



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Dual-Modality Watershed Fusion Network for Thyroid Nodule Classification of Dual-View CEUS Video

Rui Li¹[0000-0003-0635-0549], Jingliang Ruan², and Yao Lu¹

¹ School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510275, China

lirui256@mail2.sysu.edu.cn; luyao23@mail.sysu.edu.cn;

² Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China
ruanj13@mail.sysu.edu.cn

Abstract. Contrast-enhanced ultrasound (CEUS) allows real-time visualization of the vascular distribution within thyroid nodules, garnering significant attention in their intelligent diagnosis. Existing methods either focus on modifying models while neglecting the unique aspects of CEUS, or rely only single-modality data while overlooking the complementary information contained in the dual-view CEUS data. To overcome these limitations, inspired by the CEUS thyroid imaging reporting and data system (TI-RADS), this paper proposes a new dual-modality watershed fusion network (DWFN) for diagnosing thyroid nodules using dual-view CEUS videos. Specifically, the method introduces the watershed analysis from the remote sensing field and combines it with the optical flow method to extract the enhancement direction feature mentioned in the CEUS TI-RADS. On this basis, the interpretable watershed 3D network (W3DN) is constructed by C3D to further extract the dynamic blood flow features contained in CEUS videos. Furthermore, to make more comprehensive use of clinical information, a dual-modality 2D and 3D combined network, DWFN is constructed, which fuses the morphological features extracted from US images by InceptionResNetV2 and the dynamic blood flow features extracted from CEUS videos by W3DN, to classify thyroid nodules as benign or malignant. The effectiveness of the proposed DWFN method was evaluated using extensive experimental results on a collected dataset of dual-view CEUS videos for thyroid nodules, achieving an area under the receiver operating characteristic curve of 0.920, with accuracy, sensitivity, specificity, positive predictive value, negative predictive value, F1 score of 0.858, 0.845, 0.872, 0.879, 0.837, and 0.861, respectively, outperforming other state-of-the-art methods.

Keywords: Thyroid cancer · Ultrasound · Contrast-enhanced ultrasound (CEUS) · Watershed analysis.

1 Introduction

Thyroid cancer has become the fastest-growing tumor in women globally, and the key to its prevention and treatment lies in early screening [2, 9], which is mainly

conducted by grayscale ultrasound (US), supplemented by contrast-enhanced ultrasound (CEUS), doppler imaging, etc [19]. Among them, compared with the traditional US, which focuses more on morphological information (such as shape, boundary, aspect ratio, and size) [3], CEUS plays a crucial role in the diagnosis of thyroid nodules by observing the microvascular blood flow in the lesion area and the relative echo intensity changes with surrounding tissues [23]. Moreover, the CEUS thyroid imaging reporting and data system (TI-RADS) [18] pointed out that the combined use of CEUS and US had a positive effect on the risk stratification of thyroid nodules compared with the employ of US alone, helping to further improve the diagnostic performance.

Currently, the diagnosis of thyroid cancer mainly relies on radiologists observing CEUS videos back and forth and making subjective diagnoses based on their professional knowledge, which is not only time-consuming and laborious but also prone to misdiagnosis due to the contrast agent entering and disappearing within a few seconds [25]. These limitations have promoted the development of computer-assisted systems based on artificial intelligence (AI) and have made certain progress in clinical applications [8]. However, most research on the success of AI algorithms focuses on improvements at deep learning networks, ignoring the uniqueness of CEUS in clinical diagnosis. Additionally, radiologists usually make judgments based on dual-view data in clinical examinations, because US is needed to locate the nodule before the contrast agent enters, and then the blood flow movement information around and inside the nodule is observed during the entry of the contrast agent for diagnosis [12].

To better align with clinical diagnosis practices, based on the clinical CEUS TI-RADS, this study introduced watershed analysis from remote sensing, and combined dual-modality (US and CEUS) to identify the benign and malignant thyroid nodules. The main contributions are as follows: (i) We introduced the concept of watershed analysis from remote sensing field, combined with optical flow methods, to propose a watershed-based method for extracting the enhancement direction (ED) mentioned in the CEUS TI-RADS. (ii) Based on the clinical features ED, for CEUS videos, we constructed a 3D network fused watershed analysis to extract deep learning features, thereby capturing the blood flow distribution and contrast agent motion information contained in CEUS. (iii) To better align with the clinical diagnostic process, we combined dual-modality to construct a 2D+3D network, DWFN, which extracts morphological features from US images and dynamic blood flow information from CEUS videos to improve the performance of thyroid cancer classification.

