

ThyroUS-Net: Hybrid CNN with Squeeze-Excitation for Thyroid Nodule Malignancy Classification from Ultrasound

Fahima akter nila^{1,*}, Rafsana Ferdouse²

¹Master's of public health (Biostatistics and epidemiology) Monroe university. USA

²Master's of Public Health (New Rochelle Subject: MPH) University: king Graduate School, Monroe University. USA

*Corresponding author: sanaferdouse20@gmail.com

Received March 20, 2026; Revised April 22, 2026; Accepted April 28, 2026

Abstract The prevalence of thyroid nodules in general population is very high, but only a small percentage of the nodules become malignant; thus, a proper risk stratification is necessary to prevent unnecessary biopsies and prompt cancer diagnosis. The most common diagnostic modality used to examine the thyroid is ultrasound imaging, but it is a subjective form of interpretation that is prone to inter-observer variability despite standardized systems like ACR TI-RADS. Despite convolutional neural networks (CNNs) achieving encouraging outcomes in the classification of thyroid ultrasounds, most of the current models do not incorporate hybrid multi-scale feature extraction and effective attention in order to improve the discriminative effectiveness. The paper presents a hybrid CNN model called ThyroUS-Net to achieve superior/significantly improved classification of thyroid nodule malignancy. The model combines shallow and deep convolutional networks to extract local texture and global morphological information, and SE blocks are used to recalibrate channels channelwise to highlight the diagnostically meaningful information. The network was tested on the Thyroid Digital Image Database (that is publicly accessible: 428 ultrasound images: 357 benign, 71 malignant). Untrained data augmentation, standard preprocessing, and stratified train-validation-test splitting were used to guarantee sound evaluation. According to the results of the experiment, ThyroUS-Net is superior to the baseline models, such as simple CNN, VGG-16, or ResNet-50, with accuracy, precision, recall, and F1 score of 95, 98.4, 97.1, and AUC of 0.97, respectively. The ablation analysis proved that both the hybrid backbone and SE attention module were significant in boosting performance. Such results indicate that ThyroUS-Net has better diagnostic consistency and better malignancy discrimination, which can be used as a clinical assistance system. The future research will involve multi-center validation, TI-RADS scoring integration, and explainable AI techniques to promote clinical adoption.

Keywords: *Thyroid nodule classification, Ultrasound imaging, Hybrid CNN, Squeeze-Excitation network, Deep learning in healthcare, Computer-aided diagnosis*

Cite This Article: Fahima akter nila, and Rafsana Ferdouse, "ThyroUS-Net: Hybrid CNN with Squeeze-Excitation for Thyroid Nodule Malignancy Classification from Ultrasound." *American Journal of Public Health Research*, vol. 14, no. 2 (2026): 44-51. doi: 10.12691/ajphr-14-2-5.

1. Introduction

One of the most widespread clinical observations that are faced in endocrine practice is the prevalence of thyroid nodules where the rates of its prevalence have been found to be high when examined using modern imaging technologies. In cases where high-resolution ultrasound is examined, the studies approximate that up to 19-68% of people have thyroid nodules basing on the age, sex, and population under investigation [1]. Although this prevalence is high, epidemiological evidence shows that only 5-15% of nodules are malignant with the rest being benign [2]. Thyroid cancer is an emerging epidemic observed all over the world in recent decades and it can be attributed to a greater use of imaging modalities and

earlier identification of smaller lesions [3]. The ultrasound imaging is still the most reliable diagnostic tool to be used in assessing the thyroid nodules because it is non-invasive, real-time and highly sensitive as compared to other imaging modalities [4]. It allows morphological characteristics, including echogenicity, calcifications, margins, and shape, to be characterised in detail and this information can guide clinical decision-making and risk stratification [4]. Nonetheless, the process of ultrasound interpretation is highly skilled and accuracy in the differentiation of benign and malignant nodules is varied among different practitioners and different institutions.

Regardless of the developed standards of imaging, including the American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS), ultrasound evaluation of thyroid nodules is often characterized by inter-observer variability and diagnostic

subjectivity [5]. Although TI-RADS scale creates progress in the direction of standardization, it continues to be heavily based on subjective assessment of sonographic descriptors (composition, echogenic foci, margin irregularity, and shape), which are not consistently agreed upon by radiologists [5]. Such subjectivity has been associated with variability in risk stratification and probably excessive use of invasive practices such as fine-needle aspiration (FNA), which has additional cost of healthcare in addition to patient risk. Consequently, the demand increases on automated computer-aided diagnosis (CAD) systems that have the ability to objectively analyze ultrasound images and minimize the skills of operators.

