ThyGraph: A Graph-Based Approach for Thyroid Nodule Diagnosis from Ultrasound Studies

Ashwath Radhachandran[1,2], Alekhya Vittalam[1], Vedrana Ivezic[1,3], Vivek Sant[5], Shreeram Athreya[1,7], Chace Moleta[1,6], Maitraya Patel[4], Rinat Masamed[4], Corey Arnold[1,2,3,4,6,7], and William Speier[1,2,3,4]

[1] Biomedical AI Research Lab, UCLA, Los Angeles, CA, USA speier@ucla.edu
[2] Department of Bioengineering, UCLA, Los Angeles, CA, USA
[3] Medical Informatics Home Area, Graduate Programs in Bioscience, UCLA, Los Angeles, CA, USA
[4] Department of Radiological Sciences, UCLA, Los Angeles, CA, USA
[5] Division of Endocrine Surgery, University of Texas Southwestern Medical Center, Dallas, TX, USA
[6] Department of Pathology and Laboratory Medicine, UCLA, Los Angeles, CA, USA
[7] Department of Electrical and Computer Engineering, UCLA, Los Angeles, CA, USA

Abstract. Improved thyroid nodule risk stratification from ultrasound (US) can mitigate overdiagnosis and unnecessary biopsies. Previous studies often train deep learning models using manually selected single US frames; these approaches deviate from clinical practice where physicians utilize multiple image views for diagnosis. This paper introduces ThyGraph, a novel graph-based approach that improves feature aggregation and correlates anatomically proximate images, by leveraging spatial information to model US image studies as patient-level graphs. Graph convolutional networks are trained on image-based and patch-based graphs generated from 505 US image studies to predict nodule malignancy. Self-attention graph pooling is introduced to produce a node-level interpretability metric that is visualized downstream to identify important inputs. Our best performing model demonstrated an AUROC of 0.866±0.019 and AUPRC of 0.749±0.043 across five-fold cross validation, significantly outperforming two previously published attention-based feature aggregation networks. These previous studies fail to account for spatial dependencies by modeling images within a study as independent, uncorrelated instances. In the proposed graph paradigm, ThyGraph can effectively aggregate information across views of a nodule and take advantage of inter-image dependencies to improve nodule risk stratification, leading to better patient triaging and reducing reliance on biopsies. Code is available at https://github.com/ashwath-radha/ThyGraph.

Keywords: ultrasound · graph convolutional networks · multi-view · thyroid nodule · interpretability

1 Introduction

Over the past two decades, thyroid cancer incidence in the United States has nearly tripled [6]. Ultrasound (US) imaging is the primary modality used by radiologists to assess detected nodules, but this process can be subjective and time-consuming. Moreover, over 90% of detected nodules are clinically insignificant or benign, so uncertainty in radiologic evaluation can lead to unnecessary fine-needle aspiration biopsies (FNABs) [2, 10]. With the fundamental role of US in thyroid nodule diagnostics and the need for a more deterministic, non-invasive diagnostic alternative to biopsies, research has focused on developing deep learning methods that enhance nodule malignancy prediction from US [12]. In recent years, convolutional neural networks (CNNs) have become popular for automating diagnostic tasks from imaging [8, 17]. Most current approaches for thyroid nodule diagnosis select an optimal US displaying a nodule and develop a CNN-based classifier with these images [21, 11]. This approach is limited because it requires a physician to manually select a specific frame from an image study. More importantly, physicians typically use multiple image views of a nodule to make effective diagnoses. In this context, a view is defined as a specific perspective of a nodule/anatomy captured through an ultrasound image. Given the clinical context for this diagnostic task, an improved information aggregation technique across US studies could improve automated malignancy classification.

Methodology that combines information from various imaging perspectives has been explored in tangential domains, such as lung CT, chest radiographs, and breast ultrasound [22, 15, 4]. But to the best of our knowledge, there has been minimal research within the space of thyroid nodule diagnosis in leveraging features across multiple US images. Two prior studies demonstrated how cine clips could be used to improve risk stratification of thyroid nodules [18, 16]. However, the cines for these studies have highly correlated frames from a narrow anatomical region and are clipped to contain a consistent view of a nodule. Huang et al. used a transverse and longitudinal image of the thyroid, both containing a view of the nodule, to develop a model [3]. Despite demonstrating potential, this approach would require the user to choose two input images and does not leverage the remaining image study to provide additional physiological context. Wang et al. also attempt to leverage whole image studies using an attention-based feature aggregation method [14]. Most recently, Zhuang et al. proposed an ensembled attention multiple instance learning (AMIL) approach that combines US features at various image scales [23]. Despite using whole US studies and exhibiting promising results, the approaches proposed by Wang and Zhuang are trained in an AMIL framework, where each US is treated independently, disregarding their inter-image relationships.

