


ThyroidXL: Advancing Thyroid Nodule Diagnosis with an Expert-Labeled, Pathology-Validated Dataset

Hung Duong Viet^{1,3}, Huan Vu^{1,4}, Duong Phan Huong²,
Quyen Nguyen Duc¹, Hao Pham Duc^{1,3}, Toan Le Quang²,
Sy Nguyen Ba², Dung Do Tien², Sang Dinh Viet³, Cuong Nguyen Tien^{1,4},
Hoang Pham Huy¹, Hy Ngo Dien¹

¹ VNPT AI, VNPT Group, Hanoi, Vietnam

² National Hospital of Endocrinology, Hanoi, Vietnam

³ Hanoi University of Science and Technology, Hanoi, Vietnam

⁴ National Economics University, Hanoi, Vietnam


hungdv@vnpt.vn, huanv@neu.edu.vn

Abstract. Thyroid nodules are among the most prevalent endocrine disorders, with incidence rates increasing in recent years. Ultrasonography remains the primary method for thyroid nodule diagnosis due to its non-invasive nature and cost-effectiveness; however, the process is subjective and skill-intensive. To assist radiologists, Computer-Aided Diagnosis systems (CAD) have been developed to provide a second opinion. Despite these advancements, the absence of publicly available medical datasets has resulted in inconsistent validation methods, deterring comparability across studies. This paper introduces ThyroidXL, an open benchmark dataset for thyroid nodule classification, segmentation, and detection. With over 11,000 images from more than 4,000 patients, the dataset—collected and annotated by expert radiologists at the Vietnam National Hospital of Endocrinology—stands as the largest publicly available resource for thyroid nodule diagnosis in terms of both patient count and image volume. Additionally, we provide multiple deep-learning baseline models on three key tasks, including malignancy classification, thyroid nodule detection, and segmentation. The proposed dataset and benchmark can serve as a foundational resource for advancing CAD system development, fostering reproducible research, and accelerating progress in thyroid nodule diagnosis. Our dataset can be accessed at: <https://huggingface.co/datasets/hunglc007/ThyroidXL>

Keywords: Thyroid Nodules · Ultrasonography · Deep Learning.

1 Introduction

Thyroid cancer is an increasingly prevalent health issue, ranking seventh in overall incident rate and fifth among women in 2022 [1]. Advancements in medical

Corresponding author

technology have allowed early detection and treatment of thyroid cancer, thus increasing the survival rate. However, the widespread use of diagnostic imaging also contributes to increased incidence. Therefore, a key challenge for physicians is to ensure that benign or low-risk patients are not over-treated, which can be achieved by undertaking a thorough diagnosis.

Two methods are commonly utilized in thyroid cancer diagnosis, including ultrasonography (US) and fine needle aspiration cytology (FNAC). The latter is the gold standard for evaluating thyroid nodules, but the former is more commonly used due to its accessibility, affordability, and ionizing-free nature [2]. However, US interpretation requires a spatial understanding of the anatomy and accurate recognition of diagnostic-related features [2], making it a demanding, time-consuming, and subjective process. To standardize US-based classification, [3] introduced the Thyroid Imaging Reporting and Data Systems (TI-RADS) in 2009, which categorizes thyroid nodules based on their risk of malignancy using ultrasound features. Over the years, multiple variants were developed, with ACR TI-RADS [4] emerging as the most widely accepted.

To overcome the limitations of traditional thyroid cancer diagnosis, modern CAD systems leveraging data-driven algorithms have been developed over the years, and they have demonstrated near-expert performance [5,6,7,8]. Despite these advancements, the process of advancing these systems is hindered by a limited number of publicly available datasets due to the technical, ethical, and legal factors in medical data collection.

To bridge this gap, we present a new public **ThyroidXL** (XL stands for eXpert-Labeled) benchmark dataset for the training and evaluation of CAD tools in thyroid cancer diagnosis. The dataset consists of 11635 B-mode ultrasound images from 4093 patients at the Vietnam National Hospital of Endocrinology over two years, starting from February 2023. Diagnoses were biopsy-confirmed, and patient data were anonymized to ensure privacy. Ethical approval was obtained from the Research Ethics Committee, and all patients provided written consent for research use.

