust boundary delineation and effective use of anatomical prior information.	Co-registration and atlas fusion increase inference time.	The TUI dataset, obtained via a participating hospital, consists of 11,360 images from the P1 tool and 800 images with the M3 gadget.	For ASTN DSC%: P1 = 92.57%, M3 = 76.86%
Abdelrazik et al. (2025) [20]	TransUNet, ResNet-18	Reduces inter-observer variability and outperforms traditional ML with handcrafted features.	Increased computational cost due to segmentation stage.	DDTI dataset contains 480 images from 390 patients TN3k dataset contained 3,493 images, TNUI dataset consisted of 1,381 images	F1 = 76.49%, Accuracy = 66.19%, Recall = 67.13%, Precision = 89.63%
Al-Shahad et al. (2025) [21]	Improved pix2pix GAN framework	Effective feature extraction and better flexibility than semi-supervised models.	Limited generalization across diverse ultrasound devices.	747 thyroid ultrasound images: Hospital Sultan Abdul Aziz Shah, Malaysia (302 images) + open-access thyroid nodule dataset (445 images); 545 images for training and 202 images for testing	Accuracy = 97%, F1-Score = 92%, Dice = 87.97%%
Yetginler et al. (2025) [22]	V-Net based CNN integrated with SE	Enhances important feature representation, improves segmentation accuracy on complex and blurred boundaries.	Required labeled masks, evaluated on 2D ultrasound images only.	The DDTI dataset has 406 instances for training, 127 for testing, and 102 for confirmation. The TN3K dataset consisted of 2,235 instances for training, 699 for evaluation, and 559 for verification.	For DDTI: Dice = 84.51%, IoU = 76.27% For TN3K: Dice = 83.88%, IoU = 75.50%
Zhang et al. (2025) [23]	CNN, ECB, CMM, SSEM	Improved feature representation and boundary delineation.	Generalization across multi-center datasets not evaluated.	TN3K consists of 3,493 ultrasound images from 2,421 sufferers.	DSC = 83.32%, IoU = 74.73%
Zhao et al. (2025) [24]	UTV-ST Swin Transformer, KAN	Robust to noisy data, superior performance in invasiveness prediction	Higher computational complexity and validation limited to ultrasound-based data.	Clinical thyroid ultrasound dataset, Total of 346 patients were included in the study, with 352 thyroid nodules captured in 346 ultrasound videos.	Accuracy = 82.1%, Average AUC = 94.2%
Lu et al. (2025) [25]	MFSE-TransUNet	Effectively handles morphological diversity, preserves boundary details.	Its overall performance remains inferior to that of TransUNet, mainly due to the lack of explicit spatial information modeling.	TUD, TN3K, and DDTI datasets. TUD consists of 1024 images, DDIT consists of 637 images and TN3K consists of 3493 images.	For TUD, MIoU = 84.24%, Dice = 90.90%, Recall = 89.40%, Precision = 92.40%, Accuracy = 98.30%
Zhou et al. (2025) [26]	RF, SHAP-based feature selection, stacking ensemble learning, LLM	High diagnostic accuracy and strong interpretability through SHAP.	Limited to laboratory (tabular) data and does not support image-based thyroid analysis.	Screening data from 5,406 newborns collected at Children's Hospital Zhejiang University School of Medicine.	For Stage 1: AUC = 0.9914%, Sensitivity = 98.57%, Specificity = 98.60%; For Phase 2: Macro AUC = 0.9696%, Accuracy = 93.75%
Xu et al. (2025) [27]	SVM, Inception V3 CNN with LASSO feature selection	Fusion model improves diagnostic accuracy over individual models	Retrospective single-center study and performance gain is moderate.	Zhejiang Rongjun Hospital, China. 1134 ultrasound images from 630 thyroid nodule cases (589 benign, 545 malignant).	For SVM: AUC = 0.748%, Accuracy = 72.2%, For Inception V3, AUC = 0.783%
Zhang et al. (2025) [28]	ResNeT50, ResNeT101	Biopsy-confirmed expert annotations and suitable for both detection and classification	Does not cover diverse imaging devices or multi-center variability.	TN5000 dataset: 5,000 ultrasound images of thyroid nodule, 3,572 malignant	For ResNeT50, MN = 0.849%, BN = 0.771%, mAP = 0.810% For ResNeT101, MN =