Graphs provide an effective alternative for modeling US. By representing image features as nodes and accounting for inter-image relationships as edges, spatial dependencies and contextual information can be learned through the application of graph convolutional networks (GCNs). GCNs layer convolution operations to learn better node representations and incorporate local and global graph structure information [5]. Yin et al. proposed an end-to-end framework

using a joint CNN-GCN to diagnose kidney US studies [20]. In this case, feature similarity between deep features extracted from US frames was used as a metric to determine connectivity. Yan et al. were the first to use GCNs in the domain of thyroid US. In their study, a dataset-level graph was created to represent every patients' nodules and correlate images with similar features [19]. Despite the effectiveness of these methods, an approach that integrates anatomical context during graph construction and generates unique patient-level graphs could be more effective. To take advantage of untapped opportunities with thyroid US, this paper presents ThyGraph, a novel graph-based approach to model image studies and improve thyroid nodule diagnosis. The main contributions are as follows:

1. ThyGraph, a patient-level graph construction method that models US studies by correlating anatomically related images to underscore inter-image dependencies.
2. Implementation of multi-scale graphs that use information from both full frames and localized patches.
3. A visualization method that leverages attention-based graph pooling to correlate attention weights with image orientation and anatomical location, and plot the weights for better model explainability.

2 Methodology

2.1 Data Collection and Preprocessing

A retrospective, dataset of 505 US image studies from unique patients was collected from an academic medical center. Institutional Review Board (IRB# 19–001535) of University of California, Los Angeles gave ethical approval for this study. During an US examination, 20-80 images are taken that span a variety of anatomic regions and spatial orientations. Each image study includes a transverse or longitudinal view of the thyroid isthmus, and transverse (inferior, middle, and superior) and longitudinal (lateral, middle, and medial) views of the right and left thyroid lobes.

The main inclusion criterion was a fine-needle aspiration biopsy (FNAB) performed within one year of the US examination. The corresponding cytology report was acquired and following the Bethesda System for Reporting Thyroid Cytopathology, the categorization for risk of malignancy was extracted. Consistent with clinical practice, biopsies with benign cytology were labeled as benign, while those with malignant/suspicious for malignancy cytology were deemed malignant [1]. In cases of multiple biopsied nodules, the one with the highest Bethesda categorization determined the image study's label: benign if all nodules were benign and malignant if at least one nodule was malignant. This resulted in 371 benign and 134 malignant image studies. Across these samples, there were 17,626 US images. Ultrasounds were passed through a data preprocessing pipeline, which involved deidentification, negative space and artifact removal, intensity rescaling and image whitening.

2.2 State-of-the-Art Baselines

The first baseline, proposed by Wang et al. (Wang model), is an attention-based aggregation network that fuses features across an US study for thyroid nodule diagnosis [14]. The architecture is similar to a MIL framework, but unique in that each image study is randomly subsampled or oversampled each training iteration to ensure a consistent input size across all studies. The second baseline, proposed by Zhuang et al., is the multi-scale, attention-based multiple-instance learning model (MS-AMIL) ensemble model [23]. By aggregating features across whole frames and image patches, multi-scale information regarding the nodule was extracted. Their best performing model, an ensemble across features from 128x128 (MS-AMIL-patch128) and 256x256 (MS-AMIL-patch256) frame patches was replicated in our experiments.

2.3 ThyGraph

Image-Based Graph Construction The dataset can be defined as D: $\{(s_n,l_n)$; n=1,...,N$\}$ for N image studies. l_n is the label, signifying benign or malignant, for image study s_n. There are M_n ultrasound frames in image study s_n:$\{f_{nm}; m = 1, ..., M_n\}$. Each frame, f_{nm} has unique dimensions R^{wXhX1}, where w and h represent the width and height of the frame.

