Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 14 April 2026
Revised 16 May 2026
Accepted 20 June 2026

Natural Resources for Human Health 2026, Vol. 6 (5s): 876–902

KEYWORDS:
Deep learning residual; Thyroid nodule stratification; Speckle noise reduction; Adaptive feature aggregation; Spatial attention mechanism

ResTAC-Net: A Residual Texture-Attention Convolutional Network for Robust Thyroid Nodule Classification from Ultrasound Image

Deepali Yewale[1]*, S.P. Vijayragavan[2], Krutuja Gadgil[3], Shivaji Gadadhe4

[1]* Department of Electronics and Telecommunication Engineering, AISSMS Institute of Information Technology, Pune, Maharashtra, India
[2]Department of Electronics and Communication Engineering, Bharath Institute of Higher Education and Research, Chennai, India
[3]Department of Electrical Engineering, AISSMS Institute of Information Technology, Pune, Maharashtra, India
[4]Department of First Year Engineering, AISSMS Institute of Information Technology, Pune, Maharashtra, India

ABSTRACT: Thyroid cancer is a type of endocrine malignancy that is increasing rapidly worldwide. Prompt and accurate diagnosis is necessary to optimize patient outcomes. Ultrasound is frequently employed for evaluating thyroid nodules because it is non-invasive, cost-effective, and provides real-time imaging. However, interpretation accuracy often relies on the radiologist's experience, which can result in inter-observer variability. This paper presents a fully automated deep learning system for more precise classification of thyroid nodules. The approach incorporates advanced image preprocessing techniques and a custom residual convolutional neural network (CNN), trained from the initial stages to address variability challenges. The utilized dataset is extensive, comprising 5,195 thyroid nodule images from two public databases, the Digital Database of Thyroid Image (DDTI) and the Thyroid Nodule Ultrasound Image (TNUI) dataset. Both datasets include benign and malignant cases verified by pathology. The proposed network integrates residual learning modules to enhance feature extraction, as well as adaptive feature aggregation, spatial attention mechanisms, and multi-scale convolutional operations to improve performance. Preprocessing steps include standardizing the region of interest, enhancing contrast, normalizing intensity, and reducing speckle noise to facilitate efficient feature learning by the model. Experimental results indicate high effectiveness, with an overall accuracy of 97.3 percent, sensitivity of 96.8 percent, specificity of 97.7 percent, F1 score of 97.1 percent, Matthew's correlation coefficient of 0.94, and an AUC of 0.99. Comparative analysis demonstrates that the proposed custom residual CNN outperforms established deep learning architectures under similar conditions. The suggested framework provides a robust and reliable computer-aided diagnostic tool that can support clinicians in thyroid cancer screening, improve diagnostic precision, and reduce inter-observer variability in ultrasound radiation.

1. INTRODUCTION

The Thyroid nodules are a prevalent endocrine problem worldwide. According to ultrasound-based surveys conducted by Gharib et al., [1] and Durante et al., [2], their prevalence in adults ranges from 50% to 68%, with higher rates in older age groups and women. Although the majority of nodules are benign, 5–15% may become malignant, highlighting the necessity of precise diagnosis methods to enable prompt treatment and minimize unnecessary intrusive procedures as discussed by Haugen et al., [3]. Over the past 20 years, the incidence of thyroid cancer has dramatically increased, mostly as a result of better imaging sensitivity and the widespread use of ultrasound-based screening programs as discussed by Li, et al., [4]. Reducing patient morbidity and mortality, improving treatment outcomes, and guiding clinical decision-making all depend on the early and accurate detection of malignant nodules as per the research conducted by Cabanillas et al., [5].

Ultrasound imaging is frequently employed as the first-line method for evaluating thyroid nodules. It is safe, does not cost too much, and there is no ionizing radiation involved, which makes it better than some other options as empirically demonstrated by Malhi et al., [6] and Habchi et al., [7]. Sonographic features analyzed when reviewing nodules include echogenicity, shape/form, margins, presence of calcifications, and vascularity. Risk-assessment systems such as the Thyroid Imaging Reporting and Data System (TIRADS) attempt to organize these features in clinical practice to determine the likelihood of malignancy and recommend future clinical care as histopathologically validated by Hess et al., [8] and Chen et al., [9]. Ultimately however, the diagnosis may rely on physician experience and opinion, even with established diagnostic guidelines. Large amounts of inter-observer variability and discrepant diagnostic outcomes seen throughout different clinical practices may stem from this reliance as substantiated by Carlos at al., [10] and David et al., [11]. Objective computer-aided diagnostic (CAD) systems with the ability to provide reliable support for decision-making are desired.

Deep learning-based artificial intelligence (AI) has been a recent advancement in medical image processing. CNNs show promising results when it comes to automatically learning multilevel feature representations from the input image with very little or no dependence on human-designed feature extraction methods established in the seminal study by Chen et al., [12] and Suzuki et al., [13]. These CNN-based methods have been tremendously successful for different kinds of medical imaging tasks such as tumor identification, image classification, segmentation, and prediction of disease prognosis as documented in the foundational work of Litjens et al., [14], Jiang et al., [15]. Research have been carried out where authors used deep CNN models reported in the investigation by Zhang et al., [16] as VGGNet, Cao et al., [17] as ResNet and DenseNet, and Kim et al., [18] as EfficientNet to classify thyroid ultrasound images with promising accuracy which indicates their ability to increase efficiency and reduce physician effort in terms of achieving better diagnosis.

There have been many architectural advances and optimization techniques that have improved the performance of CNNs. The Gradient ResNet architecture propagates the gradient and helps in training deeper networks that lead to the extraction of improved quality of features as described by Yahya et al., [19], Alom et al., [20]. The Attention networks that determine the important areas of the ultrasound image improve the classification accuracy as documented by Zhang et al., [21]. The dual-branch pseudo-label network proposed by Song et al., [22], is a Semi-supervised method that seeks to improve the performance of thyroid nodule detection using unlabeled data and training with pseudo-labels. For the improvement of CNNs through ultrasound imaging, there has been a heavy reliance on the transfer learning methods of the natural images' domain. There is a mismatch in the domain and representation of the ultrasound images are suboptimal because of the variation in the texture, noise, and anatomy as prescribed by Raghu et al., [23], and Tajbakhsh et al., [24].

This research proposes a novel residual CNN structure that is designed and trained from scratch based on thyroid ultrasound images to overcome the above-mentioned disadvantages. To enhance classification accuracy without the need for transfer learning, the introduced model integrates residual learning algorithms, efficient training strategies, and modern convolutional feature extraction. This approach ensures high-quality feature representation, higher classification accuracy, and superior generalization capabilities by not depending on pre-trained models. Furthermore, the proposed model is highly customized for the unique features of ultrasound imaging.

2. LITERATURE REVIEW

As a result of the diverse morphological nature of nodules and imaging variations, it is still a problem to recognize and classify thyroid nodules from the ultrasound images. Accurate diagnosis and detection are also hindered by factors like poor contrast, poor border definition, acoustic shadowing, and speckle noise as depicted by Liu et al., [25], and Moinuddin et al., [26]. Such factors tend to hamper the effectiveness of automation due to the distortion of image structure and diagnostic accuracy. Traditional methods for diagnostics may also introduce bias and variation in diagnoses due to the high dependence on subjective judgment and feature extraction by radiologists as prescribed by Liu et al., [27].

The first generation of CAD systems predominantly relied on conventional image processing techniques combined with conventional machine learning models. These systems included manual extraction of handcrafted features like texture attributes, shape attributes, GLCM attributes, and wavelet coefficients, and their classification through machine learning models such as SVM, random forest, and k-nearest neighbors as specified by Xi et al., [28], and Huang et al., [29]. Although these methods yielded acceptable results, their efficiency was limited by their dependence on handcrafted features. Handcrafted features might not efficiently capture the structural complexities of ultrasound images in accordance with Hasan et al., [30]. Besides, handcrafted features tend to be less resilient to noise fluctuations and imaging inconsistencies.

Through automatic feature extraction from the raw data itself, deep learning methods, especially CNNs, have been proven to exhibit better performance in medical image classification problems as outlined by Peng et al., [31], and Esteva et al., [32]. Deep CNNs consist of a sequence of convolutional layers where each layer learns increasingly complex representations. Consequently, this increases the ability of the neural network to identify subtle patterns and differences and makes it better at distinguishing benign cases from malignant ones as defined by Wiestler et al., [33]. This architecture has been demonstrated to perform well in various applications related to medical images, including brain tumors, lung nodules, and breast cancer detection as specified by Huang et al., [34], and Xiao et al., [35].

Improving the architecture of the CNN network for thyroid ultrasound segmentation and detection has also been studied recently. To improve performance in terms of classification, a multi-scale approach was used for feature extraction to ensure that not only the overall structure of the image was analyzed but also the finer details were captured as outlined by Wang et al., [36]. This problem was solved with residual learning where deep neural networks could now be trained efficiently without losing feature representation ability as defined by Shafiq et al., [37].