				incidents and 1,428 benign cases.	0.843%, BN = 0.754%, mAP = 0.799%
Banerjee et al. (2025) [29]	TATHA, edge detection, and statistical validation for segmentation	Improves segmentation accuracy over U-Net, PSPNet, and Vision Transformers.	Relatively small dataset may not generalize well to multi-center.	DDTI Dataset, Ultrasound Images:99 cases, 134 labeled images.	Dice = 93.5%, Accuracy = 96.2%, AUC = 0.951%, Sensitivity = 97.4%, Specificity = 95.6%
Zhou et al. (2025) [30]	TNVis	Improved radiologist performance and provides 3D visualization.	Requires dynamic ultrasound videos and untested on static 2D images.	TNVs dataset, 5,228 cases (2,369 benign and 2,859 malignant).	For Internal: Dice = 90%, For External: AUC = 0.79%
Wang et al. (2025) [31]	YOLOv8, (class-weighted BCE, SIoU)	Enhanced feature extraction and better boundary localization.	Its performance is still moderate, and validation is limited to 2D ultrasound images.	The DDTI dataset has 480 images derived from 400 clinical situations.	Precision = 54%, Recall = 58.2%

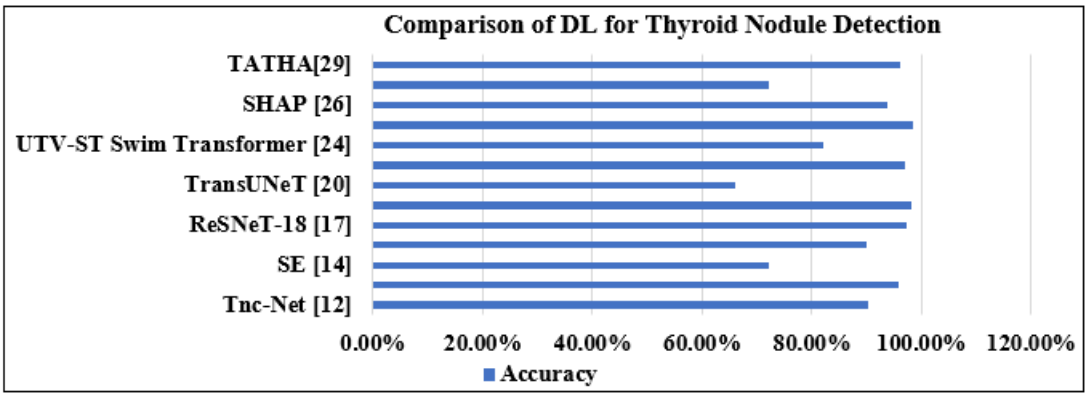

Figure 2: Comparison of DL for thyroid nodule detection

4. Performance Analysis

The performance analysis of the DL methods highlights their effectiveness in detecting, segmenting, and classifying thyroid nodules using ultrasound and other imaging modalities. Most of the studies utilized publicly available datasets such as MFSE-TransUNet, UTV-ST Swin Transformer and institution-specific datasets, reflecting a diverse range of imaging sources and clinical settings. In this section, a comparative performance evaluation is conducted across DL-based thyroid nodule detection models, focusing primarily on segmentation accuracy, classification performance, and overall diagnostic reliability.

Fig. 2 depicts the graphical representation of different DL–based thyroid nodule segmentation models evaluated using accuracy metrics. Among the compared approaches, the MFSE-TransUNet model achieves a high accuracy of, 98.30%, demonstrating superior performance over several existing methods. The strong performance of MFSE-TransUNet may be ascribed to the amalgamation of broad feature extraction, dynamic feature calibration, and edge enhancement mechanisms, which effectively preserve boundary information and capture both regional and worldwide contextual characteristics using thyroid ultrasound images.

5. Other Recommendations

Deep learning has transformed thyroid nodule diagnosis by offering improved accuracy, consistency, and support for early clinical decision-making. The models reviewed in this survey demonstrate rapid progress in segmentation, classification, and invasiveness assessment, supported by growing datasets and improved architectural designs. Despite these advances, challenges remain in model generalizability, interpretability, and multi-center robustness. Future research should prioritize lightweight and explainable models, broader dataset diversity, and multimodal integration to support reliable clinical adoption.