2 Related Work

CAD systems for thyroid cancer rely on high-quality datasets for training and evaluation. Over the years, several dedicated datasets have been introduced to support research in this domain, each with unique characteristics, advantages, and limitations.

The DDTI dataset [9] was the first open-access dataset for developing thyroid cancer diagnosis, comprising 347 ultrasound images from 299 patients at the IDIME Ultrasound Department. Being one of the first benchmark databases of US images, the dataset has certain limitations, including a limited sample size and low image resolution. In 2017, [10] introduced the TG3K dataset, which contains 3585 images from 16 ultrasonic videos. However, it was designed for thyroid gland segmentation rather than thyroid nodule classification. The Stanford CINE dataset [11] is a more recent open dataset containing 192 cine clips from

167 patients at the Stanford University Medical Center. The TN-SCUI dataset [12], released in 2020, includes 3644 images, each from a patient at the Shanghai Ruijin Hospital. While the dataset provides valuable annotated samples, it remains inaccessible for general research due to ethical approval constraints. In 2021, [13,14] proposed the TN3K dataset, which contains 3493 ultrasound images from 2421 patients collected between January 2016 and August 2020. Despite its moderate size, the dataset lacks comprehensive clinical metadata. Additionally, both the TN3K and DDTI datasets contain handwritten diameter indicators on some images. These markings can partially obscure critical anatomical structures, potentially introducing bias in model training and affecting generalization performance [16]. In 2024, [15] released a large public dataset of 8,508 thyroid ultrasound images from 842 patients. However, the limited patient count raises overfitting concerns. Furthermore, several studies also utilized thyroid ultrasound datasets, though they remain inaccessible to the broader research community [5,8,16,17]. The lack of accessibility to these datasets restricts independent validation and benchmarking, underscoring the need for large-scale thyroid ultrasound datasets.

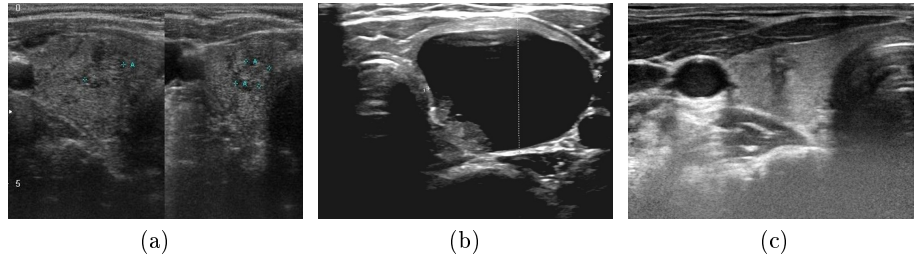


Fig. 1. Samples from public datasets. Fig. 1(a) and Fig. 1(b) depict hand-written marks from the DDTI and TN3K datasets, respectively, while Fig. 1(c) presents a sample from the ThyroidXL dataset, which contains no hand-written marks.

3 The ThyroidXL Dataset

3.1 Image Acquisition and Annotation

The ultrasound images in our datasets are collected by experienced physicians (≥ 5 years in thyroid ultrasound). The selection of patients is conducted before the study based on the following criteria: (1) normal thyroid function, (2) the thyroid nodules' diameters exceeding 5mm, and (3) no history of prior interventions, including alcohol injection therapy, radiofrequency ablation, or repeated aspiration. Before undergoing ultrasound imaging, the selected patients are informed about the study and asked to sign consent forms. The standard DICOM images are extracted from the video sequences captured using Hitachi Aloka

Arielta V70 ultrasonic image systems, with care taken to ensure that each image contains only a single nodule. In cases where multiple nodules are present, we select the most suspicious one. Images are screened for clarity, avoiding blurring or motion artifacts, and categorized using the ACR TI-RADS scoring system [4]. Fine-needle aspiration of the thyroid nodules is performed immediately following image capture under ultrasound guidance. Cytological results are classified using the Bethesda [18] system for reporting thyroid cytopathology. Most patients diagnosed with malignant nodules undergo surgical excision, and the excised nodules are marked by surgeons before being sent for pathological examination to avoid misidentification. Cytological and pathological results, along with patient demographic data, are documented to complete the research records.