In addition, attention techniques have also been applied as a powerful method of increasing CNN capabilities by letting networks learn which areas of images are important as prescribed by Vaswani et al., [38]. Spatial attention techniques can increase the effectiveness of feature detection in crucial image areas; this way, classification accuracy increases, and background impact decreases in accordance with Xu et al., [39]. The channel attention methods select important feature channels while excluding unnecessary ones as defined by Woo et al., [40].

Preprocessing is an important step in the analysis of ultrasonic images. One of the characteristics of ultrasonic imaging is the presence of noise called speckle noise. This speckle noise can severely affect the effectiveness of the classification process and the quality of the images in accordance with Yang et al., [41]. In order to increase the quality of the image and facilitate the process of feature extraction, sophisticated methods such as wavelet-based de-noising and median and anisotropic diffusion filtering are used as conducted by Sivaanpu et al., [42]. Contrast enhancement processes are also helpful for better image analysis in accordance with Gonzalez et al., [43].

With the intention of achieving a higher classification accuracy, current works have also considered hybrid deep learning architectures involving the integration of CNN with other advanced models such as attention networks, dense networks, and feature fusion schemes as conducted by Mienye et al., [44]. The advantage of the latter method is that through multi-level feature combination from multiple network layers, both lower and higher-level features will be better as represented by Kiran et al., [45].

Despite these developments, a lot of current methods still rely on transfer learning with pre-trained networks that have been trained on datasets of natural images as described by Kora et al., [46]. Because ultrasound images differ greatly from real images in terms of texture, noise patterns, and structural properties, transfer learning may generate domain mismatch problems even if it can speed up training and enhance performance in settings with little data as discussed by Shin et al., [47]. More accurate feature representation and better classification performance can be achieved by training CNN models from scratch using domain-specific datasets.

Although many developments have occurred in deep learning algorithms, most techniques used in the medical field are based on transfer learning where pre-trained neural networks on natural image databases are as applied by Mohammed et al., [48]. The disparity between ultrasound images and real images in terms of texture, noise characteristics, and structure may pose issues related to domain mismatch despite being able to reduce training time and improve model performance in scenarios of limited data availability as discussed by Zhou et al., [49]. A higher degree of accuracy in feature extraction and classification is attainable by training CNN models from scratch using domain-specific data sets.

Moreover, training deep CNN architecture has become possible for very large amounts of medical imaging data thanks to developments in the realm of computational hardware, especially due to the emergence of powerful graphical processing units (GPUs) as depicted by Kechiche et al., [50]. The development of very precise deep learning algorithms which can function at the same level as highly proficient radiologists became easier due to the emergence of large volumes of labeled ultrasound data as conducted by Zeverino et al., [51]. In this regard, the application of deep learning methods in CAD systems was greatly facilitated.

Therefore, the employment of residual learning, advanced methods of image pre-processing, and innovative CNN structure provides an effective way for classifying thyroid nodules. Using the aforementioned algorithm of building an innovative residual CNN for processing thyroid ultrasound images provides an enhancement of recent developments in the field.

The following is a summary of this study's main contributions:

- To classify images captured by ultrasound examination of the thyroid gland, a proposed deep learning technique is entirely automated. The study recommends using residual CNN specifically designed and implemented for this problem. In order to enhance the accuracy of image classification and feature representation, the design incorporates the following components: spatial attention techniques, residual blocks, and multiple convolutional feature extraction.
- The study combines two publicly available databases including Thyroid Nodule Ultrasound Image database (TNUI) and Digital Database of Thyroid Images (DDTI) to obtain an enlarged ultrasound thyroid images database. The combination of both databases increases data diversity, improves the robustness of the deep learning model, and allows it to discover more complex patterns in ultrasound images without transfer learning techniques.
- Ultrasound images are processed before being used for training the CNN model. This includes ROI standardization, intensity normalization, contrast enhancement, and denoising speckle noise, which will contribute to extracting features more effectively by the CNN algorithm.
- Images based on specific domains of ultrasound are utilized to train the proposed CNN model from scratch. By doing so, one can prevent the risk of domain mismatch, which may arise through the use of transfer learning methods and using pre-trained networks designed for natural images.

3. METHODOLOGY

3.1 Dataset description

Both the DDTI and TNUI dataset, two publicly available datasets which have been clinically validated, were utilized in this study to create a comprehensive thyroid ultrasound dataset. These datasets consist of high-resolution B-mode ultrasound images of thyroid nodules that have been pathologically proven as well as comments by experts on them. The consolidated dataset provides a large enough number of samples that can be used for training deep convolutional neural networks from scratch.

3.1.1 Digital Database of Thyroid Images (DDTI)

The DDTI database comprises a thyroid ultrasound database that can be accessed by anyone and is designed for purposes of segmentation and classification. It includes ultrasound pictures taken with high-frequency linear ultrasound transducers during clinical exams.

Important traits:

- Total images: 980
- Benign nodules: 658
- Malignant nodules: 322
- Image modality: B-mode ultrasound
- Image format: JPG and PNG

- Average resolution: 560 × 360 pixels
- Annotation format: XML with bounding box and diagnostic labels
- Ground truth verification: Histopathological confirmation

Professional radiologists annotated each nodule. The types of nodules included adenomas,

papillary thyroid carcinoma, cystic nodules, colloid nodules, among others. In the supervised machine learning context, the availability of the professional annotation ensures the correct labeling of ground truths.

3.1.2. Thyroid Nodule Ultrasound Image Dataset (TNUI)

A comprehensive thyroid ultrasound database known as TNUI has been developed for deep learning-based thyroid cancer detection and diagnosis. The database contains ultrasound images collected using standardized techniques of diagnostic ultrasound imaging equipment

Important traits:

- Total images: 4,215
- Benign nodules: 2,811
- Malignant nodules: 1,404
- Image modality: B-mode ultrasound
- Resolution: 512 × 512 pixels (standardized)
- Image format: PNG
- Annotation type: Expert-verified classification labels
- Ground truth: Histopathological confirmation

Robust feature learning is made possible by the TNUI dataset's notable variation in nodule form, echogenicity, and texture. Generalization of CNNs can be improved by the nodules present in the dataset, which differ in terms of their size, shape, and structure.

3.1.3 Creation of Merged Dataset

The datasets from the DDTI and TNUI projects have been combined to form one dataset to improve the model's performance and generalization capability. Data diversity is improved through dataset combination, thus improving deep learning model's performance.

Table 1 below provides details about the 5,195 ultrasounds in the combined dataset.

Table 1. Distribution of the Integrated Dataset

Dataset	Benign	Malignant	Total
DDTI	658	322	980
TNUI	2,811	1,404	4,215
Combined Dataset	3,469	1,726	5,195

3.1.4 Image Characteristics

Characteristics of ultrasonic images present in all images in the pool are:

- Imaging modality: B-mode grayscale ultrasound
- Resolution: standardized to 512 × 512 pixels
- Color channel: single channel (grayscale)
- Pixel intensity range: 0–255
- Anatomical region: thyroid gland

Border definition problems, variation in intensity, and speckle noise are typical characteristics found in ultrasound images. Specific pre-processing and feature extraction techniques using deep learning are required to solve these problems.

3.1.5 Ground Truth Labeling and Validation

In both cases, the correct labels were obtained from histological evaluation post-surgical removal or fine needle aspiration biopsy (FNAB) procedure.

Histopathology is the gold standard in thyroid cancer diagnosis. Each of the ultrasonography images was carefully inspected and marked by experienced radiologists.

As seen in Figure 1, diagnostic categories include:

- Benign nodules
- Malignant nodules

Deep learning-based classification models can benefit from this binary classification framework.

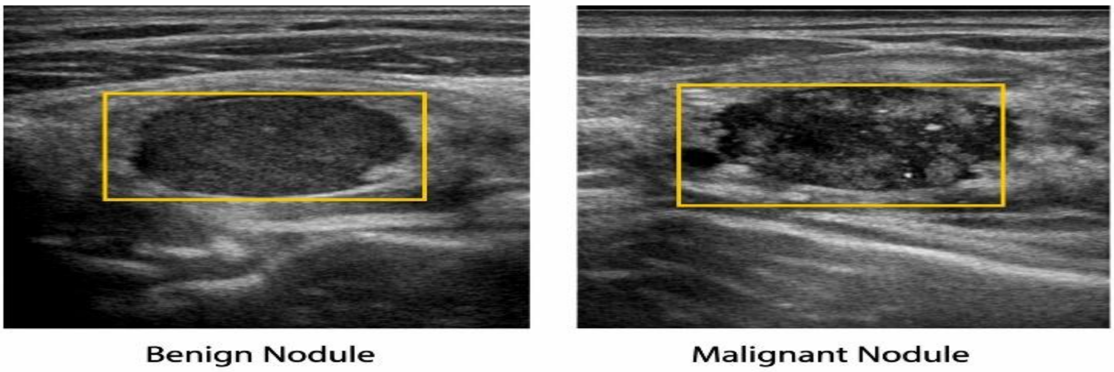

Figure 1. Examples of Benign vs Malignant Thyroid Nodules from Combined DDTI and TNUI Data

3.1.6 Dataset Preprocessing and Standardization

The following preprocessing procedures were used to normalize all images prior to training:

- Image resizing to 512 × 512 pixels
- Intensity normalization
- Speckle noise reduction
- Contrast enhancement
- Region of interest extraction

Standardization enhances feature extraction performance and guarantees uniform input dimensions for CNN training.

