

RESEARCH

Open Access



# The value of a neural network based on multi-scale feature fusion to ultrasound images for the differentiation in thyroid follicular neoplasms

Weiwei Chen<sup>1†</sup>, Xuejun Ni<sup>1†</sup>, Cheng Qian<sup>1</sup>, Lei Yang<sup>1</sup>, Zheng Zhang<sup>2</sup>, Mengdan Li<sup>1</sup>, Fanlei Kong<sup>3</sup>, Mengqin Huang<sup>4</sup>, Maosheng He<sup>5\*</sup> and Yifei Yin<sup>1\*</sup>

## Abstract

**Objective** The objective of this research was to create a deep learning network that utilizes multiscale images for the classification of follicular thyroid carcinoma (FTC) and follicular thyroid adenoma (FTA) through preoperative US.

**Methods** This retrospective study involved the collection of ultrasound images from 279 patients at two tertiary level hospitals. To address the issue of false positives caused by small nodules, we introduced a multi-rescale fusion network (MRF-Net). Four different deep learning models, namely MobileNet V3, ResNet50, DenseNet121 and MRF-Net, were studied based on the feature information extracted from ultrasound images. The performance of each model was evaluated using various metrics, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, F1 value, receiver operating curve (ROC), area under the curve (AUC), decision curve analysis (DCA), and confusion matrix.

**Results** Out of the total nodules examined, 193 were identified as FTA and 86 were confirmed as FTC. Among the deep learning models evaluated, MRF-Net exhibited the highest accuracy and area under the curve (AUC) with values of 85.3% and 84.8%, respectively. Additionally, MRF-Net demonstrated superior sensitivity and specificity compared to other models. Notably, MRF-Net achieved an impressive F1 value of 83.08%. The curve of DCA revealed that MRF-Net consistently outperformed the other models, yielding higher net benefits across various decision thresholds.

**Conclusion** The utilization of MRF-Net enables more precise discrimination between benign and malignant thyroid follicular tumors utilizing preoperative US.

**Keywords** Follicular thyroid carcinoma, Ultrasound images

<sup>†</sup>Weiwei Chen and Xuejun Ni contributed equally to this work and share first authorship.

\*Correspondence:

Maosheng He

119484526@qq.com

Yifei Yin

yinyifei4311@163.com

<sup>1</sup>Department of Medical Ultrasound, Affiliated Hospital of Nantong University, 226001 Nantong, P.R. China

<sup>2</sup>Department of Medical Ultrasound, Affiliated Hospital of Jiangsu University, 212000 Zhenjiang, P.R. China

<sup>3</sup>Department of Medical Ultrasound, Affiliated Hangzhou First People's Hospital, 310006 Hangzhou, P. R. China

<sup>4</sup>Shanghai Soundwise Information Technology Co., Ltd, 200000 Shanghai, P.R. China

<sup>5</sup>Department of Medical Ultrasound, Xuzhou Central Hospital, 221000 Xuzhou, P.R. China



## Introduction

In clinical practice, thyroid follicular neoplasms are categorized primarily as follicular thyroid adenoma (FTA) or follicular thyroid carcinoma (FTC). FTC, the second most prevalent type of differentiated thyroid cancer, accounts for approximately 10-20% of all thyroid cancers [1]. Patients typically exhibit distant metastasis through hematogenous dissemination, with the lungs and bones being the most commonly affected sites [2, 3]. According to the published literature, the annual incidence rate of follicular thyroid cancer in many countries, including the United States, is approximately 0.8 per 100,000 people, with a male-to-female ratio of 1:2.5 [4]. Although the incidence of FTC is lower than that of papillary thyroid carcinoma (PTC), FTC is associated with a greater rate of distant metastasis and mortality [5–7].

Currently, the main diagnostic approaches for thyroid follicular neoplasms include conventional ultrasound, fine-needle aspiration cytology (FNAC), diagnostic surgical excision, CT, and MRI. Conventional ultrasound examination offers advantages such as non-invasiveness, lack of radiation exposure, multi-angle imaging, and affordability. It is the most commonly employed imaging modality for diagnosing thyroid diseases in clinical practice, surpassing CT, MRI, and other methods. However, distinguishing between FTA and FTC based on nodule size, shape, echogenicity, margin characteristics, calcification patterns, and vascularity is challenging, as both entities exhibit significant overlap in ultrasonic images [8–11]. Moreover, FNAC, which is often limited by the site of puncture, faces challenges in providing a definitive pathological diagnosis due to the morphological similarities between FTA and FTC [12]. The differential diagnosis between these two tumors depends on determining the presence of vascular or extrathyroidal tissue invasion and lymph node or distant metastasis. Although a standardized clinical thyroid imaging data reporting system known as the thyroid imaging reporting and data system (TI-RADS) exists, the interpretation of ultrasound images still exhibits some subjectivity among different ultrasound practitioners, leading to potential misdiagnosis or missed diagnosis [13]. The guidelines established by the American Thyroid Association recommend diagnostic surgical excision as a well-established standard for managing follicular neoplasms or suspicious cases [14]. Pathology after surgical resection often reveals that, among patients initially diagnosed with follicular tumors, up to 80% have follicular adenomas. This finding indicates that a significant proportion of patients, despite having a benign condition, undergo diagnostic thyroid lobectomy. Therefore, it is crucial to preoperatively differentiate between these two conditions to avoid such overtreatment in patients with benign disease. Moreover,

this distinction can reduce the misdiagnosis rate in malignant patients, ultimately improving survival rates.

In recent years, artificial intelligence has rapidly advanced as an innovative tool and has found widespread application across various medical specialties [15]. In the realm of ultrasound diagnostics, researchers can employ computer algorithms to automatically extract image features that may be imperceptible to the human eye or that are challenging for ultrasound practitioners to articulate verbally. These extracted features can be translated into reliable data, providing insights into the underlying pathophysiology and offering valuable information for disease diagnosis and prognosis [16–18]. Yadav et al. compared the performance of 64 despeckling filter algorithms for analyzing ultrasound images of thyroid nodules by calculating the structure and preserving the edge. The results showed that the fast bilateral filter and edge-preserving smoothing filters had the best performance in preserving image structures, such as the edges and margins of benign and malignant thyroid tumors [19]. While numerous studies have focused on the detection and automated diagnosis of thyroid nodules to distinguish between benign and malignant cases, there is still a paucity of research on distinguishing benign and malignant thyroid follicular tumors [20–22].