Every frame within an image study also has embedded text representing that frame's location and spatial orientation within the thyroid (ex. Left Transverse Superior). An optical character recognition (OCR) algorithm was developed to process each frame and extract this text. Any frames with views of surrounding lymph nodes were not included. Fig. 1 provides a representation of the different locations and spatial orientations a frame can have. For each frame, f_{nm}, its OCR output is used to determine a 2-D coordinate, (x, y). Specifically, the x-coordinate determines whether the frame captures the isthmus, or a lateral, middle or medial view of the left or right lobes. The y-coordinate determines whether the frame is in an inferior, middle or superior position. Thus, each frame's OCR output, which equates to the frame's anatomical location, is mapped to a 2-D coordinate that allows us to estimate each frame's anatomical distance from one another.

Next, for each image study, an undirected graph is constructed in which each node is represented by a frame-level feature vector. This feature vector, R^{1X6016}, is a concatenation of the final fully connected layer of a pretrained ResNet101, DenseNet201 and ResNeXt101. Since this ensemble model demonstrated promising results in a previous study for thyroid nodule diagnosis, it was repurposed for this task [9]. Euclidean distance is then calculated between the 2-D coordinates for every pair of frames in the study following

$$d(f, f') = \|f - f'\|. \tag{1}$$

The adjacency matrix, A, to represent the graph's edges is constructed to capture the weights associated with each edge

$$A_{ij} = max(0, \alpha - d(f_i, f_j)), \tag{2}$$

where i and j represent the row and column indices of the adjacency matrix, corresponding to two different graph nodes. This definition permits the graph to assign more weight to edges between nodes whose frames share proximate coordinates, an indicator of anatomical proximity.

Patch-Based Graph Construction Next, the image-based graph construction method is extended to accommodate image patches. Each frame, f_{nm}, with an OCR-extracted location is tiled into nine patches, represented as p_{nmk} for $k \in [1, 9]$, each with dimension $R^{224X224X1}$. The patch extraction algorithm maximizes frame coverage by generating patches within the frame boundaries, avoiding negative space. Each patch is given an additional z-dimension, extending the coordinate representation of its source frame, resulting in a three-dimensional coordinate, (x, y, z). The z-coordinate represents the patch's position along the posterior-anterior axis. By leveraging the anatomical context provided by OCR-extracted location, each patch is represented within the 3-D coordinate system. Now, for each image study, it is possible to build a patch-based undirected graph, where each node is represented by a patch-level feature vector, R^{1X2048} which is extracted using a ResNet50 model, pretrained on ImageNet. Eq. (1) is adapted to calculate Euclidean distance between all pairs of patch coordinates. Then, Eq. (2) determines the graph's adjacency matrix, with the patch coordinates as inputs and a different empirically determined threshold selected accordingly. Thus, similar to the image-based graphs, each node's neighborhood inherently estimates that patch's anatomical proximity to others.

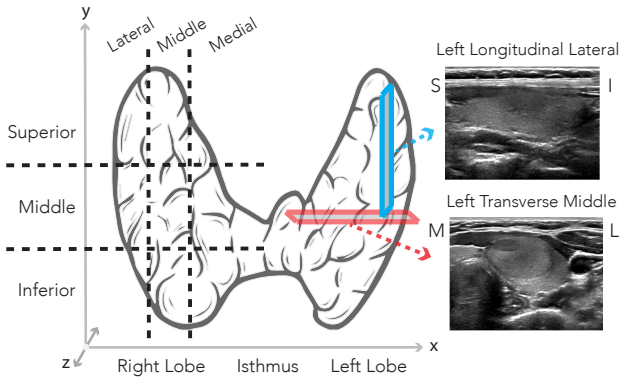

Fig. 1. Various spatial orientations and locations captured by frames within a thyroid US study. The two example images are labeled with superior (S), inferior (I), medial (M) and lateral (L) to help the viewer orient each image within the context of the 3-D coordinate system.

GCN with Self-Attention Graph Pooling The non-Euclidean nature of graphs can be leveraged through GCNs to promote learned correlations. A two-layer GCN is trained on the image-based (image-GCN) and patch-based (patch-GCN) graphs to serve as baselines. Furthermore, we explore the effect of introducing self-attention graph (SAG) pooling [7] to the patch-GCN. The resulting patch-SAG-GCN model includes SAG pooling after each graph convolution. SAG pooling calculates node-level attention scores using an auxiliary graph network and then creates an attention mask to retain a subset of nodes the network deems to be of higher value. This joint feature-based and topology-based pooling method allows the architecture to progressively identify more relevant nodes and capture localized graph information.