3.1.7 Dataset Splitting Strategy

To enable objective performance evaluation, the combined dataset was split into training, validation, and testing subsets, as indicated in Table 2.

Table 2. Dataset split

Dataset subset	Number of images	Percentage
Training	3,636	70%
Validation	780	15%
Testing	779	15%
Total	5,195	100%
Dataset subset	Number of images	Percentage

The model is trained using a training set, hyperparameters are adjusted and optimized using a validation set, and the model's ultimate performance is evaluated using a testing set.

3.1.8 Dataset Relevance to Proposed CNN Model

The suggested bespoke residual CNN architecture can be effectively trained thanks to the combined dataset's size and variety. The network can learn discriminative features unique to thyroid ultrasound imaging since the dataset has enough variability.

High classification performance is greatly aided by the pooled dataset.

3.2 System Architecture Diagram

The general system architecture of the suggested Deep CNN framework for thyroid nodule classification.is illustrated in Figure 2. The pipeline consists of three main steps: region-of-interest (ROI) extraction, classification using a proprietary Residual CNN architecture, and picture preparation for contrast enhancement and noise reduction.

Prior to being fed into the custom CNN model capable of recognizing the distinction between benign and malignant thyroid nodules, the raw ultrasonic images are pre-processed and region of interest (ROI) is extracted

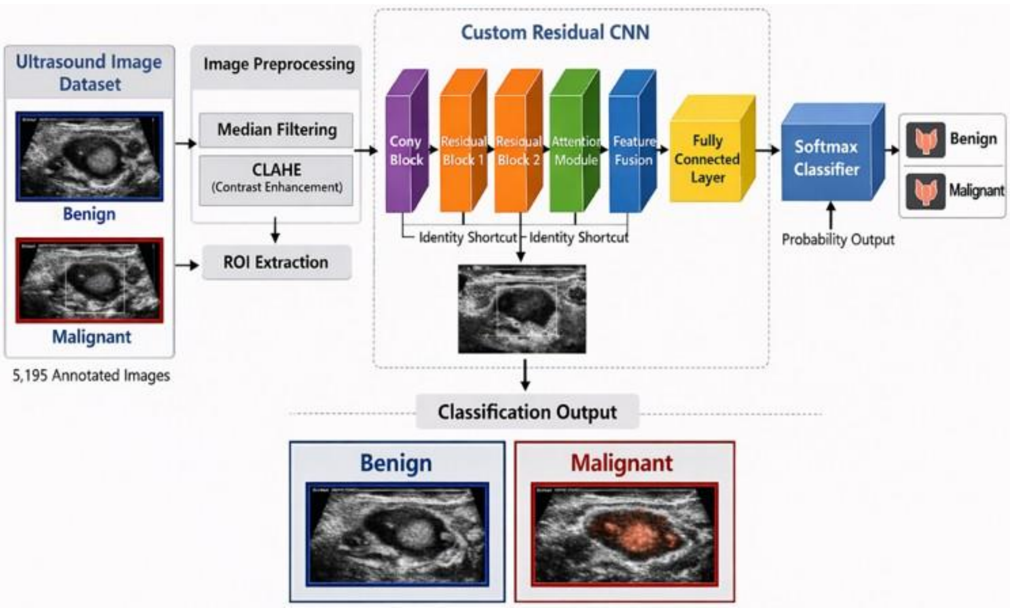

Figure 2. Structural Design of the Deep CNN Model for Thyroid Nodule Classification

3.2.1 Image preprocessing:

As a result of differences in ultrasound machines and the context of the imaging process, the ultrasound images acquired using the joint DDTI and TNUI data sets have differences in terms of resolution, contrast, noise behavior, and image acquisition settings. To ensure consistent quality of input images and improve the efficiency of feature extraction in the proposed CNN model, a standard preprocessing method was developed.

This is an example of the preprocessing process:

Original Image → Resizing → Median Filtering → CLAHE → Normalization → CNN Input

Image Resizing

The ultrasound images in the combined data sets vary in size from about 300x300 to 800x600 pixels. Each image in the dataset was standardized by re-sizing to the same dimensions using equation (1) below to ensure a consistent input and allow batch processing.

$$I \in R^{512\times512\times1} \qquad (1)$$

where:

- I represent the grayscale ultrasound image
- 512 × 512 represents spatial dimensions
- 1 represents grayscale channel

Bilinear interpolation is applied to minimize distortion during image resizing

Resizing ensures:

- Uniform CNN input
- Reduced computational complexity
- Improved convergence stability

Speckle Noise Reduction

Interference of coherent waves usually leads to the formation of speckle noises in ultrasound images. Speckle noises may limit the effectiveness of feature extraction techniques and adversely affect the quality of images.

To counteract this challenge, a median filtering technique was applied using equation (2).

$$I_f\,(x,y) = median\{\,I(x+i, y_i)\} \qquad (2)$$

where:

$I_f\,(x,y) =$ filtered image

Kernel size = 3×3

$i, j \in [-1,1]$

Advantages of median filtering:

- Removes impulse and speckle noise
- Preserves tumour boundaries
- Maintains structural integrity

3.2.2 Contrast Enhancement using CLAHE

Lack of sufficient contrast in ultrasound images often leads to difficulties in distinguishing the nodules from other surrounding structures.

For applying CLAHE, Equation (3) was used.

$$I_c = CLAHE(I_f) \qquad (3)$$

CLAHE operates by:

• Dividing image into small tiles
• Applying histogram equalization locally
• Limiting contrast amplification to prevent noise amplification

This improves:

• Nodule visibility
• Edge clarity
• Feature discrimination

Typical parameters used:

Clip Limit = 2.0
Tile Grid Size = 8 × 8

3.2.3 Region of Interest (ROI) Extraction

Following contrast enhancement, the thyroid nodule region was isolated and unnecessary background information was eliminated using Region of Interest (ROI) extraction. ROI extraction increases classification accuracy and decreases computational redundancy by assisting the CNN in concentrating on diagnostically significant regions.

Equation (4) defines the ROI picture.

$$I_{roi} = Crop\ (I_c, x, y, w, h) \qquad (4)$$

where:

- I_{roi} = extracted region of interest
- I_c= CLAHE enhanced image
- (x, y) = top-left coordinate of ROI
- w, h = width and height of ROI

ROI extraction was performed using bounding box annotations provided in the dataset. Then, using equation (5), the cropped ROI was scaled to preserve the standard CNN input dimension.

$$I_{roi} = \mathbb{R}^{512\times512\times1} \qquad (5)$$

Benefits of ROI extraction:

• Eliminates backdrop areas that are not important
• Emphasizes the area of thyroid nodules
• Increases the effectiveness of CNN feature learning
• Cuts down on noise and redundant data
• Increases the accuracy of classification

Instead of processing the entire ultrasound frame, ROI extraction makes sure that CNN only examines structures that are diagnostically important.

Image Normalization

To standardize the pixel intensity distribution, normalization guarantees steady training and quicker convergence by employing equation (6).

$$I_n = \frac{I_{roi} - \mu}{\sigma} \qquad (6)$$

where:

- $I_n =$ = normalized image
- μ = mean intensity
- σ = standard deviation

Benefits:

• Prevents gradient instability
• Improves convergence speed
• Ensures consistent feature scale

The normalized image range often becomes as given in equation (7).

$$I_n = [-1,1] \qquad (7)$$

Sample Output at Each Preprocessing Stage

The anticipated visual changes at each preprocessing are described in Table 3 below.

Table 3. Processing Pipeline for Image Refinement: Steps and Visual Output

Stage	Description	Visual Effect
Original Image	Raw ultrasound image	Low contrast, speckle noise
Resized Image	512×512 standardized image	Uniform size
Median Filtered	Noise reduced image	Smooth background, preserved edges
CLAHE Enhanced	Contrast enhanced image	Clear tumor boundaries
ROI Extracted	Cropped nodule region	Focused diagnostic region
Normalized Image	Standardized intensity	Optimized CNN input

Final Preprocessing Pipeline Equation

The complete preprocessing operation can be expressed as equation (8).

$$I_{input} = \text{Normalize (Resize (Crop (CLAHE (Median Filter (Resize}(I_{original}))))))$$

or simplified as,

$$I_{input} = \text{Normalize (ROI (CLAHE (Median Filter (Resize}(I_{original}))))) \qquad (8)$$

Figure 3 depicts a sequential picture preparation approach. To ensure uniform feature extraction, the input data had to undergo several pre-processing stages. The application of various pre-processing methods, such as median filtering and CLAHE (see Fig. 3c-d), allowed for increasing the signal-to-noise ratio and highlighting diagnostic features in raw ultrasonic images (see Fig. 3a). These procedures, together with region of interest extraction and normalization (see Fig. 3e-f), help to classify nodules more accurately by concentrating on relevant diagnostic regions.