To the best of our knowledge, Shin et al. investigated the use of machine learning by employing manual nodule segmentation to screen and extract features, subsequently incorporating them into support vector machine (SVM) and artificial neural network (ANN) classifier models [23]. The results demonstrated an accuracy of 74.1% for the ANN model and 69.0% for the SVM model. Seo et al., on the other hand, employed deep learning techniques to identify ultrasound images of thyroid follicular tumors [24]. They proposed the use of convolutional neural networks (CNNs) to detect specific morphological features within the border regions of thyroid follicular tumors, leveraging image selection subsampling and datasets provided by the tumor boundaries. Alabrak et al. utilized a CNN model to extract 886 pathological images from 43 patients' Bethesda Class IV nodules (527 FTC images, 359 FTA images) [25]. Among these, 708 images were used as the training set, and 108 images were used as the test set. The input images were color images with a resolution of 256×256 pixels. The convolutional and pooling layers automatically extract image features, ultimately providing a binary classification for FTA and FTC. The results demonstrated an accuracy of 78.0%, a sensitivity of 88.4%, a specificity of 64.0%, and an AUC of 0.870 in distinguishing and diagnosing FTA and FTC. Deng et al. employed four deep neural network (DNN) models (Resnet50, Densenet121, EfficientNet, and Resnext50), while Chan et al. compared three convolutional neural network (CNN) models (InceptionV3,

ResNe101, and VGG19) and evaluated the diagnostic efficacy of two clinical experts with over 20 years of experience [26, 27]. The outcomes of these studies collectively indicate that CNN models can serve as auxiliary diagnostic tools for preoperative differentiation between FTA and FTC. However, both of these studies possess certain limitations that necessitate further improvement. For instance, Shin and Seo et al. manually extracted sampled images along the lesion contour, potentially overlooking certain features during the extraction process. Thyroid experts emphasize the importance of considering the lesion envelope, specifically the marginal zone of the segmented region, as a critical factor for distinguishing between FTA and FTC [28, 29]. Notably, recent advancements in deep learning have revolutionized cancer diagnosis, with multi-scale fusion neural networks gaining considerable attention due to their exceptional feature extraction capabilities [30–33].

In our study, we devised a novel neural network model called the multi-scale fusion network (MRF-Net), which incorporates the fusion of multi-scale features. The primary contribution of our research lies in integrating ultrasound findings with a neural network model that incorporates multi-scale fusion. This combination aims to increase the accuracy of preoperative diagnosis of follicular thyroid tumors. The significance of our study extends to the realm of personalized treatment for patients and the rational allocation of medical resources. By doing so, we can avoid the overtreatment of benign cases and reduce the rate of misdiagnosis of malignant cases, ultimately improving overall survival rates. Herein, we provide detailed elucidations of the notable advantages offered by this model:

- The application of multi-scale image processing methods significantly enhances the structural characteristics of images, resulting in improved accuracy and efficiency in image classification.
- The incorporation of the Rescale Enhancement Image Module (REI) enables the suppression of noise interference along with the simultaneous enhancement of boundary features. This feature facilitates the identification and differentiation of various elements within the image.

## Materials and methods

### Study population

Patient data for this retrospective study were collected from two hospitals from August 2009 to June 2022. The data consisted of patients who underwent surgical pathology for thyroid follicular tumors. The collected information included patient demographic information, 2D ultrasound images, immunohistochemical results

from pathology sections, and other relevant data. Ethical approval was obtained from the institutional review board, and the requirement for informed consent was waived for the two study populations.

All patients included in the study met the following selection criteria: (1) underwent their initial surgery and received a pathological diagnosis of either FTC or FTA; (2) underwent a preoperative ultrasound examination at one of the two hospitals, and the ultrasound images obtained were of sufficient quality. Some patients were excluded from the study due to missing preoperative ultrasound image information, undetected thyroid nodules during the preoperative ultrasound examination, or inadequate image quality. Ultimately, a total of 279 patients (Fig. 1) met the eligibility criteria and were included in the study; 193 patients were diagnosed with FTA, and 86 patients were diagnosed with FTC.

### Image collection and preprocessing

Experienced radiologists acquired preoperative two-dimensional images of all thyroid nodules using an ultrasound image archive workstation. Subsequently, experienced ultrasound physicians reviewed the images and retrospectively selected the maximum transverse and longitudinal plane images for each nodule. Whenever uncertainties arose regarding the images, a senior radiologist with more than 20 years of experience was consulted for further evaluation. Finally, all the collected images were cropped to remove extraneous information, ensuring that only the nodule remained at the center of the image (Fig. 2).

### The proposed multi-scaled image learning method

In this study, we propose a deep learning network that utilizes multi-scale images for the classification of FTC and FTA. First, it is crucial to review the principles of multi-scale image analysis, as it plays a significant role in our current task. Multi-scale image analysis has the ability to enhance image features effectively, leading to improved classification accuracy.

#### *Design of multi-scale image processing blocks*

Multi-scale image processing involves decomposing and reconstructing an original image into different scales to extract features at various scales, thereby enabling comprehensive and accurate image analysis and processing. By decomposing the original image, multi-scale image processing captures information about different scales, facilitating the representation of features such as edges and textures. Various decomposition methods can be employed, including pyramid decomposition and wavelet decomposition.

The key advantages of multi-scale image processing technology are as follows: First, it enhances the reliability