During evaluation, for each patient-level graph passed through patch-SAG-GCN, we extract the attention scores generated by the first SAG pooling operation. Each score is associated with a graph node, which via our graph construction method, represents a different image patch. Since each patch has a 3-D coordinate mapping within the system in Fig.1, a coordinate-level attention can be assigned. Moreover, in a given graph, certain patches can be mapped back to the same coordinate; in this case, the aggregated attention scores are averaged to create a mean coordinate-level attention.

2.4 Model Evaluation and Implementation

All architectures, including the Wang model and MS-AMIL baselines, are trained and evaluated using stratified 5-fold cross validation. Evaluation metrics include area under the receiver operating characteristic curve (AUC), area under the precision-recall curve (AUPRC), accuracy, precision and recall. The results are reported in terms of mean and standard deviation across the five cross validation folds. All architectures, except the Wang model, were trained for 200 epochs with early stopping based on validation AUC and a patience value of 25 epochs. The Adam optimizer is used with a learning rate of 1e-4. Weighted cross-entropy loss is used to train the models. The Wang model was trained using different hyperparameters in line with the original paper. The statistical significance of the proposed models relative to the Wang model and MS-AMIL was compared at significance level 0.05 using DeLong test for AUC and Wilcoxon signed-rank test for accuracy. All experiments are implemented in PyTorch with a NVIDIA DGX-1.

3 Results and Discussion

First, the impact of ThyGraph graph construction method is evaluated. The image-GCN and the patch-GCN demonstrated 0.713 and 0.847 AUC, respectively. The patch-SAG-GCN on its own also achieved 0.847 AUC, but when ensembled with the image-GCN and patch-GCN models (GCN-Ensemble), demonstrated 0.853 AUC. The MS-AMIL-patch128 and MS-AMIL-patch256 branches demonstrated a 0.83 and 0.824 AUC, respectively, and when ensembled had

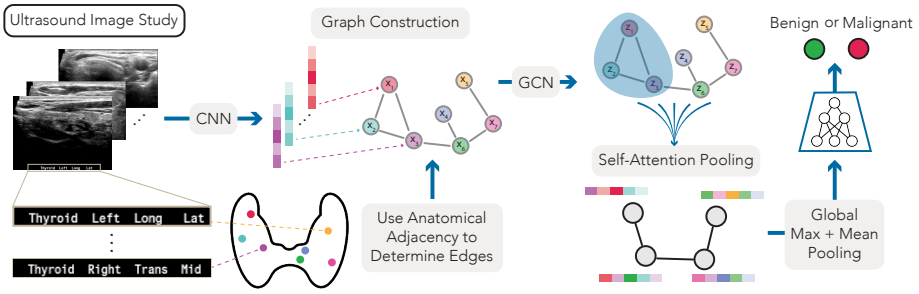

Fig. 2. Overview of proposed graph construction and model architecture.

a 0.837 AUC. With a larger patient cohort, we were able to demonstrate improved performance for MS-AMIL patch ensemble from the published results of 0.785±0.104 AUC. Despite this improvement, the patch-GCN still outperformed the MS-AMIL-patch128, MS-AMIL-patch256 and the patch ensemble.

The MS-AMIL framework, although designed to independently analyze and aggregate features across images, can still identify important input instances through its attention mechanism. Despite the framework not considering dependencies between inputs, the patterns it learns are still valuable in nodule malignancy prediction. To incorporate the contributions of the MS-AMIL model with ThyGraph, we assessed an ensemble of MS-AMIL-patch128, MS-AMIL-patch256, image-GCN, patch-GCN and patch-SAG-GCN (Full-Ensemble), which demonstrated an 0.866 AUC, significantly improving upon MS-AMIL ($p=0.03$) and Wang model ($p<0.001$). The Wang model achieved a 0.632 AUC on the in-house dataset, but was significantly outperformed by all proposed models ($p<0.001$). The performance discrepancy with the Wang model can be attributed to the use of a different training dataset, and the published results using smaller and more variably sized images studies.