2 Related Work

Related work based on US. According to the latest research results, the accuracy of AI diagnosis of thyroid cancer on US has reached 85%-90% [10, 11]. Early limitations in ultrasound technology were evident in most studies based on single grayscale US images [5]. With the continuous development of various new US technologies, many scholars began to integrate multi-modality

US (such as elastography, color Doppler, CEUS, etc.) for research, achieving good classification results [16, 24]. However, such methods are affected by the clarity of thyroid US images, and because they are carried out on static images, they may overlook many effective features and rely on the accuracy of radiologists capturing static images.

Related work based on CEUS. Compared with static US analysis, AI diagnosis of CEUS videos is more challenging, as it necessitates the inclusion of temporal features besides the spatial features of the original 2D image [15]. Early studies focused on extracting quantitative functional parameters from regional time intensity curves [17] or factor curves obtained through matrix decomposition [7]. However, these parameters can only provide limited functional information, overlooking the rich morphological features of tumor vascularization [21]. Subsequently, researchers developed deep learning models based on videos, resulting in various variants of 3D CNN: I3D [4], R (2+1) D [22], Vit3D [6], etc. However, most current research focuses solely on the deep learning features, which is a black box, thereby neglecting their significant significance in clinical diagnosis. Therefore, developing interpretable deep learning models that integrate the blood flow information contained in CEUS videos poses a critical challenge. Additionally, single modality information has inherent limitations, whereas dual-modality approaches can leverage information complementarity to offer more comprehensive features and improve diagnostic accuracy.

3 Method

As shown in Fig.1, the proposed dual-modality watershed fusion network (DWFN) primarily comprises preprocessing stage, US branch, CEUS branch, and feature fusion stage.

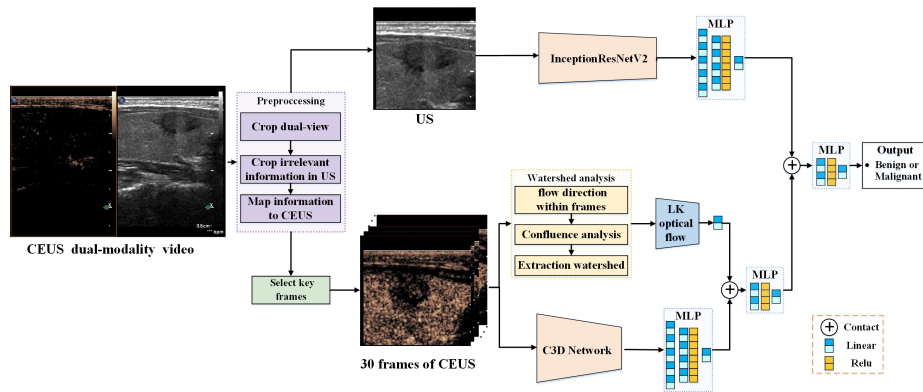


Fig. 1. Proposed framework of the Dual-modality Watershed Fusion Network (DWFN).

Preprocessing Stage. As the collected clinical data constitutes a dual-view video with a one-to-one positional relationship between US and CEUS, our initial step involves cropping all dual-view data along the central axis. Given the minimal changes in the US video during the examination, an experienced radiologist cropped the irrelevant information on the initial frame of the US video. This marked frame is directly used as input for the subsequent US branch and is mapped to each frame of the CEUS video. Additionally, considering inconsistent frame rates (ranging from 100 to 3000 frames) and redundancy between frames, we calculated the pixel intensity of all frames. We established a first-order derivative relationship between intensity and time order, automatically selecting 30 frames as input for subsequent CEUS branches.