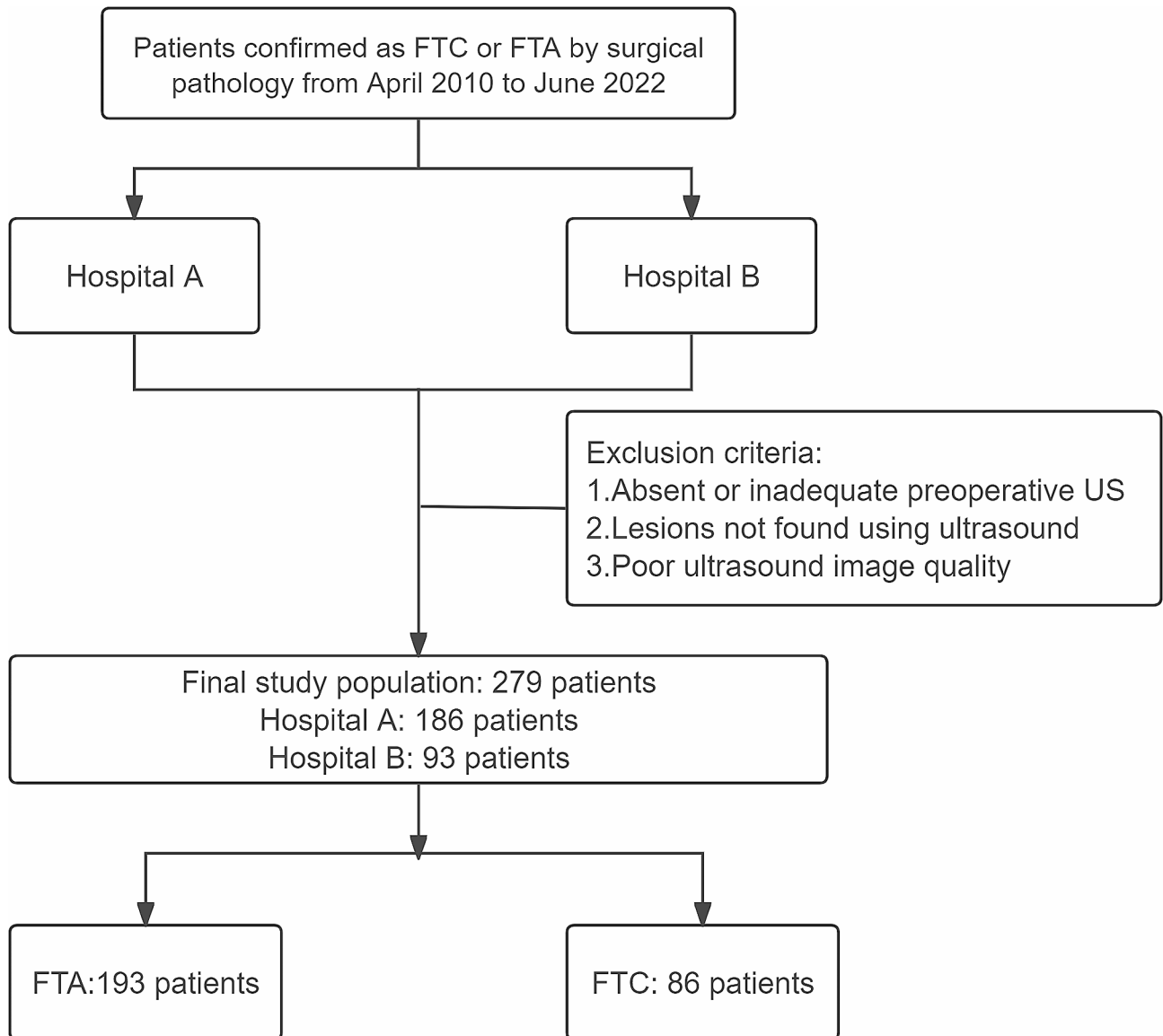


Fig. 1 A flow chart of data collection

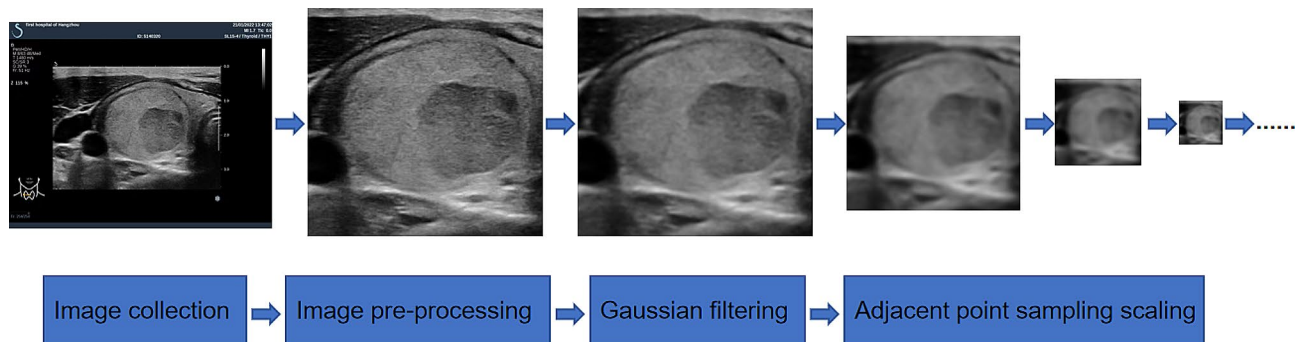


Fig. 2 A flow chart of image processing

and robustness of image features by capturing features and details at different scales through multi-scale processing. This contributes to improved reliability in feature extraction. Second, it enhances algorithm robustness by enabling adaptability to changes at different scales, thereby enhancing the algorithm’s generalization ability. Third, the impact of noise is reduced by decomposing the original image and eliminating high-frequency components, leading to reduced noise interference and improved algorithm accuracy. Finally, the algorithm efficiency is enhanced by allowing calculations to be performed at different scales.

To optimize the performance of our MRF-Net, we carefully selected the following hyperparameters for our CNN models. The learning rate was initially set to 0.001, with a dynamic adjustment mechanism that reduces the rate by a factor of 0.1 if the validation loss plateaus for more than 10 epochs. We used a batch size of 32 to balance the computational efficiency and model performance. The models were trained using the Adam optimizer due to its adaptability in adjusting learning rates for different parameters. The dropout rate was set to 0.5 in fully connected layers to prevent overfitting. Furthermore, we applied L2 regularization with a lambda value of 0.001 to penalize large weights in the network. The convolutional layers used ReLU (Rectified Linear

Unit) activation functions for introducing non-linearity, while the final output layer utilized a softmax activation function for multi-class classification. The models were trained for a total of 200 epochs, or until no significant improvement in validation accuracy was observed.

In this paper, we propose the use of the REI to enhance the head layer of the network architecture. Utilizing a Gaussian pyramid structure, similar to that in Fig. 3, the input image undergoes multi-scale processing to improve its representation.

The specific method is as follows:

1. Gaussian filtering was applied to the image, resulting in a Gaussian distribution (Fig. 2).
2. The difference between the Gaussian image and the input image is computed, and the negative values are subsequently obtained to yield the difference images.
3. Average pooling is conducted on the gaussian image, and its dimensions are reduced by half.
4. Utilize max absolute value pooling on the difference image, reducing its dimensions by half.
5. The reduced gaussian image and the difference image are combined to generate a new image.
6. The features of the new image are enhanced through structural part enhancement.
7. The enhanced image is output as the final result.

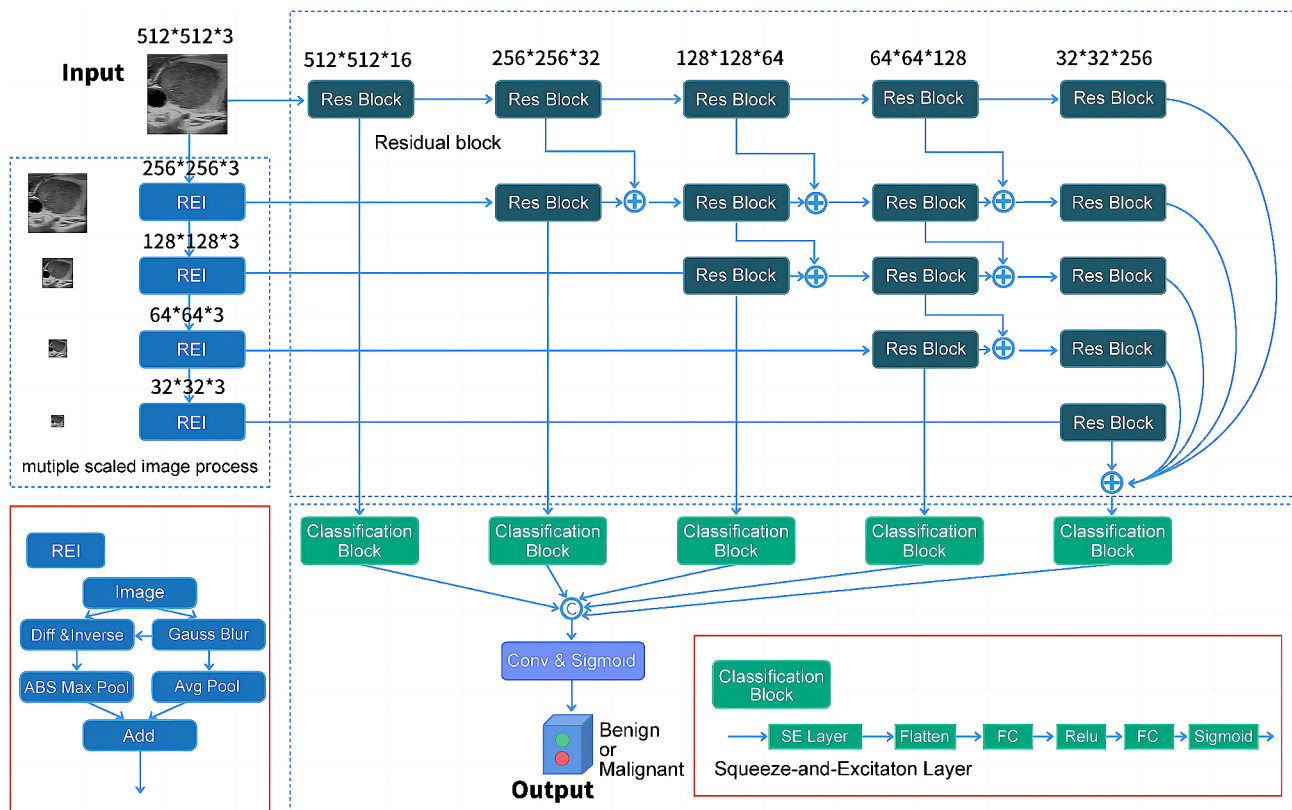


Fig. 3 Diagram of overall workflow of model training and validation

Compared with traditional multi-scale structures, there are two merits:

1. The difference image undergoes absolute maximum pooling, which effectively preserves the essential structure while suppressing noise.
2. The difference image is subsequently added to the Gaussian image, thereby enhancing the structural features of small-scale images.