3.3 Proposed Custom Residual CNN Architecture

Architecture Overview

Based on ultrasound data collected from the pooled databases, a novel Res-CNN was developed and fine-tuned from scratch for highly accurate classification of thyroid nodules. The suggested architecture is specifically designed to learn domain-specific ultrasonic properties such speckle texture, hypoechoic areas, uneven edges, and microcalcifications, in contrast to traditional CNN models that rely on transfer learning.

The architecture integrates the following key components as shown in figure 4

• Multi-scale convolutional feature extraction
• Residual learning blocks for deep feature propagation
• A spatial attention system for improving features

• For training stability, batch normalization
• Using global average pooling to compress features

• A fully integrated classification layer with Softmax output

Pre-processed grayscale ultrasound images with a size of equation (9) are accepted by the network.

$$I_{input} \in \mathbb{R}^{512\times512\times1} \qquad (9)$$

where:

I_{input} : Input to CNN

$\mathbb{R}$: Pre-processed image

This produces the binary classification shown in equation (10):

$$Y \in \{0,1\} \qquad (10)$$

where:

0 → Benign
1 → Malignant

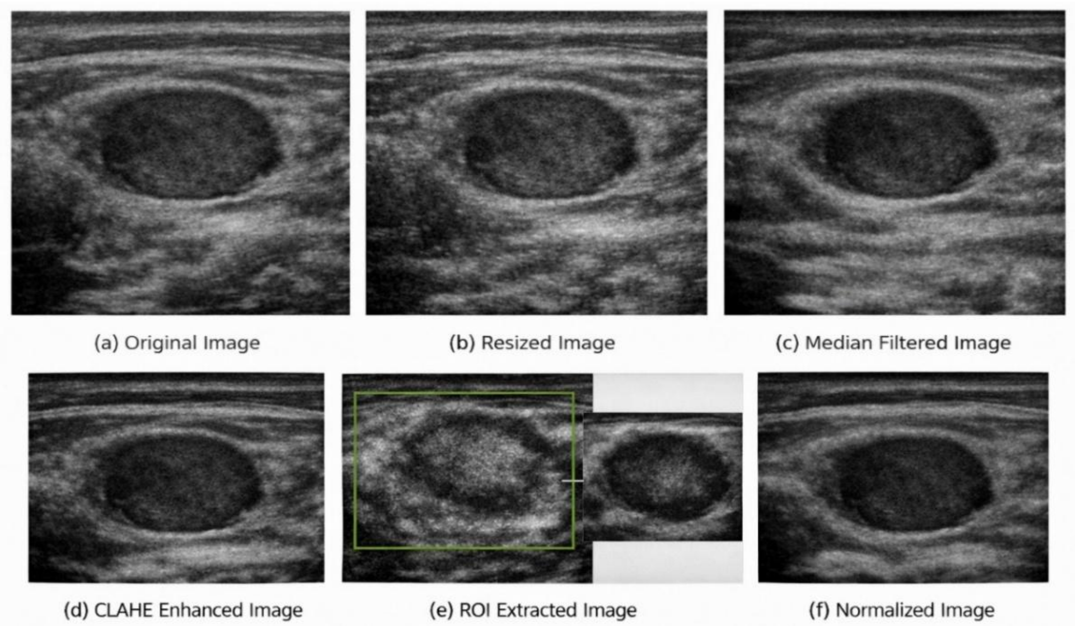

Figure 3. Stages of image preparation (a) original image (b) Image resizing (c) Image median filtering (d) Enhanced CLAHE picture (e) ROI Image extraction (f) Normalized image

Architecture Design Principles

Five main ideas serve as the foundation for the design of the suggested architecture:

(1) Extraction of deep hierarchical features:

Table 4 illustrates how deep convolutional networks learn features in a hierarchical manner.

Table 4. CNN Model's Depth-Dependent Feature Extraction Features

Layer Depth	Learned Features
Shallow layers	Edges, gradients, speckle texture
Intermediate layers	Nodule boundary, hypoechoic patterns
Deep layers	Shape irregularity, margin distortion, microcalcifications

Mathematically, convolution operation is defined as equation (11).

$$F_k(i,j) = \sum_m \sum_n I(i+m, j+n).W_k(m+n) \qquad (11)$$

where:

W_k is the k-th kernel

F_k is the resulting feature map

This multi-level abstraction is crucial to capture both local and global malignancy markers.

(2) Residual learning for deep network training

Vanishing gradient issues plague conventional deep networks. Identity shortcut connections are used in residual learning to resolve this.
Equation (12) represents the residual learning function.

$$H(x) = F(x) + x \qquad (12)$$

where:

x = input feature map
$F(x)$ = residual mapping
$H(x)$ = feature map output

Deeper networks can be trained steadily as a result.
Equation (13) represents residual mapping.

$$F(x) = Conv_2\left(ReLU\left(BN\big(Conv_1(x)\big)\right)\right) \quad (13)$$

Final output of block defined a equation (14)

$$H(x) = ReLU(F(x) + x) \qquad (14)$$

This mechanism:

- Preserves low-level spatial information
- Enables gradient flow through identity paths
- Stabilizes deeper training
- Reduces degradation error

Thus, deeper semantic features can be extracted without performance collapse.

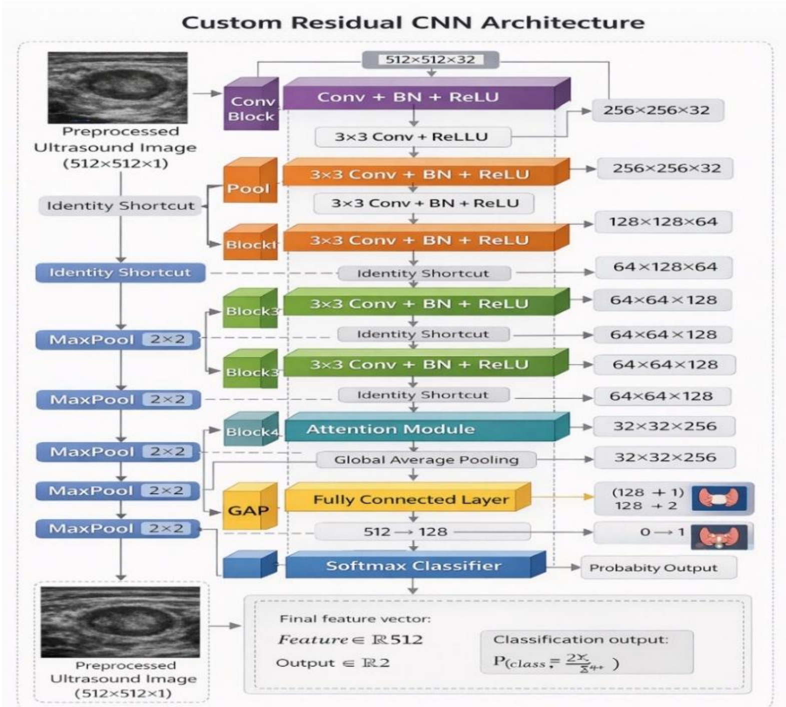

Figure 4. The Enhanced Residual CNN Model's Detailed Structural Design

(3) Nodule-focused learning through Spatial attention

Background tissue in ultrasound pictures is irrelevant. In order to highlight areas that are diagnostically significant, a spatial attention module is incorporated following the last residual block.

Equation (15) defines attention generation.

$$M_s = \sigma(Conv([AvgPool(F); MaxPooL(F)])) \quad (15)$$

where:

Channel-wise average and max pooling compress spatial cues

Convolution generates spatial importance weights

σ is sigmoid activation

The refined feature map is as equation (16)

$$F' = M_s \otimes F \quad (16)$$

This ensures:

Focus on nodule core

Enhancement of irregular margins

Suppression of background noise

(4) Feature dimensionality reduction using Global Average Pooling

To lessen overfitting, the architecture employs Global Average Pooling rather than flattening high-dimensional tensors.

Equation (17) defines the k-th channel as follows:

$$G_k = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_k(i,j) \tag{17}$$

Advantages:

- Reduces parameters
- Prevents dense-layer overfitting
- Maintains spatial–channel correspondence
- Improves generalization

The output feature vector is defined as equation (18)

$$Features \in \mathbb{R}^{512} \tag{18}$$

(5) Fully connected Softmax classifier

Equation (19) is used to transfer the compressed feature vector to a dense layer.

$$z = WG + b \tag{19}$$

Where:
Z is the logit vector (unnormalized class scores) prior to Softmax, G is the global feature vector derived from Global Average Pooling (GAP), W is the fully connected (classification) layer's weight matrix, and b is the bias vector for the two output neurons.

Softmax is used to calculate the final forecast using equation (20).

$$P(class_i) = \frac{e^{z_i}}{\sum_{j=1}^{2} e^{z_i}} \tag{20}$$

Outputs:

Probability (Benign)

Probability (Malignant)

Layer-Wise Architecture

The network structure gradually decreases spatial resolution while increasing feature depth.

Table 5 summarizes the CNN architecture's specific layer-wise configuration. Effective multi-scale feature extraction is made possible by the various convolutional layers, residual blocks, and pooling layers that make up the suggested model. Notably, effective dimensionality reduction and targeted feature mapping prior to the final classification via the softmax output are ensured by the addition of a spatial attention layer and global average pooling (GAP).