US Branch. As clinicians focus more on morphological information in US, and US videos remain largely unchanged during acquisition, this study utilizes only the preprocessed US images as input for the subsequent model to improve computational efficiency. This module uses a two-dimensional convolutional neural network based on InceptionResNetV2 [20] to extract features and employs an MLP classifier [1] to filter the extracted features. Specifically, InceptionResNetV2 consists of 1 Inception-A module (5 3x3 convolutional layers and 2 3x3 max pooling layers), 5 Inception-B modules, Reduction-A, 10 Inception-C modules, Reduction-B, 5 Inception-D modules, Reduction-C, and a 2x2 average pooling layer. All Conv layers have BatchNorm and ReLu activation. The number of channels from Conv1 to Conv5 are 32, 32, 64, 80, and 192, respectively. The 1536-dimensional features obtained from the average pooling of InceptionResNetV2 are then reduced to $d1 = 2$ dimensions using a three-layer MLP to generate a fixed-length feature while enhancing network performance.

CEUS Branch. To investigate the distribution characteristics of contrast agent in CEUS, this study utilized preprocessed 30 frames of CEUS videos. We introduced a watershed analysis [13] from the remote sensing field, calculated the contrast agent flow rate per frame, and combined with the Lucas-Kanade optical flow method [14] to estimate the motion information of the contrast agent between adjacent frames. Furthermore, we extracted the clinical features $d2$ related to enhancement direction (ED) of the contrast agent, as mentioned in CEUS TI-RADS. Detailed information can be found in the supplementary. Based on the clinical features, ED, dynamic blood flow features were extracted from CEUS videos using C3D [21], which consists of 5 convolutional layers and 5 pooling layers (each convolutional layer is followed by a pooling layer), aimed at extracting dynamic information from CEUS videos. The number of filters in the 5 convolutional layers from 1 to 5 are 64, 128, 256, 256, and 256, respectively. Similarly, using an MLP classifier, the 8192 dimensional features extracted by C3D are reduced to $d3 = 2$ dimensions.

Feature Fusion Stage. The DWFN first fused the CEUS features by concatenating the ED features ($d2$) with the spatiotemporal features ($d3$) extracted by

C3D, and then concatenates the new CEUS features (d_{new}) with the US features ($d1$). Subsequently, an MLP is used to reduce the concatenated features, ultimately achieving the classification of benign and malignant thyroid nodules.

4 Results and Discussion

This paper retrospectively collected thyroid data from 986 patients from Sun Yat-sen Memorial Hospital, Sun Yat-sen University, including 471 benign cases and 515 malignant cases. The dataset was randomly divided into 3 groups in a ratio of 6:2:2 using three-fold cross-validation approach. Further implementation details can be found in the supplementary material. To evaluate the effectiveness of our proposed method, a series of comparative and ablation experiments were conducted.

4.1 Analysis of Classification Performance

Table 1 displays the classification results of various modalities (single modality versus our proposed DWFN). The results indicate that the combined dual-modality $M_{US+CEUS}$ is superior to single-modality based solely on US and CEUS (with AUC values of 0.888 compared to 0.867 and 0.851). Furthermore, the classification performance of M_{US} is higher than that of M_{CEUS} (1.6% higher AUC), which aligns with the fact that radiologists primarily utilize US for thyroid cancer diagnosis. However, this does not undermine the significance of characteristics in CEUS, on the contrary, the blood flow information contained within CEUS videos is crucial for clinical diagnosis. The results of the CEUS modality employing watershed analysis demonstrate improvements in AUC, Accuracy, Sensitivity, Specificity, PPV, NPV, and F1 by 6% (the p value is $<.001$), 5.6%, 1%, 10.6%, 8.5%, 2.8%, and 4.7%, respectively, compared to the CEUS modality without watershed analysis. These findings validate that the ED extracted by watershed analysis can quantify the direction of contrast agent entry into blood vessels, thereby providing clinically relevant key features for the final thyroid cancer classification task. Fig.2 (a) depicts the receiver operating characteristic curve of the corresponding AUC value, with the proposed DWFN achieving the optimal AUC value of 92.0%. These results collectively confirm that the DWFN can effectively learn dual-modality features and greatly improve accuracy.

