Machine Learning Assisted Doppler Features for Enhancing Thyroid Cancer Diagnosis: A Multi-cohort Study

Yi-Cheng Zhu, MD , Hongbo Du, MSc, MPhil, Quan Jiang, MD , Tao Zhang, PhD, Xu-Juan Huang, MD, Yuan Zhang, MD, Xiu-Rong Shi, MD, Jun Shan, MD., Alaa AlZoubi, PhD

Abstract

Background: This pilot study aims at exploiting machine learning techniques to extract colour Doppler Ultrasound (CDUS) features and to build an artificial neural network (ANN) model based on these CDUS features for improving the diagnostic performance of thyroid cancer classification.

Methods: A total of 674 patients with 712 thyroid nodules (TNs) (512 from in-ternal dataset and 200 from external dataset) were randomly selected in this retrospective study. We used ANN to build a model (TDUS-Net) for classifying malignant and benign TNs using both the automatically extracted quantitative CDUS features (whole ratio, intranodular ratio, peripheral ratio, and number of vessels) and grey-scale Ultrasound (US) features defined by the ACR Thyroid Imaging Reporting and Data System (TI-RADS). Then, we compared the diagnostic performance of the model, the performance of another ANN model based on the grey-scale US features alone (TUS-Net), and that of radiologists.

Results: The TDUS-Net (0.898, 95%CI: 0.868-0.922) achieved a higher area under the curve (AUC) than that of TUS-Net (0.881, 95%CI: 0.850-0.908) in the internal tests. Compared with radiologists, TDUS-Net (AUC: 0.925, 95% CI: 0.880-0.958) performed better than radiologists (AUC: 0.810, 95% CI: 0.749-0.862) in the external tests.

Conclusions: Applying a machine learning model by combining both gray-scale US features

and CDUS features can achieve comparable or even higher performance than radiologists in classifying thyroid nodules.

Keywords: Thyroid nodules; Ultrasound; Doppler Ultrasound; Machine Learning; Artificial Neural Network

Introduction

The growth and progression of a malignant tumour largely depend on its blood flows[1]. Angiogenesis occurs in the thyroid during disease processes, including goitres, Graves' disease, thyroiditis, and cancer[2]. The multiplicity of vessels and disordered patterns can be detected by non-invasive colour Doppler ultrasonography (CDUS). CDUS images contain blood flow information over the entire area of the grey-scale ultrasound (US) image. This technique can detect the abnormal vascularity associated with thyroid cancer, which ranks ninth in global cancer incidence[3]. However, as with the current clinical practice, colour signals in CDUS thyroid images can only be evaluated subjectively or semi-subjectively[4,5], which may lead to intra- and inter- observer variability that in turn has limited the wider use of CDUS as a routine clinical tool. Therefore, an automatic and accurate quantitative criterion of tumour vascularity analysis becomes crucial for accurate cancer diagnostics.

The development of computer-aided diagnosis (CAD) systems has arisen in recent decades as non-invasive methods to supplement radiologists' interpretation and over-come the subjective interpretation limitations[6]. Yoo et al. compared the diagnostic performance of an experienced radiologist, a CAD system alone, and the radiologist assisted by the CAD system in classifying thyroid cancer[7]. The results showed that visual interpretation by the radiologist had higher specificity (95.0%), while the sensitivity for both CAD and radiologist interpretation was comparable. However, Choi et al. evaluated the diagnostic performance of their CAD system and that of an experienced radiologist for differentiating malignant and benign thyroid nodules (TNs)[8]. They found that the CAD system, in general, achieved a slightly higher sensitivity (90.7%) than what the radiologist achieved (88.4%), whereas the radiologist achieved a higher

specificity (97.4%) than that of CAD (71.8%). Nevertheless, the diagnosis of TNs using machine learning methods needs constant improvement and further studies.

Currently, there are few research in investigating quantitative doppler features for up-lifting the diagnostic performance of thyroid cancer classification. Ardakani et al. evaluated seven gray-scale US features and five CDUS features to classify hot and cold thyroid nodules using support vector machine algorithms. However, one of the CDUS features (vascularity pattern) were assessed by full visual estimation, which may not precisely reveal the value of vascularity pattern. Therefore, as a pilot study, this paper analyses the CDUS thyroid images for accurate thyroid cancer classification through three steps. Firstly, we developed automatic methods to detect and extract different quantitative doppler features of the blood flow signals from CDUS images of TNs. Secondly, we used Artificial Neural Network (ANN) technique to build a model for malignant and benign TN classification using both the extracted doppler features and grey-scale US features indicated by the ACR Thyroid Imaging Reporting and Data System (TI-RADS) scores[9]. Finally, we evaluated the contribution (i.e. the added value) of the CDUS features in improving the diagnostic accuracy in differentiating malignant from benign TNs.

Materials and Methods

Study population

A total of 489 patients (381 female and 108 male) with 512 TNs from Pudong New Area People's Hospital (PNAPH) and a total of 93 patients with 100 TNs from Heqing Com-munity Healthcare Centre (HCHC) and 92 patients with 100 TNs from Jinyang Community Healthcare Centre (JYHC) were included in this retrospective multi-centre study. All the included nodules were selected on a random basis with 50% of benign TNs and 50% of malignant TNs. We used

the data from PNAPH as the internal dataset and the data from HCHC and JYHC as the external dataset. All the included TNs underwent fine-needle aspiration biopsy (FNAB) and/or surgery pathology in one month after the US examinations. TNs with benign cytologic results received follow-up for more than 6 months. Exclusion criteria applied include 1) nodules with cytologic results of equal to or higher than Category 4 Bethesda grading and without pathology confirmation after surgery; 2) lack of a complete record of conventional US images and/or colour Doppler images; and 3) nodules with Hashimoto's disease. The patient selection workflow is shown in Figure 1. The mean age of the patients in the internal dataset and external dataset was 52.28 ± 12.48 years (29-85 years) and 53.31 ± 12.74 years (29-85 years) respectively. The ethics committees of the centres involved (PNAPH, HCHC and JYHC) approved this study and waived the requirement for written informed consent due to the nature of the retrospective study. The ethics committee of the University of Buckingham also approved this pilot study with the centres as the third-party data provider.