Table 1. Model performance across the five cross validation folds comparing the proposed approach with state-of-the-art baselines (*statistically significant compared to MS-AMIL at significance level 0.05).

Model	AUC	AUPRC	Accuracy	Precision	Recall
Wang model [14]	0.63 ± 0.04	0.39 ± 0.04	0.63 ± 0.09	0.41 ± 0.06	0.67 ± 0.15
MS-AMIL [23]	0.84 ± 0.03	0.68 ± 0.06	0.83 ± 0.04	0.68 ± 0.08	0.73 ± 0.07
image-GCN	0.71 ± 0.06	0.52 ± 0.12	0.68 ± 0.07	0.46 ± 0.06	0.73 ± 0.19
patch-GCN	0.85 ± 0.03	0.71 ± 0.09	0.81 ± 0.05	0.63 ± 0.1	**0.81** ± 0.06
patch-SAG-GCN	0.85 ± 0.03	0.7 ± 0.07	**0.85** ± 0.03	0.71 ± 0.05	0.75 ± 0.06
GCN-Ensemble	0.85 ± 0.03	0.72 ± 0.06	**0.85** ± 0.04	**0.73** ± 0.08	0.73 ± 0.08
Full-Ensemble	**0.87** ± 0.02*	**0.75** ± 0.04	0.84 ± 0.04*	0.69 ± 0.08	0.78 ± 0.09

3.1 Attention-Based Interpretability

The coordinate-level attention scores extracted from patch-SAG-GCN are visualized within a 3-D thyroid mesh[13]. Each 3-D coordinate is mapped to a corresponding location within the mesh and the coordinate-level attention scores are used to create a volumetric heat map. Hot spots indicate that patch-SAG-GCN is focusing on imaging features from those thyroid regions when making a certain prediction. Fig.3 illustrates the value of our attention-based visualization technique in enhancing model interpretability. In case a), the model places greater weight on features from the right lobe, which is also the location of the biopsied malignant nodule. Similarly, in case b), the model focuses on features from the left lobe, the location of the biopsied benign nodule. However, it also directs attention to the right inferior lobe, which according to the cytology report, is the location of a second, unbiopsied nodule.

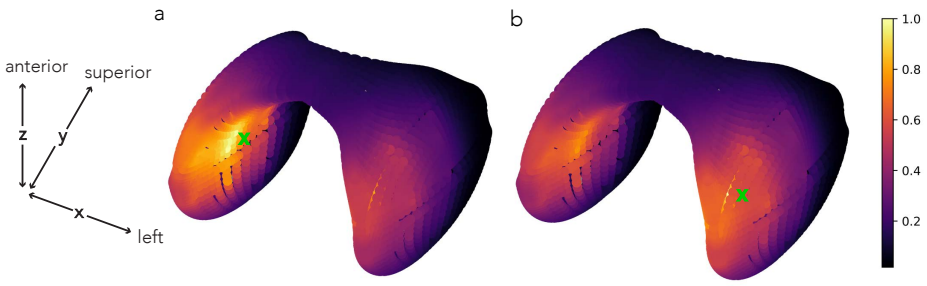

Fig. 3. Visualization of attention heatmap overlaid on a thyroid mesh model for two patients. The attention weights are scaled between 0 and 1 for each patient. The biopsied nodule location, extracted from corresponding cytology report, is marked with a green X. As indicated, the x-axis stretches from right to left, the y-axis from inferior to superior and the z-axis from posterior to anterior.

4 Conclusion

We propose ThyGraph, a novel approach to aggregate information across thyroid US image studies. The majority of past work required manual intervention to select an optimal frame to train, evaluate and prospectively use deep learning-based thyroid nodule diagnostic methods. Previous approaches seeking to fuse multi-view information typically treated each image independently, through MIL-inspired approaches, disregarding inter-image dependencies. ThyGraph not only effectively models an US study despite its non-uniform nature but specifically integrates spatial location information to better represent correlated ultrasounds. Subsequently, the patch-GCN and image-GCN are able to learn multi-scale information across the image study, and the patch-SAG-GCN

promotes model explainability, allowing us to localize suspicious regions within the thyroid. Moving forward, different patch- and frame-level feature extractors can be considered, such as vision transformers to improve on the current CNN-based technique. Another limitation of this paper is that further external validation datasets from other academic institutions could help highlight this method's strengths and weaknesses. Ultimately, the novel graph construction method of ThyGraph and model ensemble of MIL and GCN techniques, presents notable advantages in terms of automation and interpretability, enhancing our ability to effectively risk stratify thyroid nodules.