After acquiring the DICOM images, they are converted into an uncompressed PNG format and provided to experienced radiologists for annotation. For malignancy classification, labels were directly derived from cytological and pathological findings, eliminating inter-annotator variability. Three board-certified radiologists annotated the images for segmentation. Each image was independently labeled by two annotators (≥ 5 years of experience), with 8% of cases resolved by a third senior radiologist (≥ 30 years of experience) where conflicts were present. During the preprocessing stage, the non-ultrasound image regions containing patient information are cropped, and the patient IDs are remapped to ensure data confidentiality. The resulting dataset is used for further analysis.

3.2 Dataset Statistics

Dataset size: Our dataset consists of 11,635 images from 4,093 patients, split 80-20 into training and validation sets. The training set includes 9,541 images from 3,354 patients, while the validation set has 2,094 images from 739 patients. As shown in Table 1, the ThyroidXL dataset contains the highest number of images and patients among publicly available datasets.

Table 1. Number of images and patients comparison among public datasets

No.	ThyroidXL	DDTI	TN3K	TG3K	TN-SCUI
Images	11 635	347	3493	3585	3644
Patients	4093	299	2421	-	3644

Dataset distribution: Fig. 2(a) and Fig. 2(d) illustrate the age distribution of patients across the training and test sets. The data indicates that most patients fall within the age range of 30 to 60 years. The major difference between the two sets is the benign-to-malignant ratio, which is approximately 50:50 (386 benign, 353 malignant) in the test set and 65:35 (2,477 benign, 877 malignant) in the training set. This imbalance influences the distribution of TI-RADS categories across the sets, as shown in Fig. 2(b), Fig. 2(c), Fig. 2(e) and Fig. 2(f). Additionally, the dataset reflects real-world epidemiology, with female patients

outnumbering males by a factor of eight (3,650 females, 443 males), consistent with GLOBOCAN statistics for Southeast Asia [1].

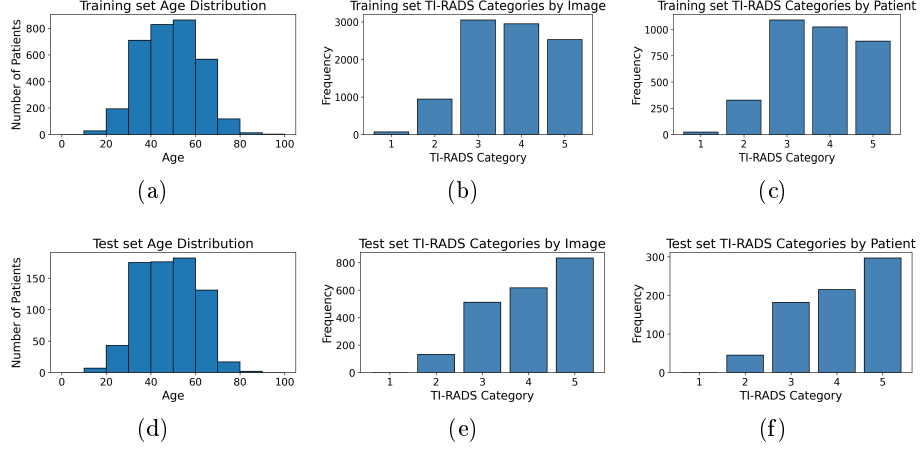


Fig. 2. Training set and test set patients' age and TI-RADS categories distributions

4 Benchmark

In this section, we conduct three types of experiments: malignancy status classification, thyroid nodule detection, and segmentation. For each task, we create a benchmark of the state-of-the-art deep learning algorithms. In addition to the common metrics, we also calculate *Sensitivity* and *Specificity* image-wise and patient-wise, as they are crucial performance metrics in medical image analysis. The formulas are as follows:

$$Specificity = \frac{\text{True Negatives}}{\text{True Negative} + \text{False Positives}} \quad (1)$$

$$Sensitivity = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (2)$$

Specifically, to determine the final classification at the patient level, we apply **Weighted Majority Voting (WMV)** across multiple images from the same patient. The formula is as follows:

$$C_{final} = \arg \max_c \sum_{i=1}^n w_i I(y_i = c) \quad (3)$$

where C_{final} is the final classification for the patient, y_i represents the classification result of image i , w_i is the confidence score of the model for that image,