Scanning technique

US machines of four different brands (i.e., Siemens, GE, Philips, and Toshiba) were used for the acquisition of the internal and external datasets. Grey-scale US and CDUS examinations were performed with the same machines. A 5-15 MHz linear transducer was applied for both grey-scale imaging and CDUS imaging. All examinations were conducted by the same radiologist with 5 years of experience in thyroid grey-scale US and CDUS for the internal dataset. The examinations of the external datasets were conducted by one radiologist with more than 15 years of thyroid ultrasound experience in HCHC and one radiologist with more than 15 years of thyroid ultrasound experience in JYHC respectively.

All patients were asked to lie in a supine position, with their necks slightly extended. Transverse and longitudinal US images were acquired for every TN. Grey-scale US features, such as size, shape, echogenicity, margin, calcifications, halo sign, and composition, were captured and assessed. Two radiologists with over 10 years of experience in thyroid US imagery graded each TN in consensus based on the ACR TI-RADS classification scheme for both internal and external datasets. In cases of disagreement, a senior radiologist with enriched experience (longer than 15 years) in thyroid US imagery was consulted, and the final TI-RADS classification was agreed upon. The CDUS settings were chosen to optimize sensitivity to low-velocity and low-volume blood flow signals. The region of interest (ROI) on each US image was scanned slowly with minimum probe pressure. All the US and CDUS images were recorded and stored in our internal database in JPG format.

Automatic Colour Doppler Feature Extraction

We stored all US thyroid images in our internal database for analysis. All CDUS measurements were limited to the ROIs of the examined TNs. A software tool written in MATLAB (version 2019b; MathWorks Inc., Natick, MA) was developed to support this study. The tool has two main functions: (a) to enable manual cropping and segmentation of ROI, and (b) to quantify the Doppler signals within the ROI. The ROI was extracted by cropping the nodular region from the CDUS image for an accurate ratio estimation. We developed a free-hand cropping tool for function (a) that enables the radiologist (user) to locate and click, on the edge of the nodule boundary, a collection of points $\boldsymbol{X'} = \{(\boldsymbol{x'}, \boldsymbol{y'})_{\boldsymbol{1}}, (\boldsymbol{x'}, \boldsymbol{y'})_{\boldsymbol{2}}, \dots, (\boldsymbol{x'}, \boldsymbol{y'})_{\boldsymbol{m}}\}$ to extract the ROI. Using the software tool, one radiologist with more than 5 years' experience in thyroid US cropped all ROIs manually. Afterwards, a senior radiologist with more than 15 years of

experience in thyroid imaging checked the cropped images.

Vascularity Ratio Estimation

The second main function (b) extracts the "ratio of vascular flow areas" within the ROI of each CDUS image through the following steps. First, pixels of various colours are detected and segmented from grey-scale pixels. The CDUS image contains blood flow information in terms of pixels of colours made from the three primary colours: red, blue, and green. We studied the intensity values in the three channels (RGB) of each pixel at the identified points located inside or on the edge of the nodule region and then identified the coloured pixels. The original CDUS image was read in RGB format as $\boldsymbol{I} = \{\boldsymbol{R}, \boldsymbol{G}, \boldsymbol{B}\}$. The coloured pixel has an inequivalent intensity distribution across the RGB channels. That enabled us to subtract the intensity values of a pixel in the red and blue channels from the value in the green channel in determining whether their difference was large enough to be considered as a coloured pixel. This analysis was run on every pixel $I_{(x,y)}$ on the studied CDUS image I within the identified ROI as

$$\boldsymbol{isColoured}(\boldsymbol{I}_{(x,y)}) := \begin{cases} \mathbf{1} & \boldsymbol{if}\ \left|\boldsymbol{G}_{(x,y)} - \boldsymbol{R}_{(x,y)}\right| > \boldsymbol{thr} \\ \mathbf{1} & \boldsymbol{if}\ \left|\boldsymbol{G}_{(x,y)} - \boldsymbol{B}_{(x,y)}\right| > \boldsymbol{thr} \\ \mathbf{0} & \boldsymbol{else} \end{cases}$$

where thr referred to a threshold that defined the minimum tolerance on the colour difference and can be defined empirically (thr is set to 48 in the present study).

A bit-map can be derived from this function which determines the coloured areas within the ROI. We have further performed closing morphological operation with a 2-pixel disk structuring element in closing the narrow bright boundaries caused by the blood flow vortex. The flag value of each pixel within the ROI after the closing morphological operation was denoted as $isColoured(x, y)'$. As we marked X_{whole} as a set of all pixel coordinates within the

ROI, the total vascularity ratio (VR_{whole}) was eventually calculated as the ratio of the coloured area against the entire ROI area.

$$VR_{\text{whole}} = \frac{\sum_{(x,y)\in X_{\text{whole}}} isColoured(I_{(x,y)})'}{|X_{\text{whole}}|}$$

Estimation of Vascularity Location (peripheral and intra-nodular areas)