When applying the Gaussian filtering as part of our image pre-processing, we used a standard deviation ( $\sigma$ ) value of 1.5 for the Gaussian kernel. This value was chosen to effectively balance the smoothing effect and preservation of image details, crucial for maintaining the integrity of features relevant to our classification task.

These steps are closely associated with the characteristics of the filtering structure in FTC and FTA. Some filtering structures exhibit unclear boundaries due to speckle noise interference. Applying a general Maxpool operation can cause these boundaries to vanish. Therefore, in multi-scale image processing, enhancing boundaries is highly important.

#### Multi-scale feature fusion module

The notion of multi-scale feature fusion was proposed for a considerable time and was initially implemented in the universal neural enhancement technology (UNET) segmentation network [34]. Additionally, in the realm of object detection, Kaiming He introduced the feature pyramid network (FPN) structure in RetinaNet, which exemplifies the concept of multi-scale feature fusion [35]. However, there are notable distinctions between these two approaches. The FPN is primarily employed for object detection, while the UNET is utilized for segmentation. The FPN produces multiple layers of output, whereas the UNET provides output solely in the final layer. Furthermore, their upsampling methods differ, with direct interpolation used in one case and up-convolution and optimizing parameters used in the other. While the FPN employs an addition operation for skip connections, the UNET utilizes concatenation.

Upon examining the dataset, we observed that the lesion targets in FTA and FTC tended to be relatively large, necessitating a substantial receptive field for accurate identification. As a result, reversing the fusion of high-level features lacks significance in this context. Drawing inspiration from the FPN, we adopted an “add” fusion method during the downsampling process of each

**Table 1** Distribution of the dataset

	Training Set	Validation Set	Test Set
FTC(label1)	226	57	42
FTA(label0)	98	24	34
Total	324	81	76

layer’s output. In contrast to UNET and FPN, we omitted the upsampling step and retained the fundamental features of FTC and FTA. By feeding the outcomes of each layer into the inference, we reduce the computational burden by eliminating the upsampling module.

#### Datasets and experimental setting

The experimental images were sourced from two hospitals. All the images were cropped to the minimum bounding square that encompassed the FTA and FTC lesions. Subsequently, the images were resized to  $256 \times 256$  pixels. Of these, 283 FTA images and 122 FTC images were allocated for the training and validation sets, respectively, while 76 images were designated as the test set (Table 1). Fivefold random cross-validation was employed for training the images. The 405 training sets were randomly divided into five subsets, with one subset reserved for validation and the remaining four subsets utilized for training (Table 2). The group that yielded superior results in the validation set was ultimately selected as the final result group. During the model training phase, the batch size was set to 2, the number of epochs was set to 1000, and the learning rate was set to  $1e-4$ . Considering that the dataset is not large, the batch size and learning rate settings are relatively small.

#### Evaluation methods

The performance of our model was assessed by calculating several metrics, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and F1 score. Additionally, we utilize a confusion matrix to calculate the true positive rate and false positive rate of the model. To evaluate the overall performance of the model, we plotted an ROC curve and calculated the area under the curve (AUC) metric. The F1 score, which is the harmonic mean of precision and recall, is considered a comprehensive metric that combines both precision and recall, making it highly useful for evaluating model performance. Furthermore, we employed decision curve analysis (DCA) to assess the value of the model. By comparing the DCA curves of different models, we can determine which model is better

**Table 2** Distribution of the data for the five subsets

	Subset 1	Subset 2	Subset 3	Subset 4	Subset 5
FTC(label1)	56	57	56	57	57
FTA(label0)	25	24	25	24	24

suites for specific decision scenarios and select the optimal decision threshold [36].

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

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

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

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}}$$

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

$$\text{F1} = \frac{2 * \text{PPV} * \text{Sensitivity}}{\text{PPV} + \text{Sensitivity}}$$

In our evaluation, TP (true positive) corresponds to the number of accurately classified FTC cases, while TN (true negative) refers to the number of accurately classified FTA cases. On the other hand, FP (false positive) and FN (false negative) indicate the number of incorrectly classified FTC/FTA cases.

Based on the aforementioned information, sensitivity (also known as recall) represents the model's ability to accurately predict FTC samples out of all samples that truly belong to the FTC category. On the other hand, specificity denotes the model's ability to correctly identify FTA samples out of all samples that genuinely fall into the FTA category. The precision (also known as the PPV) corresponds to the proportion of samples correctly classified as FTC by the model out of all samples predicted as FTC, while the NPV indicates the proportion of samples accurately identified as FTA among all samples predicted as FTA by the model. The accuracy represents the overall proportion of samples correctly predicted by the model out of the entire sample set, and the F1 value serves as an evaluation metric that considers both precision and recall simultaneously.

### Statistical analysis

Statistical analysis was performed using SPSS 22 software (SPSS, 1989; Apache Software Foundation, Chicago, IL, USA). The Delong test was also conducted to assess any significant differences in diagnostic performance among the various models. A two-sided p-value of less than 0.05 was considered to indicate statistical significance. ROC curves were generated to determine the area under the ROC curve (AUROC), cutoff values, sensitivity,