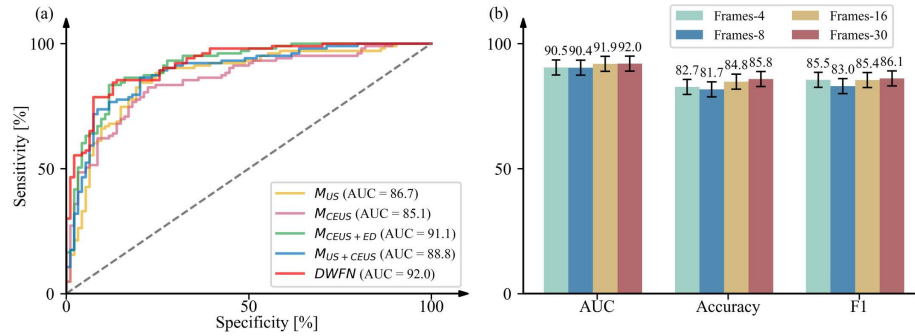
4.2 Ablation experiment

Comparison with Different US Feature Extractors A critical component of our proposed DWFN is the US feature extraction network, InceptionResNetV2, which is crucial for the final diagnosis of benign and malignant thyroid nodules, as clinicians focus more on the morphological features. For this reason, this study employed US images as model input instead of US videos. Table2 provided a quantitative comparison of US features extracted by different 2D networks in thyroid cancer classification. It can be observed that the

Table 1. The comparison of classification performance under different modality data.

Method	AUC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	F_1
M_{US}	86.0* (81.4-91.6)	82.2 (78.2-88.3)	84.5 (74.2-94.2)	79.8 (68.3-90.5)	82.1 (74.0-90.5)	82.4 (74.6-92.6)	83.3
M_{CEUS}	85.1* (79.4-90.0)	80.2 (75.1-86.3)	82.5 (65.0-89.2)	77.7 (71.0-93.0)	80.2 (74.1-91.9)	80.2 (69.7-88.1)	81.3
$M_{CEUS+ED}$	91.1 (87-94.8)	85.8 (81.7-90.9)	83.5 (77.6-95.0)	88.3 (74.5-94.0)	88.7 (79.3-94.3)	83.0 (77.1-93.7)	86.0
$M_{US+CEUS}$	88.8* (84.1-93.2)	83.2 (79.2-88.3)	86.4 (69.9-93.9)	79.8 (72.2-95.7)	82.4 (76.9-95.4)	84.3 (71.8-92.2)	84.4
DWFN	92.0 (88.2-95.6)	85.8 (81.7-91.4)	84.5 (74.3-92.5)	87.2 (80.8-96.8)	87.9 (81.7-96.6)	83.7 (75.9-92.0)	86.1

Note: All indicators are in %, AUC = area under the receiver operating characteristic curve, CI = confidence interval, PPV = positive predictive value, NPV = negative predictive value, * indicates that the P-value is less than 0.05 compared with DWFN.

**Fig. 2.** (a) Areas under the receiver operating characteristic curves (AUCs) of different methods in Table1, (b) Evaluation comparing different numbers of CEUS frames.

overall classification performance AUC of 7 different 2D networks (including InceptionResNetV2 used in DWFN) is in the range of [86.3%-92.0%], and the performance differences are not significant. Among them, the proposed DWFN achieves the best classification performance in terms of AUC, Accuracy and F1, among which, although the Sensitivity of ResNet50, ResNet101 (86.4%, 85.4%) is slightly higher (1.9%, 0.9%) than that of DWFN, their Specificity is much lower (6%, 7.4%) than DWFN. This is mainly due to the fact that Sensitivity and Specificity are a set of trade-off evaluation indicators. In comparison, DWFN has the optimal Specificity while maintaining suboptimal Sensitivity, which can better extract morphological features of US, and achieve the best classification performance due to its combination of Residual block and Inception structure.

Table 2. Comparison of different backbone 2D feature extraction network of US images.