The correlation between increased central or peripheral vascularity and thyroid cancer is somehow controversial. Some researchers suggested that increased central vascularity is a supporting feature of thyroid cancer[10-12], whereas others did not find any relationship between intra-nodular blood flow signals and thyroid malignancy[13,14]. Therefore, it is interesting to investigate whether useful features regarding blood flow locations may provide any added value in identifying thyroid cancers. To determine the blood flow in different regions of each TN, the primary ROI was then divided into peripheral and intra-nodular areas. We proposed an objective method to determine the peripheral and intra-nodular regions by defining appropriate "offsets". "Offsets" is adopted to adjust the region areas without distorting the overall shape and contour of the ROI. The inner part refers to the core region (called as "intra-nodular area") of the primary ROI, while the outer part is the peripheral of the primary ROI (called as "peripheral area") (Figure 2). The offset changed as a percentage of the largest diameter of the primary ROI. In this study, we tried a range of offsets in 1% increments. In principle, the percentage of offsets could be set from 1% to 99%. However, we did not exceed the offsets of 25% because previous study indicated that a 10% offset is already sufficiently indicative[15]. Nevertheless, we conducted this sensitivity analysis to evaluate the existing claim and probe the optimal offset for this study. At each offset level, the whole ratio (percentage of blood flow in an entire nodule), intranodular ratio (rate of blood flow in the intranodular area of a nodule), and peripheral ratio

(rate of blood flow in the peripheral area of a nodule) were measured. The final optimum offsets (21%) were determined by using Mann-Whitney analysis when all P values of whole ratio, intranodular ratio, and peripheral ratio resulted in the most significant difference ($P < 0.05$) between benign and malignant TNs.

Vascularity Number Estimation

Vascularity density is considered to have a high correlation with tumour angiogenesis[16,17]. Therefore, whether the density of blood flow signals within the thyroid nodules is a significant influencer of malignancy is under our investigation. We introduced one final CDUS feature, called the "number of vascularity". We calculated the number of the connected components on the bit-map of detected blood flow areas, with a connectivity of 8. This feature reflects various ways blood flow activities are distributed within the nodule area even it is quite constrained within a 2D slide image.

Eventual Doppler Feature Vector

We compose the colour Doppler features: i.e. whole ratio, intranodular ratio, peripheral ratio, and number of vascularity as explained above, into a combined feature vector of four dimensions. In the experimental studies to be presented later, we further combine these features with TI-RADS features for building an enhanced classification model.

Developing ANN-based Thyroid Cancer Recognition Models

In this study, we used a neural network classifier for malignant and benign thyroid nodule recognition. Before building the network, we adopted data normalisation as follows. Each of the dimensions in the feature derived was normalised into a range of [0,1] by using the Min-Max rescaling (division by range). The general normalisation formula is:

$$z' = \frac{z - z_{min}}{z_{max} - z_{min}}$$

where z and z' denote the feature value before and after the normalisation in respective and $[\boldsymbol{z_{min}}, \boldsymbol{z_{max}}]$ denotes the range of the feature. As both the TI-RADS scores and the vascularity features were designed to have a minimum value of 0, the formula was further simplified as $\boldsymbol{z' = \frac{z}{z_{max}}}$.

The network, by nature, acted as a regression solver in finding the best fitting function from the input neurons to the output neurons. We adopted a shallow feedforward network with tan-sigmoid transformation functions in the hidden layers and linear transformation function in the output layer. The generic ANN architecture consisted of an input layer of the size L (where L = 5, 6, 7, or 9, see later), a hidden layer with 10 neurons and an output layer of two neurons which produced the sigmoid likelihoods of the nodule for being malignant and benign, respectively. We used the skilled conjugated gradient back-propagation method, the Gauss-Newton algorithm, and gradient descent to train the network thyroid models.

Using the generic architecture as designed earlier, we further trained two network models for comparing the proposed CDUS features against the gray-scale US ACR TI-RADS features. The first network model (TUS-Net) adopted the 5 normalized TI-RADS scores (based on radiologists' interpretations of the US images) as the inputs. The second network model (TDUS-Net) took the combination of the 5 normalized TI-RADS scores and the 4 CDUS features (Figure 3) as the input. Figure 4 outlines the network model structures. We used 10-fold cross-validation in analysing the performance of the classifier. In other words, the given data set was divided into 90% for training and validation and 10% for testing for each fold.

Statistical Analysis

Statistical analysis was performed by using SPSS software (version 24.0; SPSS, Chicago). Grey-scale US characteristics and vascularity variances between malignant and benign TNs were evaluated using the chi-square test or Student's t-test. Categorical variables were compared using the nonparametric Wilcoxon Mann-Whitney U test, and continuous variables were compared using the independent t-test. Receiver operating characteristic (ROC) analysis was used to obtain the cut-off value of the diagnostic performance of radiologists, TUS-Net and TDUS-Net. The statistical differences between the diagnostic performance of radiologists, TUS-Net and TDUS-Net was evaluated by DeLong Test. Significance was defined as $P < 0.05$.

Results

Baseline and B-mode ultrasonic features

Of the 512 nodules, 256 (50.0%) were benign, and 256 (50.0%) were malignant. There was no significant difference in sex distribution between patients with benign and malignant TNs ($P = 0.20$). The average diameter was 15.75 ± 8.41 mm (mean ± SD; range, 3.2-69.3 mm) for the internal dataset, whereas the average diameter was 15.50 ± 8.66 mm (range, 2.8-45.29 mm) for the external dataset. There was no significant difference in size between the 2 groups (internal dataset: $P = 0.051$; external dataset: $P = 0.147$). Lobulated or irregular margins were more common in malignant TNs, which had a solid or almost completely solid composition and a taller-than-wide shape ratio ($P < 0.001$ for all) (Table 1). Benign TNs tended to have smooth margins and wider-than-tall shape ratio ($P < 0.001$ for all).