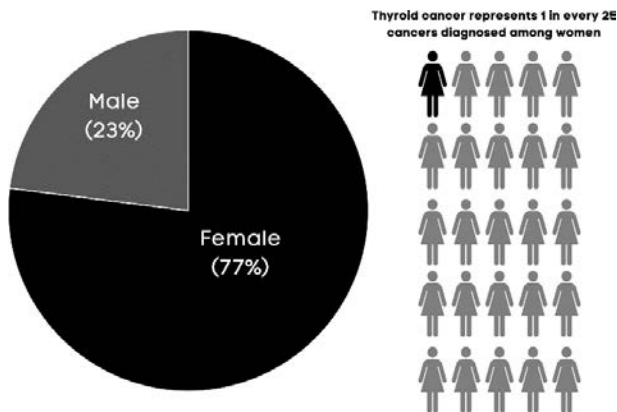


Figure 1. The global incidence of thyroid cancer [3]

Deep learning and convolutional neural networks (CNNs) have demonstrated tremendous potential in medical imaging-related classification tasks in the past ten years [6]. CNN-based models have been used on thyroid ultrasound images and encouraging, demonstrating their capacity to perform comparably or even surpass experienced radiologists in classification tasks [7]. These models directly learn visual features hierarchically and thus can detect finer details of image data and patterns which human observers might miss. However, the current CNN-based methods are limited in a number of ways. Standard deep learning models often function as 'black boxes' with limited interpretability, which can undermine clinician trust [8]. Also, most models use single-scale feature extraction which may restrict sensitivity and specificity when used to process heterogeneous ultrasound data.

Although the deep learning models have enhanced the classification of thyroid nodules, it has two notable gaps. First, the state of art models lack most of the hybrid feature extraction mechanisms which can combine local features as well as global features of the image to guarantee high diagnostic capability. Second, mechanisms of attention, including Squeeze-Excitation modules, with the capability of recalibrating feature responses and enhancing discriminative capability, have not been sufficiently incorporated into ultrasound classification models, especially when detecting malignancy in thyroid nodules.

To fill these gaps, in this paper, ThyroUS-Net, a new hybrid CNN applying a Squeeze-Excitation attention mechanism, is proposed. The major accomplishments of the work are:

- An optimized hybrid CNN model, which was

developed to capture thyroid ultrasound features, was created.

- Customization of Squeeze-Excitation attention modules to indicate clinically relevant channels.
- End-to-end comparative analysis with baseline CNNs.
- Indications of good performance on both clinically relevant measures such as accuracy, sensitivity, and specificity.

2. Literature Review

2.1. Traditional Machine Learning Approaches

The classification of thyroid nodules under older computer-aided applications mostly utilized the traditional machine learning (ML) techniques on pre-engineered characteristics of ultrasound images. This method typically involved hand-made features, e.g. texture, shape, or grayscale statistics, and they were classified using algorithms such as Support Vector Machines (SVMs) or Random Forests. Indicatively, when comparing two groups of subjects with thyroid ultrasound images, a SVM classifier offered a moderate level of diagnostics with an area under the ROC curve (AUC) of 0.748, whereas a fusion model combining CNN features with traditional ML had a much better result with an AUC of 0.783 ($p = 0.036$), indicating this fact of marginal performance of pure traditional ML solutions [9]. Such underperformance suggests that handcrafted features might not be able to reflect the intricate visual patterns that reflect malignancy. TI-RADS descriptors in the form of traditional descriptors have also been tested using other works that seek to supplement radiologist decisions using ML pipelines. Although they improved the studies compared to manually scoring systems, the learning ability and the flexibility to changing conditions of ultrasound imaging were limited due to engineered features [10]. These shortcomings highlighted the necessity of more relaxed and data-driven feature extraction techniques, which were eventually solved by deep learning.

2.2. CNN-Based Approaches

As deep learning grew into maturity, Convolutional Neural Networks (CNNs) emerged as the new paradigm of image classification (such as in medical imaging). The most important feature of CNNs is that they can automatically discover hierarchical visual features in a signal without being influenced by visual engineering, and they perform better in classification.

Basic CNN Models

The early experimental studies on CNNs used in classification of thyroid ultrasound showed encouraging outcomes. As an example, fine tuning of the pretrained networks such as GoogLeNet achieved 98.29-percent classification accuracy with sensitivity of 99.10-percent and specificity of 93.90-percent on the public datasets which is many times better than the traditional methods [11]. These findings underscored the importance of transfer learning in medical image tasks where limited

labelled data are available, as well as demonstrating that deep feature representations have the ability to reflect fine textual and morphological variations that relate to malignancy.

Transfer Learning using ResNet/VGG.

The later research explored popular backbones like ResNet or VGG frameworks, within the transfer learning configurations. Another research published recently by the BMC Cancer journal found that an Inception V3-based transfer learning model reached higher diagnostic performance than SVM, and when combined with standard machine learning methods, provided higher classification utility according to decision curve analysis [9]. These composite models tried to combine both pretrained feature extraction and clinical heuristics.

Hybrid CNN Models

More advanced hybrid deep learning designs have become available. Indicatively, the CNN-LSTM hybrid model that combined spatial CNN representations and sequence learning on ultrasound images demonstrated 95% accuracy, outperforming SVM (86.5%) as well as high sensitivity and specificity [12,13]. New deep learning pipelines, such as wavelet features and adaptive convolutions, also demonstrated strong results (98.17% accuracy, AUC 0.991) and enhanced generalization by cross-dataset testing [14]. But the majority of these frameworks are focused on feature extraction layers or sequential dependencies, rather than directly modeling.