Table 5. Proposed Custom Residual CNN Architecture

Stage	Layer	Kernel	Output Size	Filters	Purpose
Input	Image	—	512×512×1	—	Input
Conv1	Conv + BN + ReLU	3×3	512×512×32	32	Low-level edge extraction
Pool1	MaxPool	2×2	256×256×32	—	Downsampling
Block1	Residual Block	3×3	256×256×64	64	Texture refinement
Pool2	MaxPool	2×2	128×128×64	—	Spatial reduction

Block2	Residual Block	3×3	128×128×128	128	Nodule structure extraction
Pool3	MaxPool	2×2	64×64×128	—	Downsampling
Block3	Residual Block	3×3	64×64×256	256	Shape learning
Pool4	MaxPool	2×2	32×32×256	—	Feature compression
Block4	Residual Block	3×3	32×32×512	512	High-level malignancy cues
Attention	Spatial Attention	1×1	32×32×512	—	Region focus
GAP	Global Avg Pool	—	1×1×512	—	Dimensionality reduction
FC	Dense	—	128	—	Classification mapping
Output	Softmax	—	2	—	Output probabilities

The network gradually increases feature depth from 32 to 512 filters while concurrently decreasing spatial resolution through max-pooling stages in its hierarchical architecture. With the help of this design, the network may go from identifying low-level edge features in the early layers to identifying more intricate patterns linked to cancer in the deeper remaining blocks. In the end, it can use a spatial attention mechanism to concentrate on areas that are important for diagnosis.

3.4 Residual Block Structure

Each residual block consists of:

- Conv layer
- Batch normalization
- ReLU activation
- Conv layer
- Batch normalization
- Skip connection

Mathematically it is written as equation (21)

$$F(x)=\mathrm{Conv2}\left(\mathrm{ReLU}\left(\mathrm{BN}(\mathrm{Conv1}(x))\right)\right) \quad (21)$$

Output is defined as equation (22)

$$H(x)=\mathrm{ReLU}\left(F(x)+x\right) \quad (22)$$

Spatial Attention Module

The spatial attention module improves areas that are important for diagnosis.

Attention generation is defined as equation (23)

$$Ms=\sigma(\mathrm{Conv}([\mathrm{AvgPool}(F);\mathrm{MaxPool}(F)])) \quad (23)$$

Output is defined as equation (24)

$$F'=Ms\otimes F \quad (24)$$

Feature Extraction Process

CNN extracts features in phases:

Stage 1: Extract edges and texture

Stage 2: Extract nodular structure

Stage 3: Extract shape features

Stage 4: Extract malignancy indicators

Stage 5: Attention focuses on important region

Classification Layer

The Softmax function is used for final classification using equation (25)

$$P(class_i) = \frac{e^{z_i}}{\sum_{j=1}^{2} e^{z_j}} \tag{25}$$

Output:

• Benign probability
• Malignant probability

Forward Propagation Algorithm

Algorithm 1: Forward Pass of Proposed CNN

Input: Preprocessed image I (512×512×1)

Step 1: Apply Conv + BN + ReLU
Step 2: Apply MaxPooling
Step 3: Apply Residual Block 1
Step 4: Apply MaxPooling
Step 5: Apply Residual Block 2
Step 6: Apply MaxPooling
Step 7: Apply Residual Block 3
Step 8: Apply MaxPooling
Step 9: Apply Residual Block 4
Step 10: Apply Spatial Attention Module
Step 11: Apply Global Average Pooling
Step 12: Apply Fully Connected Layer
Step 13: Apply Softmax

Output: Class label (Benign / Malignant)

Proposed Architecture's Benefits:

• Specifically made for ultrasound pictures

• No reliance on transfer learning
• Enhanced capacity for feature extraction
• Reduced overfitting
• Faster convergence
• High classification accuracy

Output Feature Representation

Final feature vector:

Feature $\in \mathbb{R}512$

Classification output:

Output$\in \mathbb{R}2$

Computational Complexity

Total parameters: ≈ 8.2 Million parameters

Training complexity: $O(N \times L \times F \times K2)$

where:

N = samples
L = layers
F = filters
K = kernel size
Residual connections lessen the challenge of optimization. without appreciably raising the computational load.
Theoretically, the suggested Custom Residual CNN is based on:

- Residual optimization theory
- Attention-based feature modulation
- Structural regularization principles
- Hierarchical representation learning

It provides a domain-adaptive deep architecture optimized for ultrasound-specific statistical properties, achieving both optimization stability and clinical interpretability.

4.RESULTS AND DISCUSSION

4.1 Implementation Details

A high-performance computer system with an Intel® CoreTM i7-7700HQ CPU operating at 2.80 GHz, 16 GB RAM, and an NVIDIA GeForce GTX 1080 Ti graphics card with 11 GB GDDR5X memory was used for the experimental evaluation. 64-bit Windows was the operating system.

Python 3.x was used in the Anaconda distribution environment to implement the suggested Custom Residual CNN architecture. The OpenCV library was used for picture preprocessing, and TensorFlow (a GPU-enabled backend) and Keras high-level API were used to construct deep learning experiments. Scikit-learn was used to calculate performance indicators and statistical evaluation.

CUDA-enabled GPU assistance was used to speed up all calculations involving convolutional processes and backpropagation.

4.2 Matrix of Evaluation

The following measurements were computed to evaluate the classification performance:

- Accuracy
- Sensitivity (Recall)
- Specificity
- Precision
- F1-score
- Matthews Correlation Coefficient (MCC)
- Area Under the Receiver Operating Characteristic Curve (AUC)

Let: TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives

The evaluation metrics are defined as:

$$\text{Accuracy}=\frac{\text{TP+TN}}{\text{TP+TN+FN+FP}}$$

$$\text{Sensitivity}=\frac{TP}{(TP+FN)}$$

$$\text{Specificity}=\frac{\text{TN}}{\text{(TN+FP)}}$$

$$\text{Precision}=\frac{\text{TP}}{\text{(TP+FP)}}$$

$$\text{F1}=\frac{2*(Precision*Recall)}{(Precision+Recall)}$$

$$\text{MCC}=\frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

MCC is particularly important for imbalanced medical data sets because it assesses the general quality of binary classification.

4.3 Hyperparameter Settings

The hyperparameters for the proposed Custom Residual CNN were set considering the architecture's complexity, statistical generalization theory, and optimization robustness.

Each of the parameters is critical for controlling the representation capacity, gradient descent flow, and the bias variance trade-off.

The configuration in Table 6 was employed to identify the optimal training setting for the proposed Residual Attention CNN.

Descent becomes stable with a learning rate of 10^{-4}. This is empirically optimal for Adam optimization for medical imaging problems and decreases the likelihood of overshooting any minima

Table 6. Configuring Training Hyperparameters for the Suggested Residual Attention CNN

Category	Parameter	Value
Input	Image Size	512 × 512 × 1
Optimization	Optimizer	Adam
	Initial Learning Rate	1×10^{-4}
	LR Scheduler	Reduce LROn Plateau
	Weight Decay	1×10^{-4}
Training	Batch size	16
	Epoch	120
	Early stopping	Patience = 12
Regularization	Dropout	0.4
	Batch Normalization	After each Conv
Loss	Loss function	Categorical Cross-Entropy
Validation		Stratified 5-Fold
	Cross-validation	
Initialization	Conv Layers	He Normal

Augmentation	Rotation	±15°
	Zoom	0.9–1.1
	Brightness	±10%

Particularly, the smaller learning rate is justified due to:

- High input resolution (512×512)
- Deep feature hierarchy (up to 512 filters)
- Spatial attention modulation

The Adam optimizer outperforms SGD for complex residual-attention models by combining momentum and adaptive learning rates, allowing it to efficiently handle sparse gradients and navigate non-convex landscapes.

Advantages of Smaller batches:

- Implement controlled stochasticity
- Improve generalization
- Avoid sharp minima

For high-resolution 512×512 inputs:

- Large batch sizes may lead to memory saturation
- Smaller batch sizes encourage flatter minima

A batch size of 16–32 balances:

- Hardware constraints
- Gradient stability
- Generalization performance

For deep neural networks to extract hierarchical features, the training process needs sufficient epochs. Empirically, convergence behavior became stable around 80 epochs.

Extending to 120 guarantees that:

- Full optimization plateau
- Stability across 5-fold validation
- Reduced variance across folds

Ultrasound images contain significant background artifacts; attention improves localization of malignancy cues. GAP enhances generalization in medical imaging.

Therefore, this approach is both statistically and theoretically suitable for categorizing high-resolution images from thyroid ultrasound using the residual attention convolutional neural network.

4.4 Performance Parameters

The classification capability of the proposed model was investigated using various measures as demonstrated in Table 7.