DWFN-US Branch	AUC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	F_1
AlexNet	89.4 (84.7-93.4)	82.7 (78.7-88.3)	83.5 (72.6-98.0)	81.9 (64.9-92.7)	83.5 (73.5-92.4)	81.9 (72.9-96.9)	83.5
VGG19	91.0 (88.2-95.2)	85.3 (82.2-90.9)	82.5 (78.8-96.0)	88.3 (75.0-93.5)	88.5 (79.2-93.9)	82.2 (77.8-95.0)	85.4
GoogleNet	89.1 (84.7-93.3)	81.7 (78.2-87.8)	81.6 (71.7-94.3)	81.9 (69.1-93.1)	83.2 (75.6-92.9)	80.2 (72.3-92.8)	82.4
MobileNet	90.5 (86.1-94.4)	83.2 (80.2-89.3)	80.6 (76.9-98.9)	86.2 (66.0-92.5)	86.5 (74.0-92.7)	80.2 (76.4-98.4)	83.4
ResNet34	86.3 (81.4-91.0)	80.7 (76.6-86.8)	84.5 (73.8-92.7)	76.6 (67.0-88.5)	79.8 (73.2-88.8)	81.8 (72.9-90.5)	82.1
ResNet50	90.9 (86.6-94.8)	84.3 (80.2-89.8)	86.4 (77.6-95.6)	81.9 (71.7-91.7)	84.0 (76.2-92.1)	84.6 (76.7-94.7)	85.2
ResNet101	88.6 (83.7-92.8)	82.7 (78.7-88.3)	85.4 (75.2-94.3)	79.8 (72.1-92.3)	82.2 (76.2-92.1)	83.3 (74.6-93.2)	83.9

Comparison with Different CEUS Feature Extractors In addition, another key component of our proposed method is the CEUS feature extraction network, C3D. Since clinicians pay more attention to the blood flow changes during the contrast agent entry process, it is critical to extract deep learning features from CEUS videos using appropriate classifiers in addition to using watershed analysis to extract the ED feature. Table 3 provides a quantitative comparison of different CEUS feature extractors in thyroid cancer classification. Note that TRN based on 2D network is significantly lower than other models based on 2+1D networks and 3D networks (AUC is at least 19.8% lower), which may indicate that TRN, using 2D convolution, cannot capture both blood flow distribution and dynamic information simultaneously. Additionally, besides the

C3D used in DWFN, the I3D exhibits the best classification performance. This is attributed to its use a dual-stream CNN with dilated 3D convolution on dense RGB and optical flow sequences. The dual-stream CNN captures dynamic information, aligning with the concept of watershed analysis in this paper, further highlighting the superiority of DWFN.

Table 3. Comparison of different CEUS feature extraction of CEUS videos

DWFN- CEUS branch	AUC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	F_1
TRN	65.9 (59.4-73.1)	68.5 (62.9-75.6)	93.2 (85.4-98.1)	41.5 (31.3-53.7)	63.6 (56.6-71.8)	84.8 (74.3-95.7)	75.6
I3D	91.7 (87.6-95.2)	84.3 (81.2-90.4)	79.6 (74.3-93.4)	89.4 (75.8-95.8)	89.1 (79.2-95.7)	80.0 (74.6-92.5)	84.1
P3D	88.5 (83.8-92.8)	79.2 (76.1-86.3)	75.7 (67.6-97.9)	83.0 (60.2-92.2)	83.0 (69.9-91.7)	75.7 (705-9.64)	79.2
R3D	89.6 (85.2-93.5)	83.8 (79.7-88.8)	96.1 (83.2-100)	70.2 (63.9-84.9)	78.0 (72.5-87.4)	94.3 (81.7-100)	86.1
R(2+1)D	89.4 (84.5-93.5)	81.7 (77.7-87.8)	83.5 (68.9-95.5)	79.8 (67.3-94.6)	81.9 (74.2-93.9)	81.5 (71.6-94.2)	82.7
S3D	91.2 (86.9-94.8)	84.3 (80.7-89.8)	82.5 (79.2-98.0)	86.2 (695-92.3)	86.2 (69.5-92.3)	81.8 (77.7-97.0)	84.6
Vit3D	87.7 (82.6-92.1)	79.2 (75.1-85.8)	95.1 (62.4-99.1)	61.7 (56.6-95.0)	73.1 (68.1-93.8)	92.1 (67.5-98.4)	82.7