Doppler features

The blood flow whole ratio and ratio of vascularity at different locations of the 512 nodules and 200 external test nodules were reported in Table 2. Objectively, 3% of external test benign

nodules and 15.2% of 512 benign nodules showed no detectable flow, whereas 7% of external test malignant nodules and 12.1% of malignant thyroid nodules were absent with blood flow signals. The ratio of displayed vascularity was significantly different between benign and malignant TNs. In the internal dataset, the mean blood flow whole ratio of benign nodules was 12.51±16.82, whereas the mean blood flow whole ratio of malignant nodules was 5.03±8.78 ($P<0.001$) (Table 2). The intranodular ratio of benign thyroid nodules (11.12±15.41) was higher than that of malignant thyroid nodules (4.45±8.01) ($P<0.001$). There was also a significant difference in the peripheral ratio be-tween benign and malignant TNs ($P<0.001$). In the external dataset, the whole ratio of benign nodules (20.58±21.69) were more than two times higher than that of malignant nodules (8.55±11.88) ($P<0.001$). Both the intranodular ratio and peripheral ratio of benign nodules were significantly higher than that of malignant nodules. (all $P<0.001$) (Table 2). The ratio of vascularity displayed was not correlated with nodule size in either malignant TNs ($R=-0.069$, $P=0.269$) or benign TNs ($R=-0.044$, $P=0.484$).

There was a negative relation between the number of vessels and thyroid cancer ($R=-0.92$, $P=0.037$). The more vascularity a thyroid nodule had, the less likely it was a malignant thyroid nodule (Figure 5). The mean number of vascularity in 256 malignant nodules and 256 benign nodules was 5.17±5.78 and 16.26±8.43, respectively. There was also a negative correlation between the number of vascularity and malignant TNs ($R=-0.182$, $P=0.010$) in the external dataset. The mean number of vascularity in 100 malignant nodules and 100 benign nodules was 5.43±5.00 and 7.86±7.14. Both results confirmed a statistically sig-nificant difference regarding the number of vessels between malignant and benign TNs (internal dataset: $P=0.037$, external dataset: $P=0.011$).

Comparison of the diagnostic performance between TUS-Net, TDUS-Net and radiologists

In the internal dataset, the TDUS-Net achieved a higher sensitivity (79.18%), specificity (89.88%), and accuracy rate (84.59%) than that of TUS-Net (sensitivity:77.84%, specificity:87.8%, and accuracy rate:83.21%). The area under the curve (AUC) of TDUS-Net and TUS-Net was 0.898 (95% CI: 0.870-0.926), and 0.881 (95% CI: 0.850-0.908) respectively. There was a statistically significant difference between TDUS-Net and TUS-Net (P =0.0491). The 10-fold cross validation results were shown in Table 3. Among the 10-fold results, Model 8 achieved a higher performance with a sensitivity of 91.67%, a specificity of 96.30%, and an accuracy rate of 94.12%. In addition, Model 8 achieved the most balanced sensitivity and specificity performance. Therefore, we selected and evaluated TDUS-Net model 8 on the external test set. The TDUS-Net in the external dataset achieved a sensitivity of 78.00%, a specificity of 95.00%, and an accuracy of 86.50% (Figure 6).

Regarding the diagnostic performance of radiologists, the radiologist B with 10 years' experience achieved a higher sensitivity (78.91%), specificity (92.58%), and accuracy (85.74%) in the internal dataset than the radiologist with 5 years' experience (68.17%, 79.49%, and 73.73% respectively). For the external test datasets, the TDUS-Net achieved the highest sensitivity (78.00%), specificity (95.00%), and accuracy (86.50%), followed by the radiologist with 10 years' experience (sensitivity: 73.00%, specificity: 89.00%, and accuracy: 81.00%), the average performance of the radiologists (sensitivity: 71.00%, specificity: 82.00%, and accuracy: 76.50%), and the junior radiologist A (sensitivity: 69.00%, specificity: 75.00%, and accuracy: 72.00%) (Table 4). The senior radiologist achieved a higher specificity (92.58%) and accuracy (85.74%) than the TDUS-Net for the internal dataset. There was statistically significant

difference between TDUS-Net and the average performance of the radiologists (P =0.0001), and between TUS-Net and the average performance of the radiologists (P = 0.0014). Compared with the less experienced radiologists, the TDUS-Net achieved higher performance in classifying thyroid cancers in both internal and external dataset.

Discussion

This retrospective pilot study showed that the TDUS-Net model reached the competitive or even better diagnostic performance of the experienced radiologists, with slightly higher sensitivity in the internal dataset (AUC of 0.898 [95%CI: 0.868-0.922]) and sig-nificantly higher sensitivity, specificity, and accuracy in the external dataset (AUC of 0.925 [95%CI: 0.880-0.958)). Furthermore, we investigated whether colour Doppler can provide the additional value to uplift the diagnostic performance. Our results indicated that compared to the use of grey-scale US alone (TUS-Net), TDUS-Net had a higher di-agnostic performance than TUS-Net in terms of sensitivity, specificity, and accuracy (TDUS-Net: 79.18%, 89.88%, 84.59%; TUS-Net: 77.84%, 87.8%, 83.21%). To the best of our knowledge, this is the first multi-centre research to investigate the added value of CDUS in classifying thyroid cancers.