Table 2. Performance of Classification Models

Model	Accuracy	Precision	Recall	F1-score	Image		Patient	
					Sensitivity	Specificity	Sensitivity	Specificity
AlexNet	0.798	0.751	0.847	0.796	0.847	0.756	0.887	0.785
Resnet34	0.825	0.821	0.798	0.839	0.798	0.849	0.819	0.839
Resnet50	0.826	0.793	0.846	0.806	0.846	0.806	0.867	0.806
EfficientNet-B6	0.814	0.796	0.858	0.811	0.858	0.776	0.878	0.780
EfficientNet-B7	0.831	0.809	0.833	0.820	0.833	0.829	0.839	0.878
VGG-11 Batchnorm	0.818	0.781	0.845	0.812	0.845	0.794	0.850	0.847
VGG-13 Batchnorm	0.830	0.796	0.855	0.824	0.855	0.809	0.853	0.824

and $I(y_i = c)$ is an indicator function that is 1 if the classification of image i matches class c , otherwise 0. All experiments are conducted using a DGX node of 8 Nvidia 40GB A100 GPU.

4.1 Malignancy Status Classification

For the malignancy status classification task, we experiment with models commonly used for classification tasks, including AlexNet [19], VGG [20], ResNet [21], and EfficientNet [22].

Training Procedure: We utilize several modules for the training process, including PyTorch *, Albumentations **, and Hydra ***. Each model is trained on input images of size 224×224 using the AdamW optimizer [23] for 100 epochs, with the learning rate set to 5×10^{-4} . Regarding the loss function, the Binary Cross-Entropy (BCE) is employed.

Metrics: Model evaluation and analysis are performed using standard performance metrics, including Accuracy, Precision, Recall, and F_1 -score, with F_1 -score being the main metric.

Classification Results: The performance of classification models are summarized in Table 2. For each model, we select the two versions that yield the best results. The best F_1 -score is 0.824, achieved by the VGG-13 with batch normalization. This is followed by the EfficientNet-B7 with the F_1 -score being 0.820. Regarding medical image analysis metrics, both models show equally high sensitivity and specificity compared to the other models.

4.2 Thyroid Nodule Detection

For the thyroid nodule detection task, we experiment with six object detection models: EfficientNet-B3 [22], Faster R-CNN [24], YOLOX [25], Deformable DETR [26], and CO-DETR [27], employing two backbone architectures, including ResNet-50 [21] and Swin-L [28].

Training Procedure: Each model is trained using a specific set of hyperparameters tuned for optimal performance. The number of epochs ranges from 20

* <https://github.com/pytorch/pytorch.git>

** <https://github.com/albumentations-team/albumentations.git>

*** <https://github.com/facebookresearch/hydra.git>

Table 3. Performance of Object Detection Models

Model	mAP	mAP@50	mAP@75	Precision	F1-score	Image		Patient	
						Sensitivity	Specificity	Sensitivity	Specificity
EfficientNet-B3	0.568	0.899	0.638	0.905	0.824	0.756	0.931	0.785	0.940
Faster R-CNN	0.527	0.868	0.577	0.871	0.817	0.769	0.900	0.799	0.928
YOLOX-S	0.570	0.898	0.640	0.894	0.848	0.807	0.917	0.819	0.935
YOLOX-M	0.580	0.904	0.649	0.906	0.866	0.830	0.925	0.873	0.930
Deformable DETR	0.538	0.890	0.584	0.878	0.815	0.760	0.908	0.785	0.951
CO-DETR R50	0.586	0.902	0.661	0.897	0.810	0.739	0.926	0.754	0.953
CO-DETR Swin-L	0.613	0.904	0.702	0.899	0.833	0.777	0.924	0.802	0.935

to 200, depending on convergence. The Stochastic Gradient Descent (SGD) [29] and the Adam [30] optimizers, combined with weight decay and momentum, are utilized to optimize models’ parameters. For learning rate scheduling, we choose linear decay and cosine annealing strategies. Data augmentation techniques, including random flipping, rotation, and intensity normalization, are applied to enhance generalization.