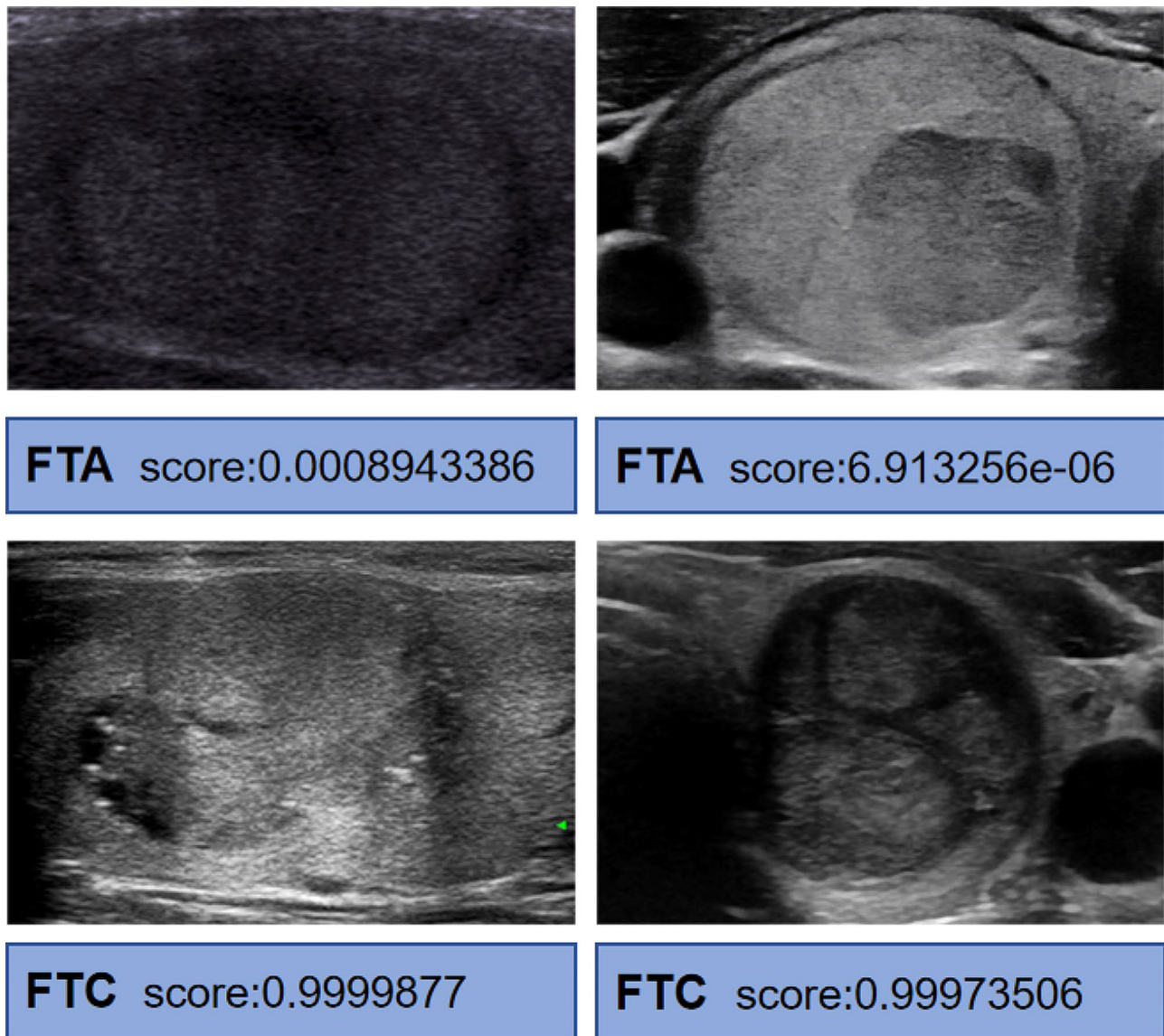
specificity, positive predictive value (PPV), and negative predictive value (NPV).

### Results

Two criteria were used to evaluate the score of the model: 0.5 was used as the cutoff for identification, a deviation in the score to 0 was considered benign, and a deviation in the score to 1 was considered malignant. Figure 4 shows the ultrasound image cases and model scores of several tests. The results showed that our model performed well and could accurately distinguish between FTC and FTA in two similar ultrasound images. Based on the experimental data provided, MRF-Net outperforms other commonly used neural network models, exhibiting the highest sensitivity (79.41%), specificity (90.24%), PPV (87.1%), and NPV (84.1%) (Table 3). The accuracy, a fundamental metric for evaluating predictive performance, was reported as 85.3% for MRF-Net, surpassing the other models presented in Table 3. This indicates that the MRF-Net model is more likely to correctly predict outcomes than are the other models. To account for potential class imbalances within the dataset, the F1 value was calculated, providing a comprehensive evaluation of the models' performance. MRF-Net achieved an impressive F1 value of 83.08%, suggesting a well-balanced prediction of both FTA and FTC outcomes, making it a reliable predictor for the dataset.

87.1

Figure 5 displays the experimental results of MobileNet V3, ResNet50, and DenseNet121, with their ROC curves closely clustered and an overall AUC ranging from 0.721 to 0.785. Among these models, MRF-Net (red line) exhibited relatively better performance, with an overall AUC of 0.848, the highest among them. A comparison of the ROC curves revealed that the AUC of MRF-Net was significantly greater than that of MobileNet V3 and ResNet50, with corresponding p values of 0.0007 and 0.0054, respectively. Moreover, DCA curve analysis demonstrated that the net benefit curve of MRF-Net consistently outperformed that of the other models across different decision thresholds (Fig. 6). These findings indicate that MRF-Net is a superior classification model, particularly for distinguishing between benign and malignant thyroid follicular tumors. This finding suggested that MRF-Net would provide greater benefits to patients and offer valuable support to doctors in making informed decisions in such cases. Furthermore, the performance of the four models was assessed by analyzing their confusion matrices. As depicted in Fig. 7, the horizontal axis represents the model's output results, with dark blue squares indicating the probability of correct predictions for FTA or FTC. Notably, MRF-Net exhibited higher true positive and true negative rates than did the other three



**Fig. 4** Examples of the case tested

**Table 3** Comparison of experimental results

(%)	MobileNetV3	Resnet50	DenseNet121	MRF-Net
Sensitivity	73.53	76.47	76.47	<b>79.41</b>
Specificity	70.73	75.61	80.49	<b>90.24</b>
PPV	67.57	72.22	76.47	87.1
NPV	76.32	79.49	80.49	<b>84.1</b>
Accuracy	72	76	78.67	<b>85.3</b>
F1 value	70.42	74.29	76.47	<b>83.08</b>

models, while its false positive and false negative rates were correspondingly lower.

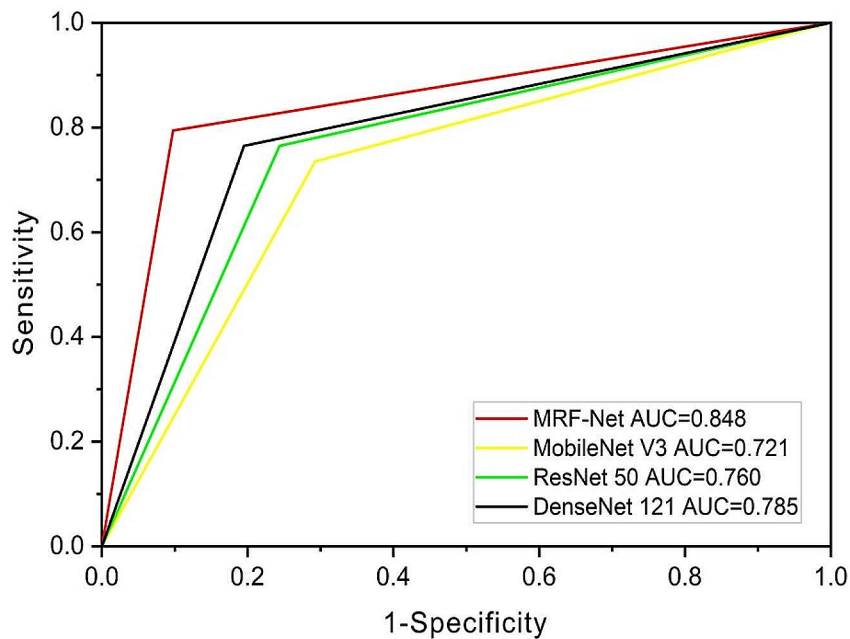
The aforementioned findings indicate that our neural network exhibits a high level of accuracy and reliability in differentiating between benign and malignant thyroid follicular tumors. Notably, the utilization of multi-scale feature fusion in our network design further enhances

its performance in this task. This research introduces a novel approach for discriminating between benign and malignant thyroid nodules in clinical medicine, offering promising prospects for practical applications.