CDUS is a technique used to evaluate tumour vascularity and has been largely used as a diagnostic tool for distinguishing benign and malignant TNs[18]. That may be due to the finding that the survival and growth of malignancies depend largely on the available blood supply[19]. Because vascularity in malignant thyroid nodules may have marked differences from that in benign nodules[20,21], it is vital to quantify vascularisation. Intra-observer and inter-observer variabilities were commonly found in vascularity analysis since only subjective or semiquantitative assessment methods have been adopted in clinical practice[22,23]. A semi-

291 objective study was developed and presented by Cosgrove et al. to identify the vascularity
292 quantity of breast nodules with a vascular score[24]. Their results showed that the presence of
293 dramatically abnormal blood flow was correlated with malignant breast tumours in terms of
294 both the number of vessels as well as the flow velocity and pattern within the malignancies.
295 Fein et al. established a new approach for quantifying vascularity in tumours[25]. They adopted a
296 computer-assisted image analysis system to calculate three colour Doppler parameters, known
297 as the ratio of colour pixels in the ROI (CPD), ratio of pixels with the limit colour indicating
298 the non-identification of pixels (LCD), and the average of all absolute colour values (MCV).
299 Their method was better than the approach of colour pixel counting reported by Cosgrove et
300 al.[24] and had a faster speed and more accurate diagnostic accuracy in classifying malignant and
301 benign nodules. However, the limitations of that method were also noticeable. For example,
302 colour values may not match the correct flow velocity due to aliasing, which affected the
303 accuracy of MCV using a low-scale maximum. In addition, the noise was another influential
304 limitation. In other words, some colours lack corresponding blood flow signals in the image,
305 which may also affect the accuracy of the vascularity evaluation. Fukunari et al.[26] developed a
306 4-grade scale to evaluate vascularity in benign TNs and thyroid follicular carcinomas. A total
307 of 7.9% of adenomatous nodules showed grade 4 (moderate to rich) blood flow signals, whereas
308 40.9% of follicular carcinomas showed moderate to rich blood flow signals. The results
309 contradicted ours because only follicular carcinomas were included in their study. In contrast,
310 our study was not limited to follicular thyroid carcinoma, but included two additional types of
311 malignant TNs, and therefore closer to the clinical reality.
312 This study proposed an entirely objective method to extract colour Doppler features in thyroid

313 nodules through using Computer Vision techniques. Four quantitative Doppler features, namely,
314 whole blood flow ratio, blood flow ratio in the peripheral area, blood flow ratio in the
315 intranodular area, and the number of vascularity, were investigated. Kim et al.[27] roughly defined
316 the blood flow locations shown on colour Doppler US images as scant, peripheral, central, and
317 mixed type. The peripheral type referred to those presented with more blood flow signals in the
318 peripheral area than those in the central area, and the central type was vice versa. Their study
319 claimed no statistically significant difference in the central and peripheral colour Doppler
320 features in benign and malignant thyroid nodules. Only one malignant and one benign nodule
321 in their study was categorised as the central type and peripheral type blood flow. However, their
322 study only included 27 solid, round, isoechoic thyroid nodules. The study sample was rather
323 small and did not represent variety of nodules. Our study showed that more blood flows were
324 found in benign nodules than in malignant nodules in terms of the whole ratio, blood flow ratio
325 in the peripheral area, and blood flow ratio in the intranodular area (all $P<0.001$). This was the
326 first study to calculate the number of vessels in a full objective way. In previous studies,
327 semiquantitative methods such as Adler's classification was widely adopted in visually
328 counting the number of vessels in breast nodules[28], solid renal tumors[29], soft tissue tumors[30],
329 cervical cancer[31], and thyroid nodules[32]. Xia et al.[32] adopted Alder's classification to divide the
330 number of vascularity in thyroid nodules into four categories, namely, grade 0 (no blood vessel),
331 grade I (low blood flow with rod-like blood flows), grade II (medium blood flow with 3-4 blood
332 vessels and at least one blood vessel is longer than the radius of the nodule), and grade III (rich
333 blood flow with > 4 blood flow signals). This grading scheme is rather crude and heuristically
334 defined. Their study showed the majority of malignant thyroid nodules presented low blood

flows whereas the majority of benign nodules were absent with blood flow or had low blood flow, which indicated that there was no significant difference between malignant and benign TNs in terms of vascularity quantity. However, the results of our study revealed that there was a negative association between the number of vessels and thyroid cancer (R=-0.92, P =0.037). The more vascularity a thyroid nodule had, the less likely it was a malignant thyroid nodule. The way of calculating the number of vessels can explain the conflictions of the results between ours (100% objective) and Xia's team (semi-subjective).

The trained TDUS-Net in our study based on 5 TI-RADS features with the 4 CDUS features showed a promising diagnostic performance. The diagnostic performance of TDUS-Net was comparable to the skilled radiologists' performance in the internal dataset and higher than that of radiologists in the external dataset tests regardless the radiologists' experience. Previous studies confirmed the assistance of colour Doppler ultrasound in machine learning for improving other cancer diagnosis, such as breast cancer. Afaf et al.[33] ex-tracted quantitative colour Doppler radiomics features (the vascular fractional area and blood flow velocity index) and further combined these two Doppler features with nine conventional ultrasound features in classifying breast cancer. Their model achieved a sensitivity of 96.9% and a specificity of 76.8%. Wu et al.[34] developed a machine learning model, which was based on nine quantitative grey-scale ultrasound features and three colour Doppler features. The three colour Doppler features were the fractional area of flow in the nodule, mean flow velocity, and flow volume in the nodule. Three vascularity features showed a statistically difference between triple-negative breast cancer and non-triple-negative breast cancer ($p<0.05$). The combined model showed the highest diagnostic performance of a sensitivity of 86.96% and a specificity of 82.91%. Similarly,

our results also confirmed the added value of colour Doppler features in improving thyroid cancer diagnosis, but our Doppler features represent the amount of vascular activities in specific locations.

Our study has certain limitations. First of all, although it is a multi-centre study, the scope is limited to reflect the pilot study nature, and hence the amount of data samples is still relatively small, especially with the external datasets. Secondly, the performance of our model is expected to increase by expanding the datasets to real-world data. The included data sample was balanced of 50% malignant nodules and 50% of benign nodules, whereas the real-world data would be more proportion of benign nodules and less proportion of malignant nodules. Thirdly, we only included thyroid nodules with pathological results confirmed by FNAB. However, in clinical practice, most benign nodules did not receive FNAB, which means a selection bias existed in our study. Also, manual ROI selection would also result in significant variability. Therefore, a larger real-world data sample with more thyroid cancer subtypes and innovative methods (such as automatic ROI selection) would be our future research.