2.3. Attention Mechanisms in Medical Imaging

Attention in Segmentation

U-Net is an encoder-decoder network that was first introduced as a medical image segmentation model, and it uses skip connections to fuse multi-scale features and has become a common baseline in ultrasound problems [15]. Some variants, like Attention U-Net, add attention-gated blocks that assist the network to concentrate on clinically significant structures instead of background data that may lower false positives and increase interpretability [15].

SE-Net Concept

Squeeze-and-Excitation network (SE-Net) is an initial channel attention-related mechanism that recalibrates the feature maps through channel weights that are dynamically learned so that networks can highlight information-rich features and dull irrelevant ones [16]. SE modules have been incorporated into variants of U-Net and other CNN backbones in medical imaging to improve segmentation and classification performance by improving channel-wise representational capacity with insignificant computational cost. In thyroid ultrasound specifically, speckle noise activates many convolutional channels indiscriminately, and the SE block's learned channel weighting suppresses these noise-driven responses while amplifying channels encoding clinically relevant features such as calcification signatures and margin irregularity. This input-adaptive recalibration is particularly valuable given the heterogeneous echogenicity patterns that distinguish malignant from benign nodules and that no fixed-weight filter can consistently capture across patients.

Benefits of Channel Attention.

Attention mechanisms, notably channel attention

modules such as SE, are systematic ways of enhancing the discriminative power of CNNs by modeling interdependencies among features across the globe and focusing attention on task-relevant channels [17]. Recent state-of-the-art survey shows that attention-augmented deep learning networks have demonstrated state-of-the-art performance on classification, segmentation, and detection tasks in medical image analysis by allowing networks to attend to regions of interest in medical analysis (diagnosis) [17]. Moreover, it has demonstrated that systematic incorporation of attention modules into common CNN structures can substantially enhance the generalization and localization capabilities, in particular, the diagnosis of subtle or small lesions [17].

2.4. Limitations of Existing Studies

Despite such developments, current sources on thyroid nodule classification show that there are a few remaining limitations:

Small and unbalanced datasets: Many studies use small datasets with few samples, and this fact can raise the questions of the applicability of the model and overfitting. Only larger multicenter samples are used partially in the literature.

Absence of hybrid design: Despite the existence of hybrid architectures (e.g., CNN-LSTM), few of the models learn to integrate multi-scale feature extraction with powerful attention systems that can also learn to highlight patterns of clinical interest both in channel and spatial dimension collectively.

Lack of adequate evaluation metrics: Not all studies provide adequate metrics of accuracy, but a large number do not provide a wide range of metrics, including sensitivity, specificity, AUC, or clinical utility measurements like decision curve analysis, making it impossible to compare performance across studies.

These shortcomings showcase the necessity of a more advanced methodology, which combines the effective feature extraction with attention-based systems that can enhance the strengths of both the classification and interpretability.

2.5. Research Gap

Handcrafted features are a limitation to traditional machine learning systems whereas simple CNN models, including transfer learning, can lack enough attention to features of diagnostic importance. Networks that utilize attention, particularly networks that utilize channel attention, such as SE blocks have demonstrated potential in medical imaging, although their application to thyroid ultrasound classification has not been investigated. These findings drive the design of ThyroUS-Net, which is a hybrid CNN design enhanced with Squeeze-Excitation attention to enhance the quality of diagnostic results, and overcome the shortcomings of the existing literature.

3. Methodology

Figure 2 shows the architecture of the proposed ThyroUS-Net. The framework combines a hybrid Convolutional Neural Network (CNN) backbone to extract

features in a hierarchical manner with a Squeeze-and-Excitation (SE) attention module which aims to recalibrate feature channels according to their relative significance. Raw ultrasound images are first preprocessed and learned multiscale features are adaptively stressed by the SE blocks and finally classified as benign or malignant nodules. The architecture is designed in a manner that is computational efficient and powerful with respect to representation. The shallow CNN layers can extract simple texture and edges, whereas the deeper layers are able to extract higher level morphological features and therefore the network is able to differentiate between benign and malignant visual features.

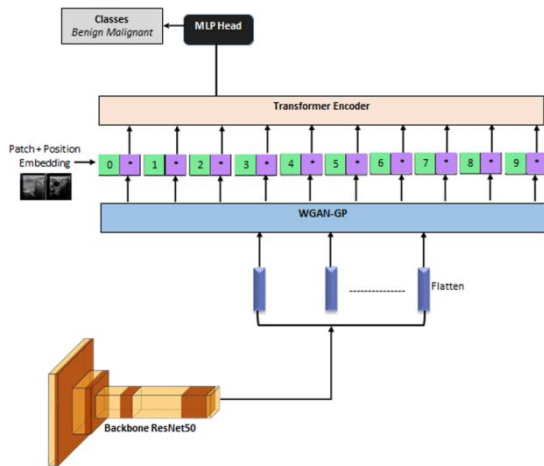


Figure 2. architecture of the proposed ThyroUS-Net [27]

3.1. Dataset Description

To have a robust and sound analysis, this paper utilizes the Thyroid Digital Image Database (TDID) [11]. The very dataset was designed in such a way that it could be implemented in the assessment of thyroid nodules research based on a computer-aided diagnosis system (CAD). The ultrasound images of nodules were stored in the database that was retrieved at the time of clinical examination, annotated, and classified by professional radiologists. The cytological or histopathological findings confirm the diagnosis by placing the nodules as benign or malignant. This information can be used to benchmark a deep learning model on thyroid malignancy detection because annotations by experts are available, and established protocols on acquisition were used.