### Discussion

Accurately distinguishing between benign and malignant thyroid follicular tumors before surgery is crucial for clinical treatment and subsequent therapy selection. If the model-assisted diagnosis suggests a greater possibility of FTA, patients and clinical physicians may choose conservative drug treatment or partial thyroidectomy to reduce unnecessary harm and improve quality of life. Despite the use of preoperative ultrasonography and fine needle aspiration cytology, the effectiveness of these methods in differentiating thyroid FTC from FTA has been limited.





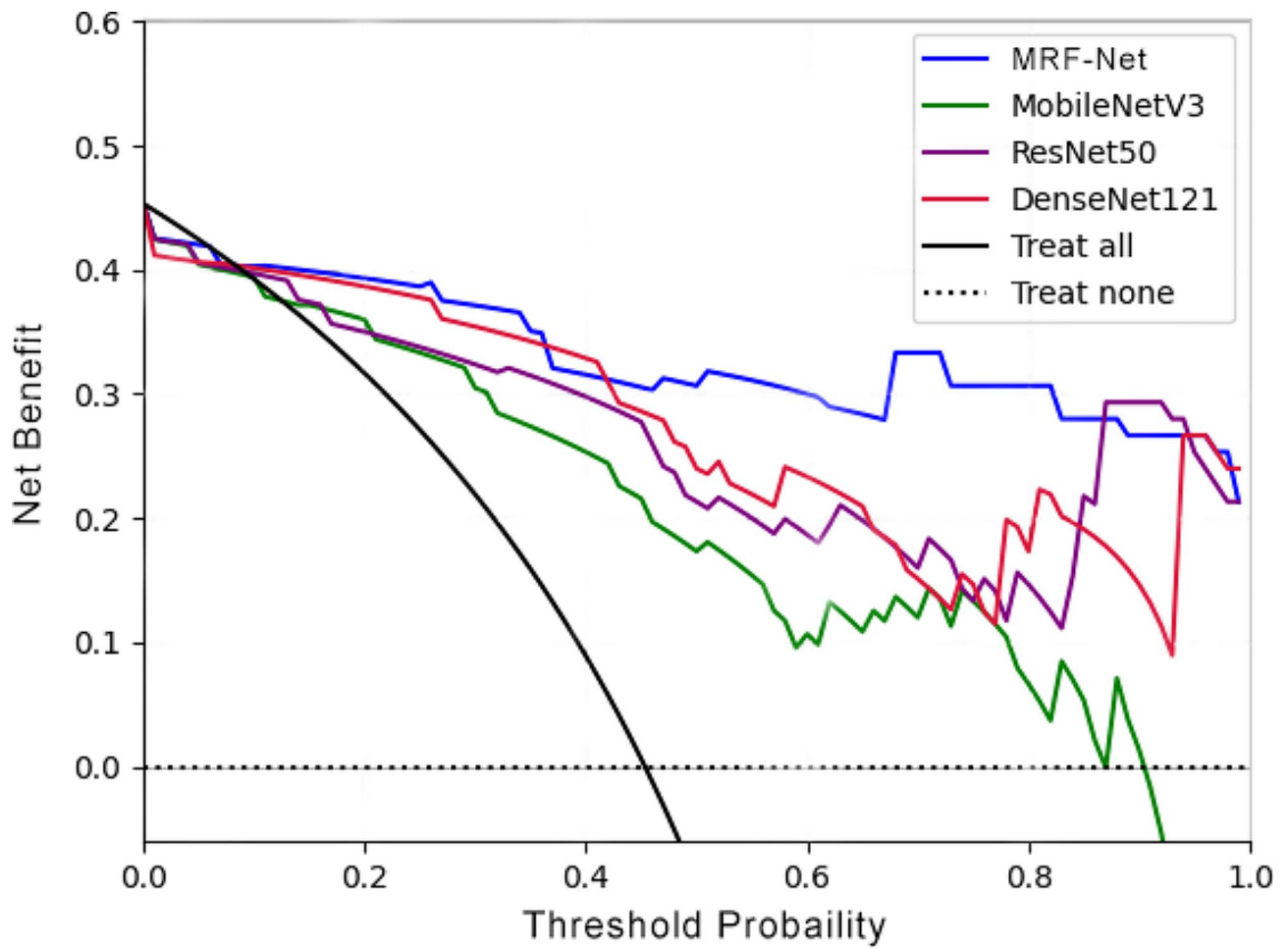
**Fig. 5** The AUC values of different models

According to the investigation results of some researchers, it is difficult to identify FTA and FTC even in pathology and other imaging examinations, and their research results show low accuracy. These tumors exhibit visual similarities on ultrasound images, posing challenges even for experienced ultrasonographers. Moreover, the diagnostic accuracy of these methods is constrained by the restricted range of puncture positions.

Currently, there is limited research on follicular thyroid tumors, and further exploration is needed. Previous studies have used machine learning to manually outline lesions and analyze the benign or malignant nature of follicular thyroid tumors, achieving an accuracy rate of 74.1%. However, this approach may result in the loss of edge image feature extraction, thereby reducing the accuracy. To address these limitations, our study proposes a novel deep learning network based on multi-scale images for FTC and FTA classification. The lesion was placed in the middle of the cropping box, and the integrity of the edge image features was preserved. In contrast to conventional methods, we enhance the head layer of the network architecture by incorporating a stretch-enhanced image (REI) module. Our approach involves employing a Gaussian pyramid-like structure to process the input image at multiple scales. We perform absolute maximum pooling on the differential image to preserve key structures, suppress noise, and reintegrate it with the Gaussian image to enhance the structural features of small-scale images. The results demonstrated a high level

of sensitivity (79.41%) and specificity (90.24%) achieved by our model, underscoring its potential for clinical application. Moreover, we also compared MRF-Net with the latest model Vision Transformer, and the results similarly showed that the former showed better results. By combining ultrasound and artificial intelligence data to construct a diagnostic prediction model based on clinical and imaging data, establishing an FTC/FTA diagnostic system is a key focus in future research on follicular tumors. However, the quantity and quality of images are crucial for artificial intelligence models, necessitating collaboration to increase multi-center database resources. Improving the accuracy of preoperative diagnosis of follicular thyroid tumors is highly important for personalized patient treatment and the rational allocation of medical resources.