Metrics: The models are assessed using COCO-style [31] evaluation metrics for object detection, including Mean Average Precision (mAP), mAP@0.5, and mAP@0.75 [32]. Regarding the classification task, to get the malignancy status of each image, we extract the class score of the detected object with the highest confidence. The classification performance is evaluated using standard clinical diagnostic metrics similar to section 4.1.

Result: Table 3 summarizes the performance of different models. Among the evaluated architectures, CO-DETR with a Swin-L backbone achieves the highest detection accuracy with an mAP of 0.613, mAP@50 of 0.904, and mAP@75 of 0.702. However, its sensitivity (0.777) and specificity (0.924) are lower compared to YOLOX-M, which achieves a sensitivity of 0.830 and specificity of 0.925, despite having a lower mAP of 0.580.

4.3 Thyroid Nodule Segmentation

For the thyroid nodule segmentation task, we opt for the U-Net [33], and U-Net++ [34], two deep learning architectures specifically designed for medical image analysis, with ResNet [21] and EfficientNet [22] backbones.

Training Procedure: Frameworks and platforms such as Segmentation Model Pytorch[†], Pytorch, and Albumentations are utilized for the model training process. The input images and masks are of size 384x480. We use the RAdam optimizer [35] to train each model for 100 epochs with the learning rate set to 5×10^{-4} .

Metrics: The performance of the segmentation models is evaluated using standard metrics, including IoU-score and Dice-score, with IoU-score being the main metric. The specificity and sensitivity in the segmentation task are calculated similarly to 4.1. To get the malignancy status from the predicted masks, we refer to the number of pixels belonging to each class.

[†] https://github.com/qubvel-org/segmentation_models.pytorch.git

Table 4. Performance of Object Segmentation Models

Model	Backbone	IoU	Dice	Accuracy	Precision	Recall	F1-score	Image		Patient	
								Sensitivity	Specificity	Sensitivity	Specificity
UNet	Resnet101	0.568	0.644	0.794	0.822	0.709	0.762	0.683	0.977	0.679	0.980
	Resnet152	0.576	0.650	0.797	0.815	0.729	0.769	0.685	0.979	0.681	0.981
	EfficientNet-B5	0.621	0.697	0.821	0.824	0.783	0.803	0.713	0.982	0.707	0.984
	EfficientNet-B6	0.640	0.716	0.842	0.814	0.856	0.835	0.743	0.982	0.745	0.984
UNet++	Resnet101	0.566	0.640	0.794	0.804	0.736	0.768	0.669	0.979	0.669	0.981
	Resnet152	0.571	0.643	0.784	0.786	0.734	0.759	0.665	0.981	0.658	0.983
	EfficientNet-B5	0.648	0.725	0.848	0.819	0.863	0.840	0.755	0.982	0.759	0.984
	EfficientNet-B6	0.641	0.717	0.838	0.832	0.816	0.824	0.749	0.981	0.752	0.984

Result: Table 4 illustrates the effectiveness of various segmentation models. Overall, the UNet++ with an EfficientNet-B5 backbone yields the best results, with IoU score, Dice score, Accuracy, and F1-score being 0.648, 0.725, 0.848, and 0.840 respectively. However, its precision, being 0.819, is lower compared to other models. Furthermore, it also ranks first in terms of specificity and sensitivity.

5 Discussions

In this study, we benchmark several state-of-the-art architectures to establish baseline performance on our ThyroidXL dataset. For thyroid nodule detection, CO-DETR with a Swin-L backbone achieves superior detection accuracy, its sensitivity and specificity are lower than those of YOLOX-M. This discrepancy can be attributed to the different optimization trade-offs between object detection and classification tasks. CO-DETR, being a transformer-based model, emphasizes high localization accuracy, which benefits the object detection task but may lead to suboptimal decision boundaries for classification. In contrast, YOLOX-M, a one-stage detector, balances both detection and classification tasks more effectively, resulting in improved sensitivity and specificity.