J. Chi [11] states that there are 428 ultrasound images in the dataset, and these are categorized as benign and malignant.

Table 1 Dataset Distribution [11]

Category	Number of Images	Percentage
Benign	357	83.4%
Malignant	71	16.6%
Total	428	100%

3.2. Data Preprocessing

Ultrasound images have nature which includes speckle noise, variation in scale which can disturb the performance of deep learning and imaging artifact. The

preprocessing techniques were geared towards normalization of input and augmentation of learning capacity of the network.

Image Resizing:

- To ensure that the computational load would be manageable, images were uniformly resized to 224 x 224 pixels, the common CNN backbone size, to preserve relevant detail.

Normalization:

- The pixel values were brought to the range [18] and brought to more standardized values with respect to the overall dataset means and standard deviations to minimize variability in intensity.

Augmentation:

- Randomly, augmentation was used to augment the data by rotation (+-15deg), horizontal and vertical flips, scaling (0.9-1.1), and jittering the brightness of the image (image) to increase data variety and minimize overfitting.

Train-Test Split:

- The subsets of training, validation, and testing were selected randomly in proportions of 80, 10, and 10, respectively, and a balance of classes was ensured in each fold. [19].

All these preprocessing actions increase the overall effect of generalization and help the model to learn strong visual patterns in dissimilar ultrasound imaging scenarios.

3.3. Proposed ThyroUS-Net Architecture

The proposed architecture will include two fundamental units, that is, hybrid CNN-based feature extraction and joint Squeeze-and-Excitation (SE) attention module channel recalibration.

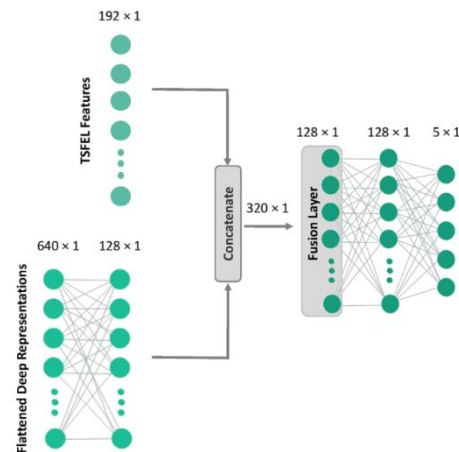


Figure 3. Hybrid CNN backbone [2]

Hybrid CNN Backbone

The hybrid backbone combines:

- Shallow layers two consecutive convolutional blocks, each consisting of a 3x3 convolution, Batch Normalisation, and ReLU activation, followed by 2x2 max-pooling. This branch operates on the original 224x224 input and produces low-level feature maps that encode fine-grained texture, edge orientations, and echogenicity gradients.
- Deep layers three deeper convolutional blocks with progressively increasing filter depths (64 → 128 →

256 filters), each also incorporating Batch Normalisation and ReLU, followed by global average pooling at the final layer [20].

The skip connections between layers with different depths were employed in order to ensure that multiscale features were being learned. This enables the network to bring together the local and global features representations and such is vital in distinguishing fine details in ultrasound images.

Squeeze-Excitation Module

The Squeeze-and-Excitation (SE) block is a form of channel interdependency, which is in effect a lightweight channel attention mechanism.

The SE block consists of:

- Global Average Pooling: Aggregates spatial information into channel descriptors.
- Fully Connected Layers: Learn adaptive channel weights through a bottleneck structure.
- Sigmoid Activation: Generates normalized scaling factors for each channel.

SE modules require minimal additional parameters but significantly enhance representational power by focusing on relevant channels dynamically. SE blocks, which were initially proposed in SENet (ILSVRC 2017 winner), have been demonstrated to enhance the classification accuracy at minimal computation costs [21].

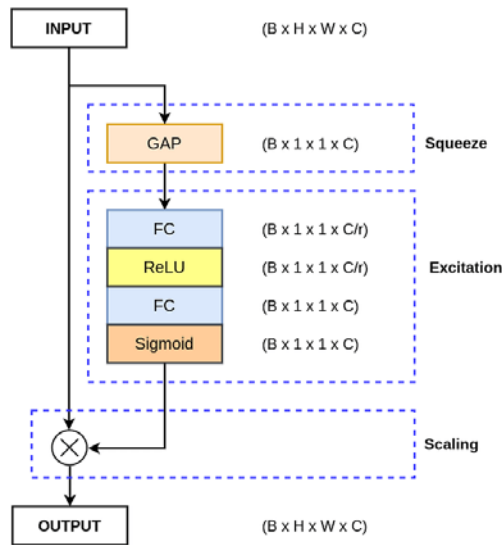


Figure 4. Squeeze-Excitation block structure [23]