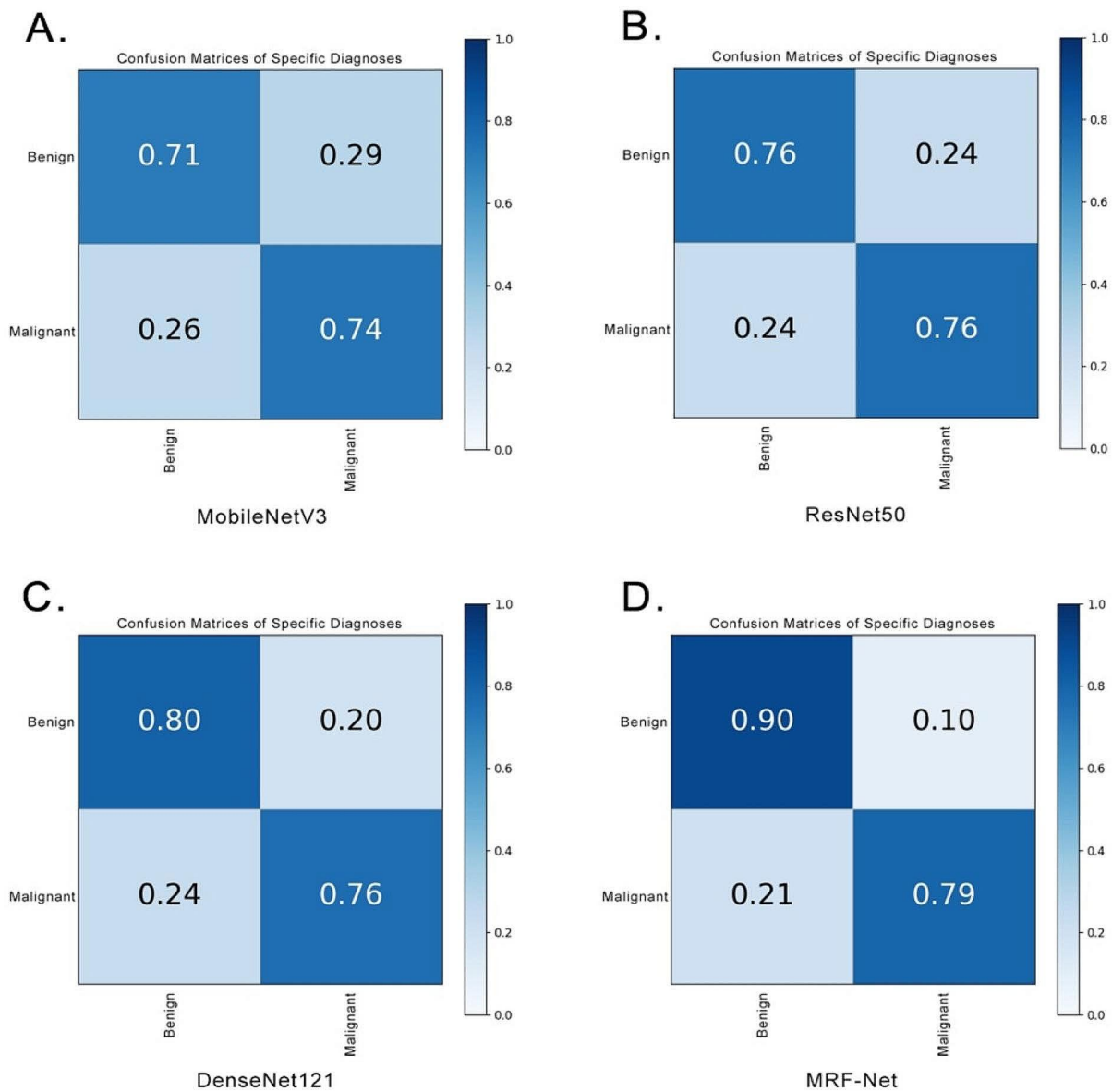
Nevertheless, our study has several limitations. First, the sample size was relatively small, consisting of 279 patients' ultrasound images, with 324 images in the training set, 81 images in the validation set, and 76 images in the test set. Therefore, expanding the dataset is necessary to enhance the generalizability of the study. Second, as a retrospective study, selection bias cannot be entirely eliminated. Finally, the ultrasound images were obtained from two hospitals, potentially leading to variations in image quality due to differences in machine models and physician expertise, which could impact the data results.



**Fig. 6** The DCA curves of different models

**Conclusion**

The neural network incorporating multi-scale integration exhibited outstanding performance in distinguishing FTC from FTA, providing valuable support for clinical diagnosis and treatment decision-making.



**Fig. 7** The confusion matrix charts for MobileNetV3 (A), Resnet50 (B), DenseNet121 (C) and MRF-Net (D)

**Abbreviations**

FTA	follicular thyroid adenoma
FTC	follicular thyroid carcinoma
PTC	papillary thyroid carcinoma
FNAC	fine-needle aspiration cytology
TI-RADS	thyroid imaging reporting and data system
SVM	support vector machine
ANN	artificial neural network
CNNs	convolutional neural networks
MRF-Net	Multi-Rescale Fusion Network
REI	Rescale Enhancement Image
UNET	Universal Neural Enhancement Technology
ReLU	Rectified Linear Unit
FPN	Feature Pyramid Network
PPV	positive predictive value
NPV	negative predictive value
AUC	area under the curve

DCA	decision curve analysis
TP	true positive
TN	true negative
FP	false positive
FN	false negative
ROC	receiver operating characteristic
AUROC	area under ROC curves
PPV	positive predictive value
NPV	negative predictive value

**Acknowledgements**

Not applicable.

**Author contributions**

Conceptualization, Y.F.Y. and M.S.H.; methodology, W.W.C.; software, C.Q. and M.Q.H.; validation, L.Y., M.D.L.; formal analysis, F.L.K.; investigation, X.J.N.; resources, W.W.C.; data curation, X.J.N. and W.W.C.; writing—original draft

preparation, W.W.C.; writing—review and editing, Y.F.Y. and M.S.H.; visualization, X.J.N.; supervision, Y.F.Y. and Z.Z.; project administration, Y.F.Y. and M.S.H.; funding acquisition, Y.F.Y. and Z.Z. All authors have read and agreed to the published version of the manuscript.

### Funding

This research was funded by China Postdoctoral Science Foundation (2022M711721), the Jiangsu University Medical Education Collaborative Innovation Fund (JDY2022002), the Fifth Phase "169 Project" Scientific Research Project of Zhenjiang City (YLJ201931), and Social Development Program of Zhenjiang City (SH2019038, SH2021028).

### Data availability

No datasets were generated or analysed during the current study.

### Declarations

#### Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki, and approved by the institutional review board of the Affiliated Hospital of Nantong University (2023-K054-01). The requirement for informed consent was waived.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that there is no conflict of interest.