Acknowledgments. This work was supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award number R21EB030691 as well as a UCLA Radiology Exploratory Research Grant. The authors would also like to thank Shawn Chen of the UCLA Biomedical AI Research Lab for help in data collection.

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

References

1. Ali, S.Z., Baloch, Z.W., Cochand-Priollet, B., Schmitt, F.C., Vielh, P., VanderLaan, P.A.: The 2023 Bethesda System for Reporting Thyroid Cytopathology. Thyroid® **33**(9), 1039–1044 (Sep 2023). https://doi.org/10.1089/thy.2023.0141
2. Durante, C., Costante, G., Lucisano, G., Bruno, R., Meringolo, D., Paciaroni, A., Puxeddu, E., Torlontano, M., Tumino, S., Attard, M., Lamartina, L., Nicolucci, A., Filetti, S.: The natural history of benign thyroid nodules. JAMA **313**(9), 926–935 (Mar 2015). https://doi.org/10.1001/jama.2015.0956
3. Huang, H., Dong, Y., Jia, X., Zhou, J., Ni, D., Cheng, J., Huang, R.: Personalized Diagnostic Tool for Thyroid Cancer Classification Using Multi-view Ultrasound. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. pp. 665–674. Lecture Notes in Computer Science, Springer Nature Switzerland, Cham (2022). https://doi.org/10.1007/978-3-031-16437-8_64
4. Kang, G., Liu, K., Hou, B., Zhang, N.: 3D multi-view convolutional neural networks for lung nodule classification. PLOS ONE **12**(11), e0188290 (Nov 2017). https://doi.org/10.1371/journal.pone.0188290
5. Kipf, T.N., Welling, M.: Semi-Supervised Classification with Graph Convolutional Networks (Feb 2017). https://doi.org/10.48550/arXiv.1609.02907
6. Kitahara, C.M., Sosa, J.A., Shiels, M.S.: Influence of Nomenclature Changes on Trends in Papillary Thyroid Cancer Incidence in the United States, 2000 to 2017. The Journal of Clinical Endocrinology and Metabolism **105**(12), e4823–4830 (Dec 2020). https://doi.org/10.1210/clinem/dgaa690
7. Lee, J., Lee, I., Kang, J.: Self-Attention Graph Pooling (Apr 2019)
8. Liu, C., Cao, Y., Alcantara, M., Liu, B., Brunette, M., Peinado, J., Curioso, W.: TX-CNN: Detecting tuberculosis in chest X-ray images using convolutional neural network. In: 2017 IEEE International Conference on Image Processing (ICIP). pp. 2314–2318. IEEE Press, Beijing, China (Sep 2017). https://doi.org/10.1109/ICIP.2017.8296695