Influence of the Number of Selected Key Frames. Furthermore, we investigated the impact of all frame numbers on the final classification. Specifically, we selected frames 4, 8, 16, 30 in an ordered and equidistant manner, presenting the corresponding results in shown in Fig.2 (b). Observing the figure, the AUC value performance of 4 frames is 0.1% higher than that of 8 frames, and the AUC performance of 16 frames is 0.1% lower than that of 30 frames. Therefore, in our implementation, selecting 30 key frames as the input of CEUS branch can capture the dynamic trend of blood flow to a large extent, achieving the highest classification performance.

5 Conclusion

This paper proposes a dual-modality watershed fusion network for distinguishing between benign and malignant thyroid cancer. We extensively evaluated the effectiveness of our method on our collected dataset of thyroid nodules, achieving optimal classification performance compared to both single-modality methodf and state-of-the-art methods without watershed analysis. Performance metrics

include an AUC of 0.92 (95% CI, 0.882-0.956), accuracy of 0.858 (95% CI, 0.817-0.914), sensitivity of 0.845 (95% CI, 0.743-0.925), specificity of 0.872 (95% CI, 0.808-0.968), PPV of 0.879 (95% CI, 0.817-0.966), NPV of 0.837 (95% CI, 0.759-0.920), and F1 score of 0.861. More importantly, the watershed analysis method we designed considers the direction of contrast agent movement in CEUS videos, leveraging clinical prior knowledge and integrating deep learning methods to construct a more interpretable model for thyroid cancer classification, ultimately improving classification performance.

Acknowledgments. This study was funded in part by the China Department of Science and Technology under Key Grant 2023YFE0204300, in part by the R&D project of Pazhou Lab (HuangPu) under Grant 2023K0606, in part by the NSFC under Grant 82371476, Grant 82371917, Grant 81801809, Grant 81830052, Grant 81827802, and Grant U1811461, in part by the China Department of Science and Technology under Key Grant 210YBXM 2020109002, in part by the Guangzhou Science and Technology bureau under Grant 2023B03J1237, in part by the Science and Technology Innovative Project of Guangdong Province under Grant 2018B030312002, in part by Health Research Major Projects of Hunan Health Commission under grant W20241010, in part by Guangdong Province Key Laboratory of Computational Science at the Sun Yat-sen University under grant 2020B1212060032, in part by Key-Area Research and Development Program of Guangdong Province under Grant 2021B0101190003.

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

References

1. Abbas, D.K., Rashid, T.A., Bacanin, K.H., Alsadoon, A.: Using fitness dependent optimizer for training multi-layer perceptron. arXiv preprint arXiv:2201.00563 (2022)
2. Cabanillas, M.E., McFadden, D.G., Durante, C.: Thyroid cancer. *The Lancet* **388**(10061), 2783–2795 (2016)
3. Cao, C.L., Li, Q.L., Tong, J., Shi, L.N., Li, W.X., Xu, Y., Cheng, J., Du, T.T., Li, J., Cui, X.W.: Artificial intelligence in thyroid ultrasound. *Frontiers in Oncology* **13**, 1060702 (2023)
4. Carreira, J., Zisserman, A.: Quo vadis, action recognition? a new model and the kinetics dataset. In: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6299–6308 (2017)
5. Chen, Y., Gao, Z., He, Y., Mai, W., Li, J., Zhou, M., Li, S., Yi, W., Wu, S., Bai, T., et al.: An artificial intelligence model based on acr ti-rads characteristics for us diagnosis of thyroid nodules. *Radiology* **303**(3), 613–619 (2022)
6. Chen, Z., Agarwal, D., Aggarwal, K., Safta, W., Balan, M.M., Brown, K.: Masked image modeling advances 3d medical image analysis. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1970–1980 (2023)
7. Feng, Y., Qin, X.C., Luo, Y., Li, Y.Z., Zhou, X.: Efficacy of contrast-enhanced ultrasound washout rate in predicting hepatocellular carcinoma differentiation. *Ultrasound in medicine & biology* **41**(6), 1553–1560 (2015)