References

[1] M. Shi, D. Nong, M. Xin, and L. Lin, “Accuracy of ultrasound diagnosis of benign and malignant thyroid nodules: A systematic review and meta-analysis,” International Journal of Clinical Practice, vol. 2022, no. 1, pp. 5056082, 2022.
[2] J. L. Ng, L. M. Escueta, and O. A. Dampil, “Malignancy in thyroid nodules with Bethesda III category on repeat fine needle aspiration biopsy,” Journal of the ASEAN Federation of Endocrine Societies, vol. 38, no. 2, p. 86, 2023.
[3] C. Durante, G. Grani, L. Lamartina, S. Filetti, S. J. Mandel, and D. S. Cooper, “The diagnosis and management of thyroid nodules: A review,” JAMA, vol. 319, no. 9, pp. 914–924, 2018.
[4] M. Chaudhary, N. Baisakhiya, and G. Singh, “Clinicopathological and radiological study of thyroid swelling,” Indian Journal of Otolaryngology and Head & Neck Surgery, vol. 71, no. Suppl 1, pp. 893–904, 2019.
[5] M. A. Savelonas, D. K. Iakovidis, I. Legakis, and D. Maroulis, “Active contours guided by echogenicity and texture for delineation of thyroid nodules in ultrasound images,” IEEE Transactions on Information