9. Peng, S., Liu, Y., Lv, W., Liu, L., Zhou, Q., Yang, H., Ren, J., Liu, G., Wang, X., Zhang, X., Du, Q., Nie, F., Huang, G., Guo, Y., Li, J., Liang, J., Hu, H., Xiao, H., Liu, Z., Lai, F., Zheng, Q., Wang, H., Li, Y., Alexander, E.K., Wang, W., Xiao, H.: 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 (Apr 2021). https://doi.org/10.1016/S2589-7500(21)00041-8
10. Ross, D.S.: Predicting Thyroid Malignancy. The Journal of Clinical Endocrinology & Metabolism **91**(11), 4253–4255 (Nov 2006). https://doi.org/10.1210/jc.2006-1772
11. Sun, J., Wu, B., Zhao, T., Gao, L., Xie, K., Lin, T., Sui, J., Li, X., Wu, X., Ni, X.: Classification for thyroid nodule using ViT with contrastive learning in ultrasound images. Computers in Biology and Medicine **152**, 106444 (Jan 2023). https://doi.org/10.1016/j.compbiomed.2022.106444
12. Tessler, F.N., Thomas, J.: Artificial Intelligence for Evaluation of Thyroid Nodules: A Primer. Thyroid® **33**(2), 150–158 (Feb 2023). https://doi.org/10.1089/thy.2022.0560, publisher: Mary Ann Liebert, Inc., publishers
13. Thingiverse.com: Kevin's Thyroid by j0z, https://www.thingiverse.com/thing:3435530
14. 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**, 101665 (Apr 2020). https://doi.org/10.1016/j.media.2020.101665
15. Wang, Y., Choi, E.J., Choi, Y., Zhang, H., Jin, G.Y., Ko, S.B.: Breast Cancer Classification in Automated Breast Ultrasound Using Multiview Convolutional Neural Network with Transfer Learning. Ultrasound in Medicine & Biology **46**(5), 1119–1132 (May 2020). https://doi.org/10.1016/j.ultrasmedbio.2020.01.001
16. Wang, Y., Li, Z., Cui, X., Zhang, L., Luo, X., Yang, M., Chang, S.: Key-frame Guided Network for Thyroid Nodule Recognition Using Ultrasound Videos. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. pp. 238–247. Lecture Notes in Computer Science, Springer Nature Switzerland, Cham (2022). https://doi.org/10.1007/978-3-031-16440-8_23
17. Wang, Z., Liu, C., Cheng, D., Wang, L., Yang, X., Cheng, K.T.: Automated Detection of Clinically Significant Prostate Cancer in mp-MRI Images Based on an End-to-End Deep Neural Network. IEEE Transactions on Medical Imaging **37**(5), 1127–1139 (May 2018). https://doi.org/10.1109/TMI.2017.2789181, conference Name: IEEE Transactions on Medical Imaging
18. Yamashita, R., Kapoor, T., Alam, M.N., Galimzianova, A., Syed, S.A., Ugur Akdogan, M., Alkim, E., Wentland, A.L., Madhuripan, N., Goff, D., Barbee, V., Sheybani, N.D., Sagreiya, H., Rubin, D.L., Desser, T.S.: Toward Reduction in False-Positive Thyroid Nodule Biopsies with a Deep Learning–based Risk Stratification System Using US Cine-Clip Images. Radiology: Artificial Intelligence **4**(3), e210174 (May 2022). https://doi.org/10.1148/ryai.210174
19. Yan, Z., Yang, B., Dong, F., Yan, M., Xu, D., Zhu, C., Cheng, G., He, L.: Ultrasound-based thyroid follicular neoplasm classification using fine-tuning and graph convolutional network. In: Twelfth International Conference on Graphics and Image Processing (ICGIP 2020). vol. 11720, pp. 493–499. SPIE (Jan 2021). https://doi.org/10.1117/12.2589377
20. Yin, S., Peng, Q., Li, H., Zhang, Z., You, X., Liu, H., Fischer, K., Furth, S.L., Tasian, G.E., Fan, Y.: Multi-instance Deep Learning with Graph Convolutional Neural Networks for Diagnosis of Kidney Diseases Using Ultrasound Imaging. Uncertainty for Safe Utilization of Machine Learning in Medical Imaging and Clinical

Image-based Procedures : first International Workshop, UNSURE 2019, and 8th International Workshop, CLIP 2019, held in conjunction with MICCAI 2019, Sh. **11840**, 146–154 (Oct 2019). https://doi.org/10.1007/978-3-030-32689-0_15

21. Zhao, S.X., Chen, Y., Yang, K.F., Luo, Y., Ma, B.Y., Li, Y.J.: A Local and Global Feature Disentangled Network: Toward Classification of Benign-Malignant Thyroid Nodules From Ultrasound Image. IEEE Transactions on Medical Imaging **41**(6), 1497–1509 (Jun 2022). https://doi.org/10.1109/TMI.2022.3140797, conference Name: IEEE Transactions on Medical Imaging
22. Zhu, X., Feng, Q.: MVC-NET: Multi-View Chest Radiograph Classification Network With Deep Fusion. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). pp. 554–558. IEEE, Nice, France (Apr 2021). https://doi.org/10.1109/ISBI48211.2021.9434000
23. Zhuang, L., Ivezic, V., Feng, J., Shen, C., Radhachandran, A., Sant, V., Patel, M., Masamed, R., Arnold, C., Speier, W.: Patient-level thyroid cancer classification using attention multiple instance learning on fused multi-scale ultrasound image features. AMIA Annual Symposium Proceedings **2023**, 1344–1353 (Jan 2024)