8. Feng, Y., Yang, F., Zhou, X., Guo, Y., Tang, F., Ren, F., Guo, J., Ji, S.: A deep learning approach for targeted contrast-enhanced ultrasound based prostate cancer detection. *IEEE/ACM transactions on computational biology and bioinformatics* **16**(6), 1794–1801 (2018)
9. 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: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 665–674. Springer (2022)
10. Jeong, E.Y., Kim, H.L., Ha, E.J., Park, S.Y., Cho, Y.J., Han, M.: Computer-aided diagnosis system for thyroid nodules on ultrasonography: diagnostic performance and reproducibility based on the experience level of operators. *European radiology* **29**, 1978–1985 (2019)
11. Jin, Z., Zhu, Y., Zhang, S., Xie, F., Zhang, M., Guo, Y., Wang, H., Zhu, Q., Cao, J., Luo, Y.: Diagnosis of thyroid cancer using a ti-rads-based computer-aided diagnosis system: a multicenter retrospective study. *Clinical Imaging* **80**, 43–49 (2021)
12. Kondo, S., Takagi, K., Nishida, M., Iwai, T., Kudo, Y., Ogawa, K., Kamiyama, T., Shibuya, H., Kahata, K., Shimizu, C.: Computer-aided diagnosis of focal liver lesions using contrast-enhanced ultrasonography with perflubutane microbubbles. *IEEE transactions on medical imaging* **36**(7), 1427–1437 (2017)
13. Li, L., Yang, J., Wu, J.: A method of watershed delineation for flat terrain using sentinel-2a imagery and dem: A case study of the taihu basin. *ISPRS International Journal of Geo-Information* **8**(12), 528 (2019)
14. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: *IJCAI’81: 7th international joint conference on Artificial intelligence*. vol. 2, pp. 674–679 (1981)
15. Lueck, G.J., Kim, T., Burns, P.N., Martel, A.L.: Hepatic perfusion imaging using factor analysis of contrast enhanced ultrasound. *IEEE transactions on medical imaging* **27**(10), 1449–1457 (2008)
16. 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. *The Lancet Digital Health* **3**(4), e250–e259 (2021)
17. Rognin, N.G., Arditi, M., Mercier, L., Frinking, P.J., Schneider, M., Perrenoud, G., Anaye, A., Meuwly, J.Y., Tranquart, F.: Parametric imaging for characterizing focal liver lesions in contrast-enhanced ultrasound. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* **57**(11), 2503–2511 (2010)
18. Ruan, J., Xu, X., Cai, Y., Zeng, H., Luo, M., Zhang, W., Liu, R., Lin, P., Xu, Y., Ye, Q., et al.: A practical ceus thyroid reporting system for thyroid nodules. *Radiology* **305**(1), 149–159 (2022)
19. Strouthos, C., Lampaskis, M., Sboros, V., Mcneilly, A., Averkiou, M.: Indicator dilution models for the quantification of microvascular blood flow with bolus administration of ultrasound contrast agents. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* **57**(6), 1296–1310 (2010)
20. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Proceedings of the AAAI conference on artificial intelligence*. vol. 31 (2017)
21. Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3d convolutional networks. In: *Proceedings of the IEEE international conference on computer vision*. pp. 4489–4497 (2015)

22. Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., Paluri, M.: A closer look at spatiotemporal convolutions for action recognition. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 6450–6459 (2018)
23. Wan, P., Liu, C., Chen, F., Qin, J., Zhang, D.: Identifying quantitative and explanatory tumor indexes from dynamic contrast enhanced ultrasound. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24. pp. 638–647. Springer (2021)
24. Yang, W.T., Ma, B.Y., Chen, Y.: A narrative review of deep learning in thyroid imaging: current progress and future prospects. *Quantitative Imaging in Medicine and Surgery* **14**(2), 2069 (2024)
25. Zhang, C., Liu, D., Huang, L., Zhao, Y., Chen, L., Guo, Y.: Classification of thyroid nodules by using deep learning radiomics based on ultrasound dynamic video. *Journal of Ultrasound in Medicine* **41**(12), 2993–3002 (2022)