Received: 16 November 2023 / Accepted: 13 March 2024

Published online: 27 March 2024

### References

1. Carling T, Udelsman R. Follicular neoplasms of the thyroid: what to recommend. *Thyroid*. 2005;15(6):583–7.
2. Cihan BY, Koc A, Tokmak TT. The role of Radiotherapy in Skull Metastasis of Thyroid Follicular Carcinoma. *Klin Onkol*. 2019;32(4):300–2.
3. Sugino K, Ito K, Nagahama M, Kitagawa W, Shibuya H, Ohkuwa K, Yano Y, Uruno T, Akaishi J, Kameyama K, et al. Prognosis and prognostic factors for distant metastases and tumor mortality in follicular thyroid carcinoma. *Thyroid*. 2011;21(7):751–7.
4. Kitahara CM, Sosa JA. The changing incidence of thyroid cancer. *Nat Rev Endocrinol*. 2016;12(11):646–53.
5. Ito Y, Hirokawa M, Masuoka H, Yabuta T, Fukushima M, Kihara M, Higashiyama T, Takamura Y, Kobayashi K, Miya A, et al. Distant metastasis at diagnosis and large tumor size are significant prognostic factors of widely invasive follicular thyroid carcinoma. *Endocr J*. 2013;60(6):829–33.
6. Hirokawa M, Ito Y, Kuma S, Takamura Y, Miya A, Kobayashi K, Miyauchi A. Nodal metastasis in well-differentiated follicular carcinoma of the thyroid: its incidence and clinical significance. *Oncol Lett*. 2010;1(5):873–6.
7. Kwon MR, Shin JH, Park H, Cho H, Kim E, Hahn SY. Radiomics based on thyroid Ultrasound can predict distant metastasis of follicular thyroid carcinoma. *J Clin Med* 2020, 9(7).
8. Kuo TC, Wu MH, Chen KY, Hsieh MS, Chen A, Chen CN. Ultrasonographic features for differentiating follicular thyroid carcinoma and follicular adenoma. *Asian J Surg*. 2020;43(1):339–46.
9. Yoon JH, Kim EK, Youk JH, Moon HJ, Kwak JY. Better understanding in the differentiation of thyroid follicular adenoma, follicular carcinoma, and follicular variant of papillary carcinoma: a retrospective study. *Int J Endocrinol*. 2014;2014:321595.
10. Chng CL, Kurzwinski TR, Beale T. Value of sonographic features in predicting malignancy in thyroid nodules diagnosed as follicular neoplasm on cytology. *Clin Endocrinol (Oxf)*. 2015;83(5):711–6.
11. Seo HS, Lee DH, Park SH, Min HS, Na DG. Thyroid follicular neoplasms: can sonography distinguish between adenomas and carcinomas? *J Clin Ultrasound*. 2009;37(9):493–500.
12. Goyal A, Patel V. Multiple periauticular nodules diagnosed as gout on fine-needle aspiration cytology. *Indian J Med Res*. 2019;149(5):682–3.
13. Hoang JK, Middleton WD, Farjat AE, Langer JE, Reading CC, Teeffey SA, Abinanti N, Boschini FJ, Bronner AJ, Dahiya N, et al. Reduction in thyroid nodule biopsies and Improved Accuracy with American College of Radiology Thyroid Imaging Reporting and Data System. *Radiology*. 2018;287(1):185–93.
14. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, Pacini F, Randolph GW, Sawka AM, Schlumberger M, et al. 2015 American Thyroid Association Management Guidelines for adult patients with thyroid nodules and differentiated thyroid Cancer: the American Thyroid Association Guidelines Task Force on thyroid nodules and differentiated thyroid Cancer. *Thyroid*. 2016;26(1):1–133.
15. Shen YT, Chen L, Yue WW, Xu HX. Artificial intelligence in ultrasound. *Eur J Radiol*. 2021;139:109717.
16. Yadav N, Dass R, Virmani J. Objective assessment of segmentation models for thyroid ultrasound images. *J Ultrasound*. 2023;26(3):673–85.
17. Kriti, Virmani J, Agarwal R. Assessment of despeckle filtering algorithms for segmentation of breast tumours from ultrasound images. *Biocybern Biomed Eng*. 2019;39(1):100–21.
18. Rajeshwar Dass, Yadav N. Image Quality Assessment parameters for Despeckling Filters. *Procedia Comput Sci*. 2020;167:2382–92.
19. Yadav N, Dass R, Virmani J. Despeckling filters applied to thyroid ultrasound images: a comparative analysis. *Multimed Tools Appl*. 2022;81(6):8905–37.
20. Peng S, Liu Y, Lv W, Liu L, Zhou Q, Yang H, Ren J, Liu G, Wang X, Zhang X, et al. Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study. *Lancet Digit Health*. 2021;3(4):e250–9.
21. Ko SY, Lee JH, Yoon JH, Na H, Hong E, Han K, Jung I, Kim EK, Moon HJ, Park VY, et al. Deep convolutional neural network for the diagnosis of thyroid nodules on ultrasound. *Head Neck*. 2019;41(4):885–91.
22. Tong WJ, Wu SH, Cheng MQ, Huang H, Liang JY, Li CQ, Guo HL, He DN, Liu YH, Xiao H, et al. Integration of Artificial Intelligence decision aids to reduce workload and enhance efficiency in thyroid nodule management. *JAMA Netw Open*. 2023;6(5):e2313674.
23. Shin I, Kim YJ, Han K, Lee E, Kim HJ, Shin JH, Moon HJ, Youk JH, Kim KG, Kwak JY. Application of machine learning to ultrasound images to differentiate follicular neoplasms of the thyroid gland. *Ultrasonography*. 2020;39(3):257–65.
24. Seo JK, Kim YJ, Kim KG, Shin I, Shin JH, Kwak JY. Differentiation of the follicular neoplasm on the Gray-Scale US by Image Selection Subsampling along with the marginal outline using convolutional neural network. *Biomed Res Int*. 2017;2017:3098293.
25. Alabrak MMA, Megahed M, Alkhouly AA, Mohammed A, Elfandy H, Tahoun N, Ismail HA-R. Artificial Intelligence Role in subclassifying cytology of thyroid follicular neoplasm. *Asian Pac J cancer Prevention: APJCP*. 2023;24(4):1379–87.
26. Deng CW, Li D, Feng M, Han DY, Huang QQ. The value of deep neural networks in the pathological classification of thyroid tumors. *Diagn Pathol*. 2023;18(1):11.
27. Chan WK, Sun JH, Liou MJ, Li YR, Chou WY, Liu FH, Chen ST, Peng SJ. Using deep convolutional neural networks for enhanced Ultrasonographic Image diagnosis of differentiated thyroid Cancer. *Biomedicine*. 2021;9(12):14.
28. Yadav N, Dass R, Virmani J. Deep learning-based CAD system design for thyroid tumor characterization using ultrasound images. *Multimed Tools Appl* 2023.
29. Shapiro NA, Poloz TL, Shkurupij VA, Tarkov MS, Poloz VV, Demin AV. Application of artificial neural network for classification of thyroid follicular tumors. *Anal Quant Cytol Histol*. 2007;29(2):87–94.
30. Elizar E, Zulkifley MA, Muharar R, Zaman MHM, Mustaza SM. A Review on Multiscale-Deep-Learning Applications. *Sensors (Basel)* 2022, 22(19).
31. Gao Z, Sun X, Liu M, Dang W, Ma C, Chen G. Attention-based parallel multiscale convolutional neural network for visual evoked potentials EEG classification. *IEEE J Biomed Health Inf*. 2021;25(8):2887–94.
32. Agnes SA, Anitha J, Pandian SIA, Peter JD. Classification of Mammogram images using Multiscale all convolutional neural network (MA-CNN). *J Med Syst*. 2019;44(1):30.
33. Ansari MY, Yang Y, Balakrishnan S, Abinahed J, Al-Ansari A, Warfa M, Almkadad O, Barah A, Omer A, Singh AV, et al. A lightweight neural network with multiscale feature enhancement for liver CT segmentation. *Sci Rep*. 2022;12(1):14153.
34. Li Z, Wang H, Han Q, Liu J, Hou M, Chen G, Tian Y, Weng T. Convolutional Neural Network with Multiscale Fusion and Attention Mechanism for Skin Diseases Assisted Diagnosis. *Comput Intell Neurosci* 2022, 2022:8390997.
35. Krithika Alias AnbuDevi M, Suganthi K. Review of semantic segmentation of medical images using modified architectures of UNET. *Diagnostics (Basel)* 2022, 12(12).

36. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak.* 2006;26(6):565–74.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.