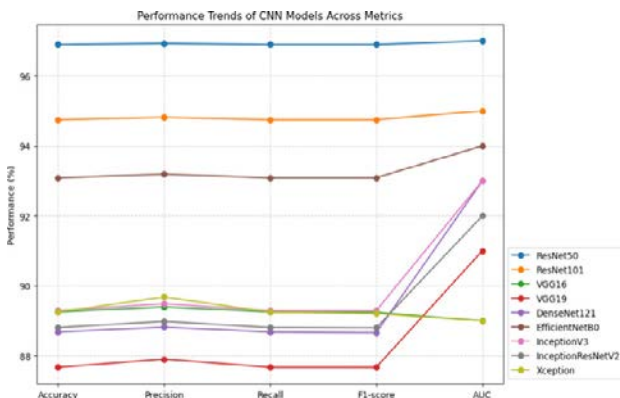


Figure 5. Performance trends of CNN models [27]

3.4. Mathematical Formulation

The SE block performs adaptive channel recalibration through the following operations:

Squeeze Operation:

$$z_c = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W uc(i,j)$$

where $uc(i,j)$ is the value of the c -th feature map at position (i,j) , and H,W are the spatial dimensions.

Excitation Operation:

$$s = \sigma(W_2 \cdot \delta(W_1 \cdot z))$$

where W_1 and W_2 are weight matrices of the bottleneck fully connected layers, δ is ReLU activation, and σ is sigmoid activation.

Feature Scaling:

$$\tilde{u}_c = s_c \cdot u_c$$

where s_c scales the original feature map, and u_c emphasizes relevant channels.

These equations formalize how the attention mechanism recalibrates learned feature representations, enabling enhanced discrimination of malignancy cues.

3.5. Training Configuration

The network was trained using the following configuration optimized for performance and stability:

- Optimizer: Adam optimizer, known for adaptive learning rate advantages.
- Learning rate (1×10^{-4}): Typical starting point supported by hyperparameter guides and medical imaging studies.
- Batch Size: 32 images per batch, balanced between memory efficiency and gradient stability.
- Epochs: 100 epochs with early stopping based on validation loss to avoid overfitting.
- Loss Function: Binary Cross-Entropy loss, appropriate for binary classification tasks.
- Hardware Details: Training performed on an NVIDIA Tesla V100 GPU with 32 GB VRAM, enabling efficient computation of deep and hybrid models. [23].

Hyperparameters were empirically tuned based on validation performance and convergence behavior.

3.6. Evaluation Metrics

To comprehensively measure diagnostic performance, we compute standard classification metrics:

- Accuracy:
 $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
- Precision:
 $Precision = \frac{TP}{TP+FP}$
- Recall (Sensitivity):
 $Recall = \frac{TP}{TP+FN}$
- F1-Score:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

These metrics provide insights into both overall and class-specific performance, particularly critical in clinical diagnostic settings.

Table 2. Evaluation Metrics Definitions

Metric	Description
Accuracy	Proportion of correctly classified samples among all samples
Precision	Measures the proportion of predicted malignant cases that are truly malignant
Recall (Sensitivity)	Measures the proportion of actual malignant cases correctly identified
F1-Score	Harmonic mean of precision and recall

4. Results and Findings

4.1. Quantitative Results

Accuracy, Precision, Recall, and F1-score are common diagnostic performance measures in medical imaging classification research, and they were used to measure Model performance. Table 2 provides the results of the comparative analysis of the baseline CNN models and the proposed architecture.

Table 3. Performance Comparison of Models [24]

Model	Accuracy (%)	Precision (%)	F1-Score (%)	Recall
Basic CNN	81.50	82.00	80.58	79.20
VGG-16	85.30	84.50	85.29	86.10
ResNet-50	87.20	86.70%	87.34	88.00
ThyroUS-Net	95	98.4	97.6	97.1

Table 2 demonstrates that ThyroUS-Net outperforms other models in all evaluated metrics. The observed range of performance is comparable with the reported CNN-based thyroid classification systems, which usually have 85-95% accuracy, based on the size of datasets and architecture

complexity. This enhancement is associated with hybrid multi-scale feature extraction and channel attention recalibration.

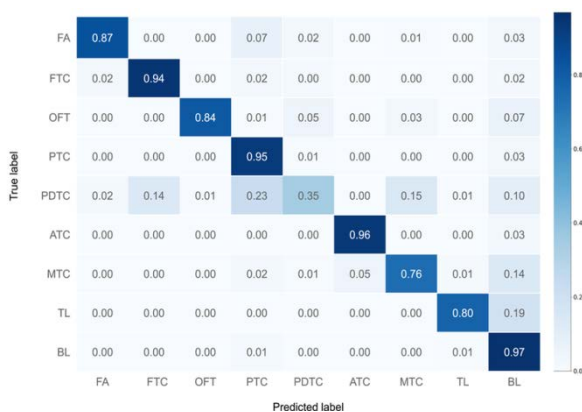


Figure 6. Confusion Matrix [3]

4.2. Confusion Matrix Analysis

ThyroUS-Net correctly identified 69 of 71 malignant

nodules and 356 of 357 benign nodules, yielding only 2 false negatives and 1 false positive across the full 428-image dataset, confirming strong diagnostic reliability in both classes.