Conclusions

In conclusion, the automatic extracted colour Doppler features can provide added value in differentiating benign and malignant thyroid nodules. Applying a machine learning model combining both conventional ultrasound features and colour Doppler features can reach comparable or even higher than radiologists in terms of sensitivity, specificity, and accuracy in classifying thyroid cancer. The improved technical performance has significant potential for enhancing the ability of radiologists in thyroid cancer diagnosis, especially for junior radiologists.

Fund

This research was funded by Outstanding Young Medical Talent Training Program of Pudong Health Bureau of Shanghai, grant number PWRq2020-40.

Reference

1. Ribatti D.; Judah F. A pioneer in the study of angiogenesis. Angiogenesis 2008, 11, 3-10.
2. Ramsden J. Angiogenesis in the thyroid gland. J Endocrinol 2000, 166, 475-480.
3. Bray F; Ferlay J; Soerjomataram I; Siegel RL; Torre LA; Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of in-cidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018, 68, 394-424.
4. Chen M; Fu X; Shen Y. Evaluation of Multimode Color Doppler Flow Imaging in the Diagnosis of Solid Renal Tumor. Con-trast Media Mol Imaging 2021 April 1, 2021:6656877.
5. Adler DD; Carcon PL; Rubin JM; Quinne-Reid D. Doppler ultrasound color flow imaging in the study of breast cancer: Pre-liminary findings. Ultrasound Med Biol 1990, 16, 553-559.
6. Zhao WJ; Fu LR; Huang ZM; Zhu JQ; Ma BY; Tarantino G. Effectiveness evaluation of computer-aided diagnosis system for the diagnosis of thyroid nodules on ultrasound: A systematic review and meta-analysis. Medicine 2019, 98, e16379.
7. Yoo YJ; Ha EJ; Cho YJ; Kim HL; Han M; Kang SY. Computer-aided diagnosis of thyroid nodules via ultrasonography: initial clinical experience. Korean J Radiol 2018, 19, 665-672.
8. Choi YJ; Baek JH; Park HS; Shim WH; Kim TY; Shong YK; Lee JH. A computer-aided diagnosis system using artificial intel-ligence for the diagnosis and characterization of thyroid

401 nodules on ultrasound: Initial clinical assessment. Thyroid 2017, 27, 546-552.
402 9. Tessler FN; Middleton WD; Grant EG; Hoang JK; Berland LL; Teefey SA; Cronan JJ;
403 Beland MD; Desser TS; Frates MC; Hammers LW; Hamper UM; Langer JE; Reading CC;
404 Scoutt LM; Stavros AT. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White
405 Paper of the ACR TI-RADS Committee. J Am Coll Radiol 2017, 14, 587-595.
406 10. Papini E; Guglielmi R; Bianchini A; Crescenzi A; Taccogna S; Nardi F; et al. Risk of
407 malignancy in nonpalpable thyroid nod-ules: predictive value of ultrasound and color-Doppler
408 features. J Clin Endocrinol Metab 2002, 87,1941-1946.
409 11. Cappelli C; Castellano M; Pirola I; Cumetti D; Agosti B; Gandossi E; et al. The predictive
410 value of ultrasound findings in the managment of thyorid nodules. QJM 2007, 100, 29-35.
411 12. Cappelli; Pirola I; Cumetti D; Micheletti L; Tironi A; Gandossi E; et al. Is the
412 anteroposterior and transverse diameter ratio of nonpalpable thyroid nodules a sonographic
413 criteria for recommending fine-needle aspiration cytology? Clin Endorinol 2005, 63, 689-693.
414 13. Shimamoto K; Sakuma S; Ishigaki T; Makino N. Intratumoral blood flow: evaluation with
415 color Doppler echography. Radi-ology 1987, 165, 683-685.
416 14. Argalia G; D'Ambrosio F; Lucarelli F; Mignosi U,;Giuseppetti GM; Passarini G; et al.
417 Echo Doppler in the characterisation of thyorid nodular disease. Radiol Med 1995, 89, 651-657.
418 15. MH Wu; CN C; KY Chen; MC Ho; HC Tai; YC Chung; CP Lo; A Chen; KJ Chang.
419 Quantitative analysis of dynamic power Doppler sonograms for patients with thyroid nodules.
420 Ultrasound Med Biol 2013, 39, 1543-1551.
421 16. Weidner N. Current pathologic methods for measuring intratumoral microvessel density
422 within breast carcinoma and other solid tumors. Breast Cancer Res Treat 1995, 36, 169-180.

17. Elston CW; Ellis IO. Pathological prognostic factors in breast cancer. The value of histological grade in breast cancer: expe-rience from a large study with long-term follow-up. Histopathology 2002, 41, 154-161.

18. Ralls PW; Mayekawa DS; Lee KP; Colletti PM; Radin DR; Boswell WD; Halls JM. Color-flow Doppler sonography in Graves disease: "thyroid inferno". AJR Am J Roentgenol 1988, 150, 781-784.

19. Ye C; Feng C; Wang S; Liu X; Lin Y; Li M. Antiangiogenic and antitumor effects of endostatin on follicular thyroid carcinoma. Endocrinology 2002, 143, 522–3528.

20. Foschini M; Papotti M; Parmeggiani A; Tallini G; Castaldini L; Meringolo D; Eusebi V. Three-dimensional reconstruction of vessel distribution in benign and malignant lesions of thyroid. Virchows Arch 2004, 445, 189-198.

21. Gritzmann N; Koischwitz D; Rettenbacher T. Sonography of the thyroid and parathyroid glands. Radiol Clin North Am 2000, 38, 1131-1145.