Technology in Biomedicine, vol. 13, no. 4, pp. 519–527, 2009.
[6] A. C. Nachiappan, Z. A. Metwalli, B. S. Hailey, R. A. Patel, M. L. Ostrowski, and D. M. Wynne, "The thyroid: Review of imaging features and biopsy techniques with radiologic-pathologic correlation," Radiographics, vol. 34, no. 2, pp. 276–293, 2014.
[7] F. Bini, A. Pica, L. Azzimonti, A. Giusti, L. Ruinelli, F. Marinozzi, and P. Trimboli, "Artificial intelligence in thyroid field—A comprehensive review," Cancers, vol. 13, no. 19, p. 4740, 2021.
[8] Y. Habchi, H. Kheddar, Y. Himeur, and M. C. Ghanem, "Machine learning and transformers for thyroid carcinoma diagnosis," Journal of Visual Communication and Image Representation, vol. 115, p. 104668, 2025.
[9] S. Peng, Y. Liu, W. Lv, L. Liu, Q. Zhou, H. Yang, ... & H. Xiao, "Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study", The Lancet Digital Health, Vol. 3, no. 4, pp.250-259, 2021.
[10] A. Aljohani and N. Alharbe, "Generating synthetic images for healthcare with novel deep pix2pix GAN," Electronics, vol. 11, no. 21, p. 3470, 2022.
[11] V. V. Vadhiraj, A. Simpkin, J. O'Connell, N. Singh Ospina, S. Maraka, and D. T. O'Keeffe, "Ultrasound image classification of thyroid nodules using machine learning techniques," Medicina, vol. 57, no. 6, p. 527, 2021.
[12] J. Cao, Y. Zhu, X. Tian, and J. Wang, "Tnc-Net: Automatic classification for thyroid nodules lesions using convolutional neural network," IEEE Access, vol. 12, pp. 84567–84578, 2024.
[13] G. Swathi, A. Altalbe, and R. P. Kumar, "QuCNet: Quantum-inspired convolutional neural networks for optimized thyroid nodule classification," IEEE Access, vol. 12, pp. 27829–27842, 2024.
[14] S. Lee, J. K. Lee, and J. Lee, "Multihead neural network for multiple segmented images-based diagnosis of thyroid-associated orbitopathy activity," IEEE Access, vol. 12, pp. 43862–43873, 2024
[15] D. Yang, J. Xia, R. Li, W. Li, J. Liu, R. Wang, J. You, "BFG&MSF-Net: Boundary feature guidance and multi-scale fusion network for thyroid nodule segmentation ultrasound imaging with improved YOLOv5 neural network," IEEE Access, vol. 12, pp. 22662–22670, 2024.
[16] J. Liu, J. Mu, H. Sun, C. Dai, Z. Ji, and I. Ganchev, "BFG&MSF-Net: Boundary feature guidance and multi-scale fusion network for thyroid nodule segmentation," IEEE Access, vol. 12, pp. 78701–78713, 2024.
[17] J. Kumar, S. N. Panda, D. Dayal, and M. Sharma, "Enhancing thyroid nodule assessment with deep learning and ultrasound imaging," e-Prime—Advances in Electrical Engineering, Electronics and Energy, vol. 11, p. 100894, 2025.
[18] Y. Sharifi, M. D. Ashgzari, S. Shafiei, S. R. Zakavi, and S. Eslami, "Using deep learning for thyroid nodule risk stratification from ultrasound images," WFUMB Ultrasound Open, vol. 3, no. 1, p. 100082, 2025.
[19] X. Li, Y. Zhu, Y. Fan, X. Wei, R. Zhang, Y. Tian, and Z. Liu, "Ultrasound image segmentation of thyroid nodule via latent semantic feature co-registration," Biomedical Signal Processing and Control, vol. 108, p. 107971, 2025.
[20] O. Abdelrazik, M. Elsayed, N. Wahab, N. Rajpoot, and A. Shephard, "A deep learning framework for thyroid nodule segmentation and malignancy classification from ultrasound images," arXiv preprint, pp. arXiv:2511.11937, 2025.
[21] H. F. Al-Shahad, R. Yaakob, N. M. Sharef, H. Hamdan, and H. A. Hassan, "An improved pix2pix generative adversarial network model to enhance thyroid nodule segmentation," Journal of Advanced Information Technology, vol. 16, pp. 37–48, 2025.
[22] B. Yetginler and İ. Atacak, "An improved V-Net model for thyroid nodule segmentation," Applied Sciences, vol. 15, no. 7, pp. 3873, 2025.
[23] Z. Zhang, L. Li, C. Zhao, P. Ren, and R. Zhang, "Asymmetric network based on CNN and attention mechanisms for thyroid nodule segmentation," IEEE Access, vol. 13, pp. 141908–141919, 2025.
[24] Y. Zhao, Y. Li, Y. Zhang, X. Yan, G. Yin, and L. Liu, "Enhancing thyroid nodule assessment with UTV-ST Swin Kansformer: A multimodal approach to predict invasiveness," IEEE Access, vol. 13, pp. 29081–29090, 2025.
[25] Y. Lu, J. Jing, W. Zhang, and Y. Kong, "MFSE-TransUNet: A thyroid nodule ultrasound image segmentation network integrated with dynamic feature calibration and edge enhancement," IEEE Access, vol. 13, pp. 137209–137218, 2025.
[26] Y. Zhou, J. Zhang, C. Chen, H. Mao, and R. Yang, "Newborn screening for thyroid dysfunction through explainable SHAP feature selection and stacking ensemble with LLM assisted diagnosis," IEEE Access, vol. 13, pp. 188552–188563, 2025.
[27] Y. Xu, M. Xu, Z. Geng, J. Liu, and B. Meng, "Thyroid nodule classification in ultrasound imaging using deep transfer learning," BMC Cancer, vol. 25, no. 1, p. 544, 2025.
[28] H. Zhang, Q. Liu, X. Han, L. Niu, and W. Sun, "Tn5000: An ultrasound image dataset for thyroid nodule detection and classification," Scientific Data, vol. 12, no. 1, p. 1437, 2025.
[29] T. Banerjee, D. P. Singh, D. Swain, S. Mahajan, S. Kadry, and J. Kim, "A novel hybrid deep learning approach combining deep feature attention and statistical validation for enhanced thyroid ultrasound segmentation," Scientific Reports, vol. 15, no. 1, p. 27207, 2025.
[30] Y. Zhou, C. Chen, J. Yao, J. Yu, B. Feng, L. Sui, and D. Xu, "A deep learning- based ultrasound diagnostic tool driven by 3D visualization of thyroid nodules," npj Digital Medicine, vol. 8, no. 1, p. 126, 2025.
[31] S. Wang, Z. A. Zhao, Y. Chen, Y. J. Mao, and J. C. W. Cheung, "Enhancing thyroid nodule detection in ultrasound images: A novel YOLOv8 architecture with a C2FA module and optimized loss functions," Technologies, vol. 13, no. 1, 2025.