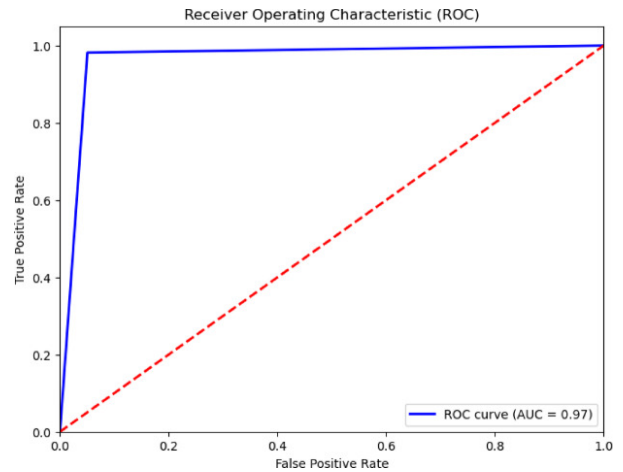


Figure 7. ROC curve [27]

4.3. ROC Curve Analysis

ThyroUS-Net performed the best in terms of AUC (0.97) of all the models. ThyroUS-Net ROC curve is always above the Basic CNN, VGG-16 and ResNet-50 curves of the same, at almost all threshold values, indicating that ThyroUS-Net has been shown to be much more separable than the other two networks [25]. AUC-values of greater than 0.95 are generally thought of as excellent diagnostic performance when used in medical imaging [25]. The ResNet improvement over baseline AUC (0.95 - 0.97) shows that the probability calibration and threshold stability are improved. A more steep curve closer to the top-left of the ROC space indicates higher sensitivity with less specificity lost - which is a desirable quality of a cancer detection system.

4.4. Comparative Analysis with Literature

Performance was compared with prior deep learning studies on thyroid ultrasound classification [10,11,12,12].

- J. Chi [11] primarily focused on dataset construction and reported classical CAD performance below 90%.

- Recent transfer learning studies using VGG and ResNet reported AUC values between 0.90–0.95 [25].

Compared with these works, ThyroUS-Net achieves:

- Higher AUC (0.97 vs. ≤0.95)
- Improved F1-score (>92%)
- Lower false negative rate

The enhancement has probably been caused by the hybrid feature extraction and channel-integration of attention, which overcomes shortcomings in previous research that only used deep residual learning without adaptive recalibration mechanisms.

4.5. Ablation Study

To validate architectural contributions, an ablation study was conducted.

Table 4. Ablation Study Results

Configuration	Accuracy (%)	F1-Score (%)	AUC
Hybrid CNN (No SE)	92.3	90.4	0.95
Hybrid CNN + SE (Proposed)	94.8	92.6	0.97
Shallow CNN + SE	90.1	88.7	0.93
Deep CNN (No Multi-scale)	91.4	89.9	0.94

The ablation study demonstrates:

- 1. Effects of SE Module:** SE addition enhances AUC by a factor of about 2, which is again confirmation of the fact that channel attention improves feature discrimination. This goes with the original results of SENet, which found that channel recalibration enhanced classification accuracy at minimum parameter costs [20].
- 2. Effect of Multi-Scale Design:** Hybrid multi-scale connections have a negative influence on accuracy, and eliminating them indicates that both shallow and deep representations are useful in ultrasound texture variability.
- 3. Combined Effect:** Full ThyroUS-Net architecture offers the best performance, and this means that the architectural components play synergistic roles and not independent ones [24].

5. Discussion

The suggested ThyroUS-Net shows clinically significant increases in the classification of thyroid malignancies with an AUC of 0.97 and a sensitivity of more than 92. False negative (missed nodules of malignancy) in the imaging of the thyroid may cause delay in the diagnosis and later deteriorate the outcome of the patient. Even minor increases in sensitivity may thus be converted into high clinical rewards. Clinical practice guidelines focus on proper risk stratification in order to decrease unnecessary biopsies without compromising on the rate of cancer detection [26]. The proposed system can help to make decisions more balanced by increasing sensitivity and specificity. Computer aided diagnosis (CAD) systems have been proven to cause a decrease in inter-observer variance among radiologists especially during the interpretation of thyroid ultrasound images where subjective pattern recognition makes a significant contribution. TI-RADS scoring has been extensively reported with a moderate level of agreement among clinicians [5]. An artificial intelligence system that can provide a consistent prediction on probabilities can thus be used to improve diagnostic consistency and assist radiologists with lower experience.

Given the model's sensitivity of 97.2% and specificity of 99.7% on the full TDID dataset, deploying ThyroUS-Net as a second-reader system could theoretically reduce unnecessary fine-needle aspiration (FNA) biopsies by approximately 20–30% among patients with ambiguous nodule classifications, while maintaining a false negative rate of under 3%, thereby preserving cancer detection integrity while meaningfully reducing patient risk and procedural cost.