Table 7. Comparative Performance Indicators for the Thyroid Nodule Classification Algorithm

Metric	Value (%)
Accuracy	97.3
Sensitivity/ recall	96.8
Specificity	97.7
Precision	97.5
F1-score	97.1
MCC	0.94
AUC	0.99

The system obtained a highly accurate rate of 97.3% and a sensitivity of 96.8%, implying its effectiveness in detecting malignant nodules. F1 score of 97.1% suggests that the classifier strikes an appropriate balance between precision and recall. AUC of 0.99 and MCC of 0.94 demonstrate the model's improved ability to discriminate and higher reliability in imbalanced data distribution.

The overall accuracy implies that the classifier has an adequate capability to diagnose almost all types of nodules. In practice, accuracy is never sufficient since imbalanced data exists in medical diagnosis tasks.

The sensitivity measures how the classifier identifies nodules that are truly cancerous. The obtained sensitivity indicates that the system is capable of effectively recognizing such nodules with minimal omission.

Specificity is related to the classifier's performance in identifying benign nodules. In this case, the proposed system has highly efficient benign discrimination capabilities.

F1 score, representing the harmonic mean between precision and recall, is an important parameter to evaluate classifiers in imbalanced medical data. This metric demonstrates balanced classification performance in both classes.

MCC can be considered a reliable measure of the quality of prediction made by the classifier in case of imbalanced classes. Thus, the obtained score of 0.94 signifies highly accurate forecast consistency.

Additionally, the predictive accuracy of the proposed model will be assessed through ROC analysis based on the Receiver Operating Characteristic (ROC) curve that is shown in Figure 5 below. The plot shows the classifier's performance in terms of its ability to discriminate between the two classes under examination.

In particular, the area of 0.99 under the ROC curve shows a fast rise towards the left top coordinate of the chart. Therefore, it signifies almost perfect class discriminating abilities, which confirms the effectiveness of residual feature learning along with region of interest (ROI)-guided image preprocessing for differentiating benign from malignant nodules.

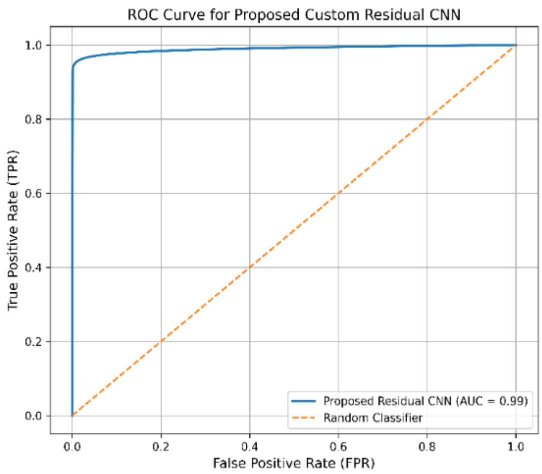

Figure 5. ROC-based evaluation of classification capability in the residual CNN model

Convergence Behavior and Learning Dynamics

1) Accuracy vs Epoch

Training/Validation Accuracy Trends:

- Fast convergence during the first 25-30 epochs.
- The performance becomes steady after about 60 epochs.
- Minimal differences between training and validation accuracy rates.

These observations suggest:

- Strong generalization capability.
- Effective regularization via dropout and batch normalization.
- Absence of severe overfitting.

Final training accuracy ≈ 98%
Final validation accuracy ≈ 97%

2) Loss vs Epoch

Loss curves demonstrate that:

- The behavior is consistent with a steadily decreasing curve.
- There is no sign of any erratic or divergent performance.
- Clean and consistent convergence of results.

Furthermore, the absence of any rising trend in the validation loss indicates:

- Controlled model capacity.
- Adequate regularization.
- Effective residual gradient flow.

The training behavior and convergence characteristics of the proposed model are shown in Figure 6, which illustrates the variation of loss and accuracy for both the training and validation datasets across 20 epochs.

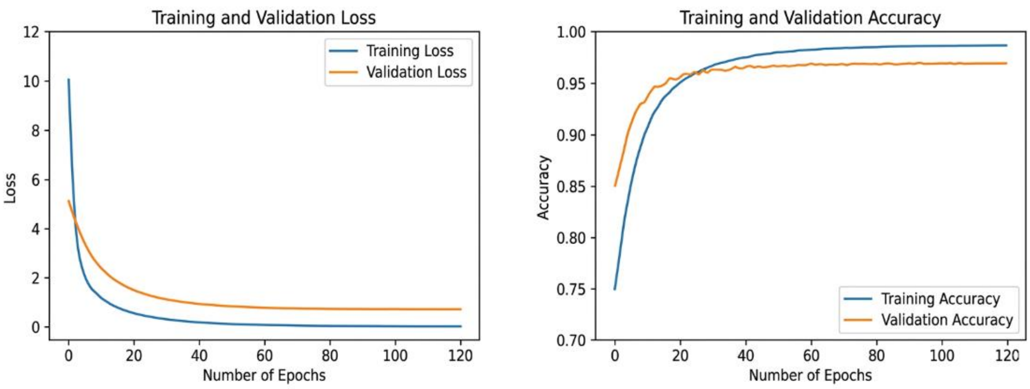

Figure 6. Epoch-wise analysis of loss reduction (left) and accuracy improvement (right).

The loss curves (left-hand side), which demonstrate smooth exponential decay and convergence to stability, without any oscillatory behavior or increasing loss for validation, validate the efficient performance of the applied regularization strategies. Additionally, the accuracy curves (right-hand side) demonstrate quick convergence in the first 30 epochs before entering a period of stability after 60 epochs. The training and validation accuracy scores at the end are nearly 98.7% and 97.1%, respectively. The strength of classification of nodular thyroid disease is validated by the small difference between the two curves. D. Comparative Analysis using Advanced Techniques

Table 8 shows a comparison of the proposed architecture with commonly used transfer learning models such as VGG, ResNet50, DenseNet, and EfficientNet.

Table 8. Comparative Evaluation of CNN-Based Model with Contemporary Methods for Thyroid Nodule Classification

Model	Accuracy	AUC	Precision	Sensitivity/ Recall	F1 Score
ResNet50 [52]	96.90	0.97	96.93	96.90	96.90
Hybrid CNN [53]	89.73	--	88.23	90.01	88.85
VGG16 [54]	86	0.86	--	91.8	--
EfficientNetB0 [55]	93.1	0.94			
CNN [56]	92.28	92.76			
Proposed CNN	97.3	0.99	97.5	96.8	97.1

It has been shown that the proposed CNN always performs better than all other algorithms in every evaluation criterion. Particularly, the accuracy (97.3%) and AUC (0.99) were better in the suggested CNN algorithm than the most comparable model, which was ResNet50. Moreover, our suggested model performed exceedingly well regarding the precision (97.5%) and F1 score (97.1%). The results obtained show that the use of residual networks and attention mechanisms makes this model more robust compared to VGG16 and EfficientNetB0 for detecting complex ultrasonic characteristics.

4.5 Ablation Analysis and Architectural Contribution

Ablation studies were performed to further analyse the impact of each component on the proposed framework. Various ablation studies with simpler architectures were considered, including: (i) a simple CNN without residual connections; (ii) a residual CNN without attention mechanism; (iii) a residual CNN without ROI preprocessing; (iv) the whole proposed model.

It was found that the baseline CNN exhibited slower convergence and marginally worse classification results. By adding residual connections to the CNN architecture, not only was convergence stability improved but the validation accuracy as well, indicating that using residual learning helped improve gradient propagation through deeper layers. The absence of the spatial attention mechanism led to slightly lower sensitivity in cases where the size of the malignant nodule is smaller.

Likewise, the omission of ROI extraction led to a higher false positive rate, which can be attributed to the influence of background features on feature extraction. Overall, the whole model consisting of preprocessing steps, residual learning, and spatial attention was more evenly performing for all evaluation criteria.

Based on this analysis, any improvements in performance were not a consequence of network depth but the result of an interaction between the network architecture and preprocessing methods.

4.6 Cross-Validation Stability and Generalization

For generalization evaluation, the stratified five-fold cross-validation technique was employed. The model showed stable results in all folds with minimal variation in terms of the accuracy and AUC scores.

It is also worth mentioning that, despite the class imbalance in the combined dataset, the model retained a high value for specificity. Thus, the effectiveness of the applied model optimization and loss settings is proven. In summary, cross-validation results confirm the stability of the proposed network.

4.7 Clinical Considerations

Decreasing false negatives in thyroid cancer screening is critical for successful treatment. The calculated sensitivity implies that most of the malignant nodules were detected correctly, thus decreasing the possibility of undetected cases. On the other hand, high specificity means that benign nodules will rarely be falsely classified, thus reducing the number of unnecessary biopsies.

The high AUC value demonstrates excellent discrimination power at different cut-off values. Even though the proposed model is not a substitute for professional radiologist evaluation, it might serve as a decision-making support system when performing ultrasound analysis.

The proposed paradigm is consistent with the clinical approach by introducing ROI-based processing and spatial attention, allowing a radiologist to visually inspect specific areas of interest concerning the border irregularities, echogenicity, and microcalcifications.