22. Samanci C; Ozkose B; Ustabasioglu FE; Erol BC; Sirolu S; Yılmaz F; Ozkose ZG; Yılmaz H; Kara SC; Kicik Caliskan R; Gulsen F. The Diagnostic Value of Superb Microvascular Imaging in Prediction of Uterine Artery Embolization Treatment Response in Uterine Leiomyomas. J Ultrasound Med 2021, Feb 18. doi: 10.1002/jum.15647. Epub ahead of print. PMID: 33599335.

23. Chung J; Lee YJ; Choi YJ; Ha EJ; Suh CH; Choi M; Baek JH; Na DG. Clinical applications of Doppler ultrasonography for thy-roid disease: consensus statement by the Korean Society of Thyroid Radiology. Ultrasonography 2020, 39, 315-330.

24. Cosgrove D; Bamber J; Davey J; McKinna J; Sinnett H. Color Doppler signals from breast

445 tumors. Radiology 1990, 176, 175-180.

446 25. Fein M; Delorme S; Weisser G; Zuna I; van Kaick G. Quantification of color Doppler for

447 the evaluation of tissue vascularisa-tion. Ultrasound Med Biol 1995, 21, 1013-1019.

448 26. Fukunari N; Nagahama M; Sugino K; Mimura T; Ito K; Clinical evaluation of color

449 Doppler imaging for the differential di-agnosis of thyroid follicular lesions. World J Surg 2004,

450 28, 1261-1265.

451 27. Kim DW; June SJ; Eom JK; Kang T. Color Doppler Features of Solid, Round, Isoechoic

452 Thyroid Nodules without malignant sonographic features: a prospective cytopathological study.

453 Thyroid 2013, 23, 472-476.

454 28. Li S; Chen B; Fang X; Wan M; Cao D; Xu H; Liu L; Wei Q; Yu J. A better understanding

455 of testicular and/or epididymal tuber-culosis based on clinical, ultrasonic, computed

456 tomography, and magnetic resonance imaging features at a high-volume in-stitute in the modern

457 era. Quant Imaging Med Surg 2021, 11, 2465-2476.

458 29. Chen M; Fu X; Shen Y. Evaluation of Multimode Color Doppler Flow Imaging in the

459 Diagnosis of Solid Renal Tumor. Con-trast Media Mol Imaging 2021, Apr 1;2021:6656877.

460 30. Wu M; Hu Y; Ren A; Peng X; Ma Q; Mao C; Hang J; Li A. Nomogram Based on

461 Ultrasonography and Clinical Features for Predicting Malignancy in Soft Tissue Tumors.

462 Cancer Manag Res 2021, 13, 2143-2152.

463 31. Che D; Yang Z; Wei H; Wang X; Gao J. The Adler grade by Doppler ultrasound is

464 associated with clinical pathology of cer-vical cancer: Implication for clinical management.

465 PLoS One 2020, 15, e0236725.

466 32. Xia S; Yao J; Zhou W; et al. A computer-aided diagnosing system in the evaluation of

thyroid nodules – experience in a spe-cialized thyroid center. World J Surg Oncol 2019, 17, 210.

33. Moustafa AF; Cary TW; Sultan LR; Schultz SM; Conant EF; Venkatesh SS; Sehgal CM. Color Doppler Ultrasound Improves Machine Learning Diagnosis of Breast Cancer. Diagnostics (Basel) 2020, 10, 631.

34. Wu T; Sultan LR; Tian J; Cary TW; Sehgal CM. Machine learning for diagnostic ultrasound of triple-negative breast cancer. Breast Cancer Res Treat 2019, 173, 365-373.

Figures

Figure 1. Flowchart of patient selection in the study.

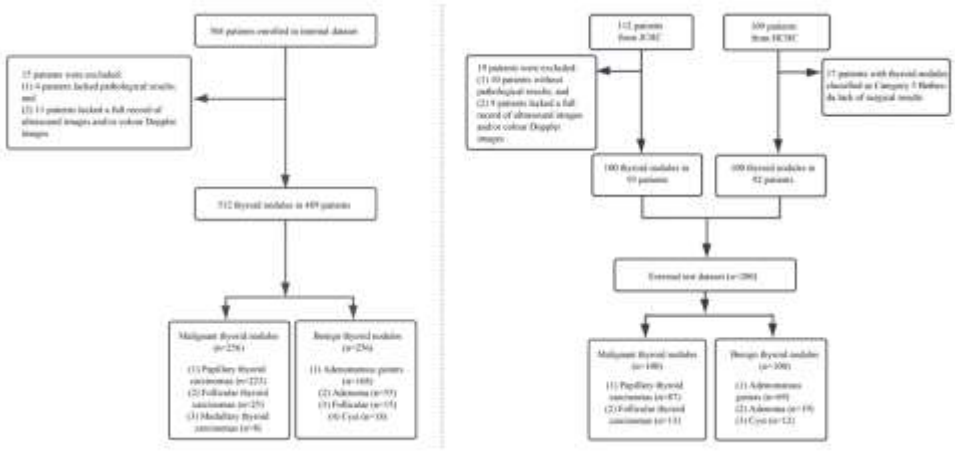


Figure 2. Illustration of the segmentation of peripheral and intranodular regions of a thyroid nodule. (A) refer to the original ROI of a thyroid nodule; (B) refer to the secondary ROI obtained when defining the best n% offset; (C) refer to the intranodular regions of the thyroid nodule; and (D) refer to the peripheral area of the thyroid nodule.