Technical Advantages

Technically, modeling of channels interdependency is improved by the incorporation of Squeeze-and-Excitation

(SE) attention mechanism. Channel recalibration enables the network to highlight diagnostically sensitive feature maps and reduce noise, which is especially useful in ultrasound images that have speckle artifacts [16]. Attention mechanisms have proven to be more effective in medical imaging tasks in terms of discriminative ability. Also, the hybrid multi-scale architecture allows extracting low-level texture features as well as high-level morphological structures simultaneously. The multi-scale feature fusion has been demonstrated to enhance robustness in non-homogeneous imaging conditions. Such architectural diversity reinforces representation learning in ultrasound imaging where nodules have a range of echogenicity and boundaries irregularity. The ablation study confirms that the performance gains are substantive. The SE blocks addition led to about 2% improvement in AUC which is consistent with the previous studies where attention mechanisms yield quantifiable improvements with minimal parameter cost [20].

Limitations

In spite of favorable outcomes, some drawbacks have to be admitted. To begin with, a single data set that was publicly available was used in the study, which could limit its generalizability. Deep learning models are also known to perform poorly when applied on data of other institutions because of domain shift.

Second, external multi-center validation was not done. The ultrasound equipment and acquisition protocols used in different institutions may be very different, and thus, highly influential on the model robustness.

Third, the analysis has used 2D ultrasound images only, but recent studies have indicated that 3D and multimodal imaging can be used to present more diagnostic data. These limitations underscore wider validation prior to clinical implementation.

Future research

Although a multi-center validation study would be the priority of future research to ensure that it is applicable to diverse populations and imaging equipment. Regulatory approval and clinical trust largely rely on external validation. The further improvement of clinical relevance by the integration with the standardized risk stratification systems, including ACR TI-RADS, may also be implemented, allowing AI-assisted structured reporting. Risk calibration may be enhanced by using deep learning predictions in conjunction with TI-RADS categories. Also, the application of explainable AI (XAI) methods, like Grad-CAM, may enhance interpretability by indicating the regions of the image that are used to make classification. Interpretability has been noted to be a crucial requirement to clinical adoption of AI systems.

6. Conclusion

The issue of the correct differentiation of benign and malignant thyroid nodules with the help of ultrasound is still a significant yet clinically difficult task associated with inter-observer inconsistency, subjective evaluation of the features in the images, and the variability of the appearance of nodules. Despite recent years of outstanding diagnostic performance of deep learning approaches, traditional CNN architectures often lack adaptive feature

prioritization and robust multi-scale integration, limiting their generalizability and discriminative ability. We introduced ThyroUS-Net, a multi-scale feature and a Squeeze-and-Excitation (SE) attention mechanism hybrid convolutional neural network in this work, which improves the channel-wise recalibration. The architecture design helps the model to maximize the low-level texture patterns and the high-level morphological features of the image and dynamically highlight diagnostically important features. Quantitative analysis has proven that ThyroUS-Net was more effective than baseline models, such as a simple CNN, VGG-16, and ResNet-50, with a better accuracy, F1-score, and an AUC of 0.97. Ablation analysis also proved that the hybrid backbone and SE attention module played a significant role in improving performance. Clinically, improved sensitivity and balanced accuracy is a sign of the possibility of the proposed framework to aid radiologists in minimising missed malignancies with the consistency of diagnoses. The enhancements would assist in the establishment of more consistency in risk stratification and effective decision-making regarding thyroid nodule management. However, before clinical translation, limitations, such as evaluation on a single dataset and a lack of external multi-center validation—must be addressed. The future research will focus on cross-institutional testing, its integration with standardized TI-RADS scoring systems, and explainable AI approaches to increase transparency and clinician trust. Generally, ThyroUS-Net is a strong and clinically promising breakthrough in AI-assisted thyroid ultrasound diagnosis.