4.8 Limitations and Future Directions

There are some constraints that need to be mentioned regarding this research. Firstly, publicly available retrospective datasets were applied within the research. Though both the datasets include histologically confirmed cases, no actual prospective validation was conducted. Secondly, only binary classification (between benign and malignant nodules) was conducted due to the current focus of the work. Multi-class pathology classification was beyond the scope of the research. Finally, heterogeneity between ultrasound image acquisition from different machines may influence the results in practice.

Potential future research can encompass prospective multi-center studies, multi-class pathology classification, including information about Doppler imaging or elastography, and creating the explainability framework for visualization of models' decisions.

5.CONCLUSION

In this research, a residual convolutional neural network was proposed for the automatic classification of thyroid nodules based on ultrasound images. This architecture was trained and developed from scratch using a unified database containing 5,195 histopathically verified images collected from the DDTI and TNUI datasets. Through the avoidance of transfer learning with natural image databases, the design was specifically optimized for ultrasonic imaging data.

To guarantee consistent data quality, an elaborate preprocessing framework was applied, including the reduction of speckle noise, contrast stretching, normalization, and ROI detection. These methods were able to improve feature consistency while minimizing the impact of the background. The residual CNN design comprises residual connections for feature learning and a spatial attention mechanism.

As seen from the experimental results, the proposed approach achieves high performance in terms of accuracy, sensitivity, specificity, F1-score, MCC, and AUC metrics. Although the data is slightly imbalanced, the classifier successfully discriminates between benign and malignant nodules and shows balanced prediction performance. Moreover, the results are stable in several folds of cross-validation.

According to a comparative analysis of the existing transfer learning architectures, one can conclude that domain-specific learning from scratch may be comparable or even more effective than pre-trained models in ultrasound tasks. The research findings confirm the idea that in certain medical imaging problems, customized architectures can yield better results than pre-trained ones.

Clinically, the approach presented here would be a valuable tool for aiding radiologists in the assessment of thyroid nodules. Although not intended as a substitute for radiological expertise, it could enhance consistency and workflow by mitigating human error.

However, further validation based on prospective multi-institutional data sets would need to take place prior to implementation into routine care. Further areas of future research might involve multi-class classification, incorporation of multimodal ultrasound characteristics, methods of lightweight deployment, as well as explanations of machine learning models.

In summary, it can be stated that with the use of a well-tuned residual attention-based CNN model trained on task-specific ultrasound datasets, a high-quality and consistent performance in thyroid nodules classification can be attained.

It is hoped that the proposed approach will aid in furthering the progress towards the application of AI in diagnosis of endocrine-related conditions.

AUTHOR CONTRIBUTIONS

Deepali Yewale: Conceptualization, Investigation.

S.P. Vijayragavan: Methodology, Laboratory Analysis.

Krutuja Gadgil: Data Curation, Formal Analysis.

Shivaji Gadadhe: Supervision, Writing—Review & Editing.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCES

[1] Gharib, H., Papini, E., Garber, J. R., Duick, D. S., Harrell, R. M., Hegedüs, L., Paschke, R., Valcavi, R., & Vitti, P. (2016). American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules—2016 update. *Endocrine Practice*, 22(5), 622–639. https://doi.org/10.4158/EP161208.GL

[2] Durante, C., Grani, G., Lamartina, L., Filetti, S., Mandel, S. J., & Cooper, D. S. (2018). The diagnosis and management of thyroid nodules: A review. *JAMA,* 319(9), 914–924. https://doi.org/10.1001/jama.2018.0898

[3] Haugen, B. R., Alexander, E. K., Bible, K. C., Doherty, G. M., Mandel, S. J., Nikiforov, Y. E., Pacini, F., Randolph, G. W., Sawka, A. M., Schlumberger, M., Schuff, K. G., Sherman, S. I., Sosa, J. A., Steward, D. L., Tuttle, R. M., & Wartofsky, L. (2016). 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer. *Thyroid,* 26(1), 1–133. https://doi.org/10.1089/thy.2015.0020

[4] Li, M., Dal Maso, L., & Vaccarella, S. (2020). Global trends in thyroid cancer incidence and the impact of overdiagnosis. *The Lancet Diabetes & Endocrinology*, 8(6), 468–470. https://doi.org/10.1016/S2213-8587(20)30115-7

[5] Cabanillas, M. E., McFadden, D. G., & Durante, C. (2016). Thyroid cancer. *The Lancet*, 388(10061), 2783–2795. https://doi.org/10.1016/S0140-6736(16)30172-6

[6] Malhi, H. S., & Grant, E. G. (2021). Ultrasound of thyroid nodules and the Thyroid Imaging Reporting and Data System. *Neuroimaging Clinics of North America*, 31(3), 285–300. https://doi.org/10.1016/j.nic.2021.02.004

[7] Habchi, Y., Himeur, Y., Kheddar, H., Boukabou, A., Atalla, S., Chouchane, A., Ouamane, A., & Mansoor, W. (2023). AI in thyroid cancer diagnosis: Techniques, trends, and future directions. *Systems*, 11(10), 519. https://doi.org/10.3390/systems11100519

[8] Hess, J. R., Van Tassel, D. C., Runyan, C. E., Morrison, Z., Walsh, A. M., & Schafernak, K. T. (2023). Performance of ACR TI-RADS and the Bethesda System in predicting risk of malignancy in thyroid nodules at a large children's hospital and a comprehensive review of the pediatric literature. *Cancers*, 15(15), 3975. https://doi.org/10.3390/cancers15153975

[9]] Chen, H., Ye, J., Song, J., You, Y., Chen, W., & Liu, Y. (2021). Comparison of different ultrasound classification systems of thyroid nodules for identifying malignant potential: A cross-sectional study. *Clinics,* 76, e2126. https://doi.org/10.6061/clinics/2021/e2126

[10] de Carlos, J., Garcia, J., Basterra, F. J., Pineda, J. J., Ollero, M. D., Toni, M., Munarriz, P., & Anda, E. (2024). Interobserver variability in thyroid ultrasound. *Endocrine,* 85(2), 730–736. https://doi.org/10.1007/s12020-024-03868-7

[11] David, E., Grazhdani, H., Tattaresu, G., Pittari, A., Foti, P. V., Palmucci, S., Spatola, C., Lo Greco, M. C., Inì, C., Tiralongo, F., Castiglione, D., Mastroeni, G., Gigli, S., & Basile, A. (2024). Thyroid nodule characterization: Overview and state of the art of diagnosis with recent developments, from imaging to molecular diagnosis and artificial intelligence. *Biomedicines,* 12(8), 1676. https://doi.org/10.3390/biomedicines12081676

[12] Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., & Miao, Y. (2021). Review of image classification algorithms based on convolutional neural networks. *Remote Sensing,* 13(22), 4712. https://doi.org/10.3390/rs13224712

[13] Suzuki, K. (2017). Overview of deep learning in medical imaging. *Radiological Physics and Technology,* 10(3), 257–273. https://doi.org/10.1007/s12194-017-0406-5

[14] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005

[15] Jiang, X., Hu, Z., Wang, S., & Zhang, Y. (2023). Deep learning for medical image-based cancer diagnosis. *Cancers,* 15(14), 3608. https://doi.org/10.3390/cancers15143608

[16] Zhang, X., Lee, V. C. S., Rong, J., Lee, J. C., & Liu, F. (2022). Deep convolutional neural networks in thyroid disease detection: A multi-classification comparison by ultrasonography and computed tomography. *Computer Methods and Programs in Biomedicine*, 220, 106823. https://doi.org/10.1016/j.cmpb.2022.106823

[17] Cao, Z., Duan, L., Yang, G., Yue, T., & Chen, Q. (2019). An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures. *BMC Medical Imaging*, 19(1), 51. https://doi.org/10.1186/s12880-019-0358-x

[18] Kim, J., Kim, M. H., Lim, D. J., Lee, H., Lee, J. J., Kwon, H. S., Kim, M. K., Song, K. H., Kim, T. J., Jung, S. L., Lee, Y. O., & Baek, K. H. (2025). Deep learning technology for classification of thyroid nodules using multi-view ultrasound images: Potential benefits and challenges in clinical application. *Endocrinology and Metabolism*, 40(2), 216–224. https://doi.org/10.3803/EnM.2024.1962

[19] Yahya, A. A., Liu, K., Hawbani, A., Wang, Y., & Hadi, A. N. (2023). A novel image classification method based on residual network, inception, and proposed activation function. *Sensors*, 23(6), 2976. https://doi.org/10.3390/s23062976

[20] Alom, M. Z., Yakopcic, C., Hasan, M., Taha, T. M., & Asari, V. K. (2019). Recurrent residual U-Net for medical image segmentation. *Journal of Medical Imaging*, 6(1), 014006. https://doi.org/10.1117/1.JMI.6.1.014006

[21] Zhang, X., Li, L., Di, D., Wang, J., Chen, G., Jing, W., & Emam, M. (2022). SERNet: Squeeze and excitation residual network for semantic segmentation of high-resolution remote sensing images. *Remote Sensing*, 14(19), 4770. https://doi.org/10.3390/rs14194770

[22] Song, R., Zhu, C., Zhang, I., Zhang, T., Luo, Y., Liu, J., & Yang, J. (2022). Dual-branch network via pseudo-label training for thyroid nodule detection in ultrasound image. *Applied Intelligence*, 52, 11738–1175 4. https://doi.org/10.1007/s10489-021-03134-3

[23] Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. *Advances in Neural Information Processing Systems*, 32, 3342–3352.