Notably, models often exhibit a trade-off between high sensitivity and low specificity or vice versa. This suggests that while they effectively identify benign nodules, they may struggle with detecting malignant ones or the other way around, leading to either underdiagnosis of malignancies or misclassification of benign cases. This trade-off highlights the importance of selecting models based on clinical priorities, where a balance between sensitivity and specificity is essential. Additionally, the use of Weighted Majority Voting significantly improves models' sensitivity and specificity when evaluating at the patient level. These findings highlight the importance of considering patient-level aggregation when evaluating diagnostic AI models.

Compared to previous studies [9,10,11,12,13,14,15], the ThyroidXL dataset offers a larger sample size, improved image quality, and rich clinical metadata. However, the dataset still shows a slight imbalance in the malignant-benign ratio, which may affect model performance. Moreover, models trained on our dataset might not generalize well to images from different equipment or clinical settings. Thus, future work should consider data augmentation, resampling, and domain adaptation strategies to overcome these challenges.

6 Conclusions

In this paper, we present ThyroidXL, a dataset of ultrasound images specifically designed for thyroid nodule diagnosis. The dataset consists of 11635 ultrasound images from 4093 patients at the Vietnam National Hospital of Endocrinology. Additionally, we provide the evaluation of state-of-the-art deep learning models on three key tasks, including malignancy status classification, thyroid nodule detection, and segmentation. In addition to commonly used metrics, we compute *Specificity* and *Sensitivity* at both the image and patient levels, as these are critical performance indicators in medical image analysis. Consequently, we anticipate that the ThyroidXL dataset will serve as a substantial contribution to advancing research in this domain.

Acknowledgments. This study was funded by the Vietnam Ministry of Science and Technology (grant number KC-4.0-43/19-25).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R., Soerjomataram, I. & Jemal, A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal For Clinicians*. **74**, 229-263 (2024)
2. Huang, H., Dong, Y., Jia, X., Zhou, J., Ni, D., Cheng, J. & Huang, R. Personalized diagnostic tool for thyroid cancer
3. Horvath, E., Majlis, S., Rossi, R., Franco, C., Niedmann, J., Castro, A. & Dominguez, M. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. *The Journal Of Clinical Endocrinology and Metabolism*. **94**, 1748-1751 (2009)
4. Tessler, F., Middleton, W., Grant, E., Hoang, J., Berland, L., Teefey, S., Cronan, J., Beland, M., Desser, T., Frates, M. & Others ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee. *Journal Of The American College Of Radiology*. **14**, 587-595 (2017)
5. Li, X., Zhang, S., Zhang, Q., Wei, X., Pan, Y., Zhao, J., Xin, X., Qin, C., Wang, X., Li, J. & Others Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. *The Lancet Oncology*. **20**, 193-201 (2019)
6. Guan, Q., Wang, Y., Du, J., Qin, Y., Lu, H., Xiang, J. & Wang, F. Deep learning based classification of ultrasound images for thyroid nodules: a large scale of pilot study. *Annals Of Translational Medicine*. **7**, 137 (2019)
7. Anari, S., Tataei Sarshar, N., Mahjoori, N., Dorosti, S. & Rezaie, A. Review of deep learning approaches for thyroid cancer diagnosis. *Mathematical Problems In Engineering*. **2022**, 5052435 (2022)
8. Wang, L., Zhang, L., Zhu, M., Qi, X. & Yi, Z. Automatic diagnosis for thyroid nodules in ultrasound images by deep neural networks. *Medical Image Analysis*. **61** pp. 101665 (2020)