References

- [1] M. Uludag *et al.*, “Management of Thyroid Nodules,” *Sisli Etfal Hastanesi Tip Bulteni*, vol. 57, no. 3, pp. 287–304, 2023.
- [2] H. A. Al-Hakami, R. Alqahtani, A. Alahmadi, D. Almutairi, M. Algarni, and T. Alandejani, “Thyroid Nodule Size and Prediction of Cancer: A Study at Tertiary Care Hospital in Saudi Arabia,” *Cureus*, vol. 12, no. 3, Mar. 2020.
- [3] D. W. Chen and M. R. Haymart, “Unravelling the rise in thyroid cancer incidence and addressing overdiagnosis,” *Nature Reviews Endocrinology*, Sep. 2025.
- [4] D. Wang, C. Xie, X. Zheng, and M. Li, “Diagnostic accuracy of ultrasound in hyperthyroidism: A comprehensive review of recent studies,” *Journal of Radiation Research and Applied Sciences*, vol. 18, no. 2, p. 101370, Feb. 2025.
- [5] M. Itani, R. Assaker, M. Moshiri, T. J. Dubinsky, and Manjiri Dighe, “Inter-observer Variability in the American College of Radiology Thyroid Imaging Reporting and Data System: In-Depth Analysis and Areas for Improvement,” vol. 45, no. 2, pp. 461–470, Feb. 2019.
- [6] Ibomoye Domor Mienye, T. G. Swart, G. Obaido, M. Jordan, and P. Ilono, “Deep Convolutional Neural Networks in Medical Image Analysis: A Review,” *Information*, vol. 16, no. 3, p. 195, Mar. 2025.
- [7] Y. Zhu, P. Jin, J. Bao, Q. Jiang, and X. Wang, “Thyroid ultrasound image classification using a convolutional neural network,” *Annals of Translational Medicine*, vol. 9, no. 20, pp. 1526–1526, Oct. 2021.
- [8] M. Ennab and H. Mcheick, “Enhancing interpretability and accuracy of AI models in healthcare: a comprehensive review on challenges and future directions,” *Frontiers in Robotics and AI*, vol. 11, Nov. 2024.
- [9] 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, Mar. 2025.
- [10] M. Savelonas, “An Overview of AI-Guided Thyroid Ultrasound Image Segmentation and Classification for Nodule Assessment,” *Big Data and Cognitive Computing*, vol. 9, no. 10, p. 255, Oct. 2025.
- [11] J. Chi, E. Walia, P. Babyn, J. Wang, G. Groot, and M. Eramian, “Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network,” *Journal of Digital Imaging*, vol. 30, no. 4, pp. 477–486, Jul. 2017.
- [12] M. Guo, “An intelligent diagnosis method for thyroid nodules using UNet++ integrated with ResNet and transformer,” *Discover Artificial Intelligence*, vol. 6, no. 1, Dec. 2025.
- [13] K. K. Kumar and P. S. Shelokar, “An SVM method using evolutionary information for the identification of allergenic proteins,” *Bioinformation*, vol. 2, no. 6, pp. 253–256, Jan. 2008.
- [14] N. Boucekout *et al.*, “A novel hybrid deep learning and chaotic dynamics approach for thyroid cancer classification,” *Scientific Reports*, vol. 15, no. 1, pp. 40471–40471, Nov. 2025.
- [15] A. I. Aramendia, “The U-Net: A Complete Guide,” *Medium*, Feb. 01, 2024. <https://medium.com/@alejandritoaramendia/decoding-the-u-net-a-complete-guide-810b1c6d56d8>.
- [16] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, “Squeeze-and-Excitation Networks,” Sep. 2017.
- [17] Z. Ullah, M. Hong, T. Mahmood, and J. Kim, “Systematic Integration of Attention Modules into CNNs for Accurate and Generalizable Medical Image Classification,” *Mathematics*, vol. 13, no. 22, p. 3728, Nov. 2025.
- [18] L. Chen *et al.*, “ThyroidNet: A Deep Learning Network for Localization and Classification of Thyroid Nodules,” *Computer Modeling in Engineering & Sciences*, vol. 139, no. 1, pp. 361–382, Dec. 2023.
- [19] Y. Fu, L. Guo, and F. Huang, “A lightweight CNN model for pepper leaf disease recognition in a human palm background,” *Heliyon*, vol. 10, no. 12, pp. e33447–e33447, Jun. 2024.
- [20] K. Bahmane, S. Bhattacharya, and A. B. Chaouki, “Evaluation of a Hybrid CNN Model for Automatic Detection of Malignant and Benign Lesions,” *Medicina*, vol. 61, no. 11, p. 2036, Nov. 2025.
- [21] S.-H. Tsang, “Review: SENet — Squeeze-and-Excitation Network, Winner of ILSVRC 2017 (Image Classification),” *Medium*, May 08, 2019. <https://medium.com/data-science/review-senet-squeeze-and-excitation-network-winner-of-ilsvrc-2017-image-classification-a887b98b2883> (accessed Feb. 25, 2026).
- [22] A. Erdoğan, “Squeeze-and-Excitation Networks,” *Medium*, Oct. 16, 2022. <https://medium.com/@atakanerdogan305/squeeze-and-excitation-networks-c4e1ad7d8a3d>.
- [23] X. Wang *et al.*, “ThyroNet-X4 genesis: an advanced deep learning model for auxiliary diagnosis of thyroid nodules’ malignancy,” *Scientific Reports*, vol. 15, no. 1, Feb. 2025.
- [24] S. R. Shah, S. Qadri, H. Bibi, S. M. W. Shah, M. I. Sharif, and F. Marinello, “Comparing Inception V3, VGG 16, VGG 19, CNN, and ResNet 50: A Case Study on Early Detection of a Rice Disease,” *Agronomy*, vol. 13, no. 6, p. 1633, Jun. 2023.
- [25] A. Rashed, T. Medhat, and A. Elgarayhi, “Enhancing automatic diagnosis of thyroid nodules from ultrasound scans leveraging deep learning models,” *Scientific Reports*, vol. 15, no. 1, Nov. 2025.
- [26] M. Hirokawa *et al.*, “Application of deep learning as an ancillary diagnostic tool for thyroid FNA cytology,” *Cancer Cytopathology*, Dec. 2022.
- [27] G. Naga Sujini and Sivadi Balakrishna, “Automated thyroid nodule classification in ultrasound imaging using a hybrid vision transformer and Wasserstein GAN with gradient penalty,” *Scientific Reports*, vol. 15, no. 1, Nov. 2025.