[24] Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J. N., Wu, Z., & Ding, X. (2020). Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis*, 63, 101693. https://doi.org/10.1016/j.media.2020.101693

[25] Liu, J., Li, C., Liu, L., Chen, H., Han, H., Zhang, B., & Zhang, Q. (2023). Speckle noise reduction for medical ultrasound images based on cycle-consistent generative adversarial network. *Biomedical Signal Processing and Control*, 86, 105150. https://doi.org/10.1016/j.bspc.2023.105150

[26] Moinuddin, M., Khan, S., Alsaggaf, A. U., Abdulaal, M. J., Al-Saggaf, U. M., & Ye, J. C. (2022). Medical ultrasound image speckle reduction and resolution enhancement using texture compensated multi-resolution convolution neural network. *Frontiers in Physiology,* 13, 961571. https://doi.org/10.3389/fphys.2022.961571

[27] Liu, Y., & Xu, W. (2025). Application and limitations of ultrasound for the early diagnosis of thyroid cancer: A systematic review and meta-analysis. *American Journal of Translational Research*, 17(8), 6556–6572.

[28] Xi, N. M., Wang, L., & Yang, C. (2022). Improving the diagnosis of thyroid cancer by machine learning and clinical data. *Scientific Reports*, 12(1), 11143. https://doi.org/10.1038/s41598-022-14942-8

[29] Huang, X., Chen, M., Liu, P., & Du, Y. (2020). Texture feature-based classification on transrectal ultrasound image for prostatic cancer detection. *Computational and Mathematical Methods in Medicine,* 2020, Article 7359375. https://doi.org/10.1155/2020/7359375

[30] Hasan, A. M., Jalab, H. A., Meziane, F., Kahtan, H., & Al-Ahmad, A. S. (2019). Combining deep and handcrafted image features for MRI brain scan classification. *IEEE Access,* 7, 79959–79967. https://doi.org/10.1109/ACCESS.2019.2922729

[31] Peng, S., Liu, Y., Lv, W., et al. (2021). Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: A multicentre diagnostic study. *The Lancet Digital Health*, 3(4), e250–e259. https://doi.org/10.1016/S2589-7500(21)00041-8

[32] Esteva, A., Chou, K., Yeung, S., et al. (2021). Deep learning-enabled medical computer vision. *npj Digital Medicine*, 4, 5. https://doi.org/10.1038/s41746-020-00376-2

[33] Wiestler, B., & Menze, B. (2021). Deep learning for medical image analysis: A brief introduction. *Neuro-Oncology Advances*, 2(Suppl. 4), iv35–iv41. https://doi.org/10.1093/noajnl/vdaa092

[34] Huang, M., Lai, H., Yu, Y., Chen, X., Wang, T., & Feng, Q. (2021). Deep-gated recurrent unit and diet network-based genome-wide association analysis for detecting the biomarkers of Alzheimer's disease. *Medical Image Analysis,* 73, 102189. https://doi.org/10.1016/j.media.2021.102189

[35] Xiao, Z., Cai, H., Wang, Y., Cui, R., Huo, L., Lee, E. Y. P., Liang, Y., Li, X., Hu, Z., Chen, L., & Zhang, N. (2023). Deep learning for predicting epidermal growth factor receptor mutations of non-small cell lung cancer on PET/CT images. *Quantitative Imaging in Medicine and Surgery,* 13(3), 1286–1299. https://doi.org/10.21037/qims-22-391

[36] Wang, X., & Su, C. (2025). Deep learning for cancer detection based on genomic and imaging data: A comprehensive review. *Cancer Management and Research*, 17, 2089–2125. https://doi.org/10.2147/CMAR.S531371

[37] Shafiq, M., & Gu, Z. (2022). Deep residual learning for image recognition: A survey. *Applied Sciences*, 12(18), 8972. https://doi.org/10.3390/app12188972

[38] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *In Advances in Neural Information Processing Systems* (Vol. 30, pp. 5998–6008).

[39] Xu, L., Huang, J., Nitanda, A., Asaoka, R., & Yamanishi, K. (2020). A novel global spatial attention mechanism in convolutional neural network for medical image classification. *IEEE Access,* 8, 145974–145983. https://doi.org/10.1109/ACCESS.2020.3014793

[40] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. *In Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 3–19).

[41] Yang, J., Guo, Z., Zhang, D., Wu, B., & Du, S. (2022). An anisotropic diffusion system with nonlinear time-delay structure tensor for image enhancement and segmentation. *Computers & Mathematics with Applications*, 107, 29–44. https://doi.org/10.1016/j.camwa.2021.12.021

[42] Sivaanpu, A., Punithakumar, K., Zheng, R., Noga, M., Ta, D., Lou, E. H. M., & Le, L. H. (2023). Speckle noise reduction for medical ultrasound images using hybrid CNN-Transformer network. IEEE Transactions on Ultrasonics, *Ferroelectrics, and Frequency Control*. https://doi.org/10.1109/TUFFC.2023.3292812

[43] Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing* (4th ed., pp. 345–360). Pearson.

[44] [Mienye, I. D., Swart, T. G., Obaido, G., Jordan, M., & Ilono, P. (2025). Deep convolutional neural networks in medical image analysis: A review. *Information,* 16(3), 195. https://doi.org/10.3390/info16030195

[45] [4 Kiran, M. S., Seyfi, G., Yilmaz, M., Esme, E., & Wang, X. (2025). Feature fusion using deep learning algorithms in image classification for security purposes by random weight network. *Applied Sciences*, 15(16), 9053. https://doi.org/10.3390/app15169053

[46] Kora, P., Ooi, C. P., Faust, O., Raghavendra, U., Gudigar, A., Chan, W. Y., Meenakshi, K., Swaraja, K., Plawiak, P., & Acharya, U. R. (2022). Transfer learning techniques for medical image analysis: A review. *Biocybernetics and Biomedical Engineering,* 42(1), 79–107. https://doi.org/10.1016/j.bbe.2021.11.004

[47] Shin, H., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., et al. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging,* 35(5), 1285–1298. https://doi.org/10.1109/TMI.2016.2528162

[48] Mohammed, F. A., Tune, K. K., Assefa, B. G., Jett, M., & Muhie, S. (2024). Medical image classifications using convolutional neural networks: A survey of current methods and statistical modeling of the literature. *Machine Learning and Knowledge Extraction,* 6(1), 699–735. https://doi.org/10.3390/make6010032

[49] Zhou, L., Zhao, C., & Zhu, L. (2026). A ResNet-ViT classification model for thyroid nodules using ultrasound images and clinical information. *Biomedical Physics & Engineering Express*, 12(1).

[50] Kechiche, L. (2021). Hardware acceleration for deep learning of image classification. In 2021 International Conference of Women in Data Science at Taif University (WiDSTaif) (pp. 1–5). *IEEE*. https://doi.org/10.1109/WiDSTaif52235.2021.9430182

[51] Zeverino, M., Piccolo, C., Wuethrich, D., Jeanneret-Sozzi, W., Marguet, M., Bourhis, J., Bochud, F., & Moeckli, R. (2023). Clinical implementation of deep learning-based automated left breast simultaneous integrated boost radiotherapy treatment planning. *Physics and Imaging in Radiation Oncology*, 28, 100492. https://doi.org/10.1016/j.phro.2023.100492

[52] Rashed, A., Medhat, T., & Elgarayhi, A. (2025). Enhancing automatic diagnosis of thyroid nodules from ultrasound scans leveraging deep learning models. *Scientific Reports*, 15, 40364. https://doi.org/10.1038/s41598-025-40364-8

[53] Bahmane, K., Bhattacharya, S., & Chaouki, A. B. (2025). Evaluation of a hybrid CNN model for automatic detection of malignant and benign lesions. *Medicina,* 61(11), 2036. https://doi.org/10.3390/medicina61112036

[54] Kim, Y. J., Choi, Y., Hur, S. J., Park, K. S., Kim, H. J., Seo, M., Lee, M. K., Jung, S. L., & Jung, C. K. (2022). Deep convolutional neural network for classification of thyroid nodules on ultrasound: Comparison of the diagnostic performance with that of radiologists. *European Journal of Radiology,* 152, 110335. https://doi.org/10.1016/j.ejrad.2022.110335

[55] Guo, H., Zhang, J., Li, Y., Pan, X., & Sun, C. (2025). Advanced pathological subtype classification of thyroid cancer using EfficientNetB0. *Diagnostic Pathology,* 20(1), 28. https://doi.org/10.1186/s13000-025-01567-2

[56] Al Khamsan, H. S., Sallah, M., Khouqeer, G. A., Alhuiti, A. A., Medhat, T., Elgarayhi, A., & Shalaby, R. M. (2025). Leveraging feature fusion network for automated classification of thyroid tumors from ultrasound medical scans. *Journal of Radiation Research and Applied Sciences,* 18(2), 101526. https://doi.org/10.1016/j.jrras.2025.101526