9. Pedraza, L., Vargas, C., Narváez, F., Durán, O., Muñoz, E. & Romero, E. An open access thyroid ultrasound image database. *10th International Symposium On Medical Information Processing And Analysis*. **9287** pp. 188-193 (2015)
10. Wunderling, T., Golla, B., Poudel, P., Arens, C., Friebe, M. & Hansen, C. Comparison of thyroid segmentation techniques for 3D ultrasound. *Medical Imaging 2017: Image Processing*. **10133** pp. 346-352 (2017)
11. AIMI, S. Thyroid Ultrasound CINE Clip Dataset. , <https://aimi.stanford.edu/datasets/thyroid-ultrasound-cine-clip>
12. Zhou, J., Jia, X., Ni, D., Noble, A., Huang, R., Tan, T. & Van, M. Thyroid Nodule Segmentation and Classification in Ultrasound Images. (Zenodo,2020,3)
13. Gong, H., Chen, G., Wang, R., Xie, X., Mao, M., Yu, Y., Chen, F. & Li, G. Multi-Task Learning For Thyroid Nodule Segmentation With Thyroid Region Prior. *2021 IEEE International Symposium On Biomedical Imaging*. pp. 257-261 (2021)
14. Gong, H., Chen, J., Chen, G., Li, H., Li, G. & Chen, F. Thyroid region prior guided attention for ultrasound segmentation of thyroid nodules. *Computers In Biology And Medicine*. **155** pp. 106389 (2023)
15. Hou, X., Hua, M., Zhang, W., Ji, J., Zhang, X., Jiang, H., Li, M., Wu, X., Zhao, W., Sun, S. & Others An ultrasonography of thyroid nodules dataset with pathological diagnosis annotation for deep learning. *Scientific Data*. **11**, 1272 (2024)
16. Chi, J., Walia, E., Babyn, P., Wang, J., Groot, G. & Eramian, M. Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. *Journal Of Digital Imaging*. **30** pp. 477-486 (2017)
17. Liu, T., Xie, S., Yu, J., Niu, L. & Sun, W. Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features. *2017 IEEE International Conference On Acoustics, Speech And Signal Processing (ICASSP)*. pp. 919-923 (2017)
18. Cibas, E. & Ali, S. The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid*. **27**, 1341-1346 (2017)
19. Krizhevsky, A., Sutskever, I. & Hinton, G. Imagenet classification with deep convolutional neural networks. *Advances In Neural Information Processing Systems*. **25** (2012)
20. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*. (2014)
21. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition*. pp. 770-778 (2016)
22. Tan, M. & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. *International Conference On Machine Learning*. pp. 6105-6114 (2019)
23. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. *ArXiv Preprint ArXiv:1711.05101*. (2017)
24. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions On Pattern Analysis And Machine Intelligence*. (2017,6)
25. Ge, Z., Liu, S., Wang, F., Li, Z. & Sun, J. YOLOX: Exceeding YOLO Series in 2021. *ArXiv Preprint ArXiv:2107.08430*. (2021)
26. Zhu, X., Su, W., Lu, L., Li, B., Wang, X. & Dai, J. Deformable DETR: Deformable Transformers for End-to-End Object Detection. *International Conference On Learning Representations*. (2021)
27. Zong, Z., Song, G. & Liu, Y. Detrs with collaborative hybrid assignments training. *Proceedings Of The IEEE/CVF International Conference On Computer Vision*. pp. 6748-6758 (2023)

28. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S. & Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. *The IEEE/CVF International Conference On Computer Vision*. pp. 10012-10022 (2021)
29. Robbins, H. & Monro, S. A stochastic approximation method. *The Annals Of Mathematical Statistics*. pp. 400-407 (1951)
30. Kingma, D. & Ba, J. Adam: A method for stochastic optimization. *ArXiv Preprint ArXiv:1412.6980*. (2014)
31. Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. & Zitnick, C. Microsoft COCO: Common Objects in Context. *European Conference On Computer Vision (ECCV)*. (2014)
32. Everingham, M., Van Gool, L., Williams, C., Winn, J. & Zisserman, A. The Pascal visual object classes challenge. *International Journal Of Computer Vision*. **88** pp. 303-338 (2010)
33. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing And Computer-assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18. pp. 234-241 (2015)
34. Zhou, Z., Rahman Siddiquee, M., Tajbakhsh, N. & Liang, J. Unet++: A nested u-net architecture for medical image segmentation. *Deep Learning In Medical Image Analysis And Multimodal Learning For Clinical Decision Support: 4th International Workshop, DLMIA 2018, And 8th International Workshop, ML-CDS 2018, Held In Conjunction With MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings* 4. pp. 3-11 (2018)
35. Liu, L., Jiang, H., He, P., Chen, W., Liu, X., Gao, J. & Han, J. On the variance of the adaptive learning rate and beyond. *ArXiv Preprint ArXiv:1908.03265*. (2019)