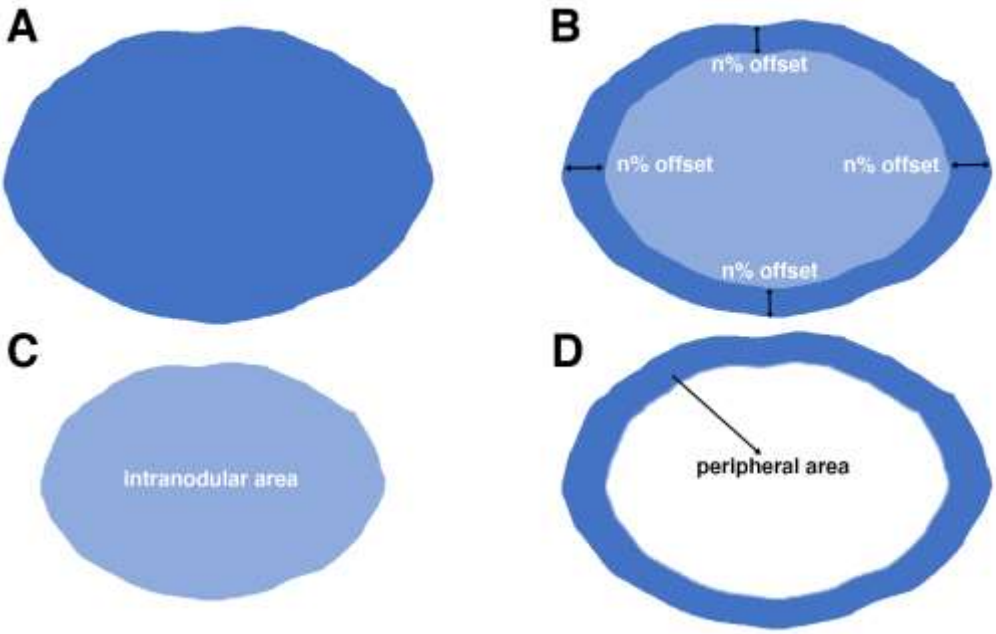


Figure 3. Illustration of 4 colour Doppler features. (upper) conventional ultrasound images; (left) number of vascularity; (the second left) whole ratio of blood flow; (the third left) blood flow ratio in peripheral areas; (right) blood flow ratio in intranodular areas.

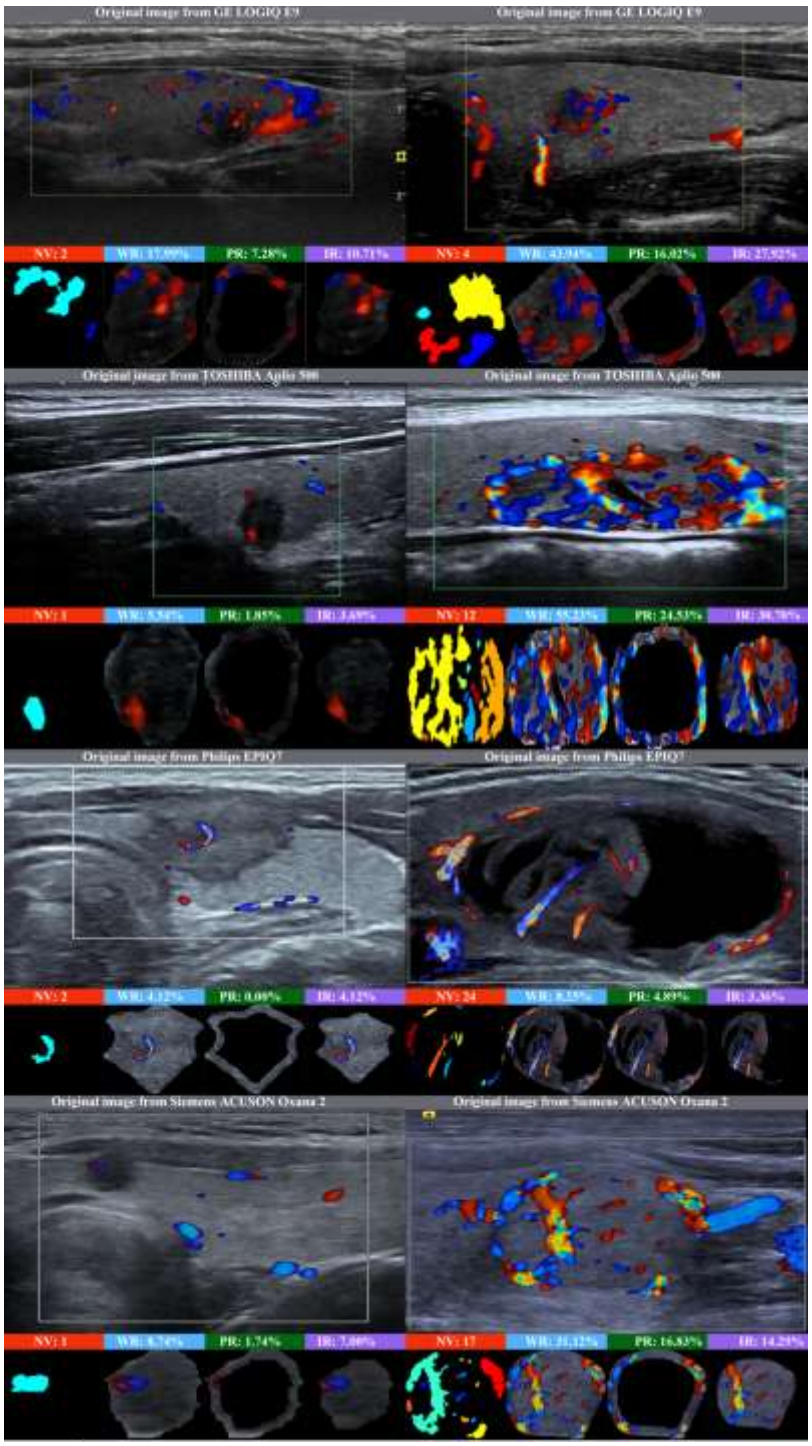


Figure 4. Study workflow of building the TUS-Net and TDUS-Net.

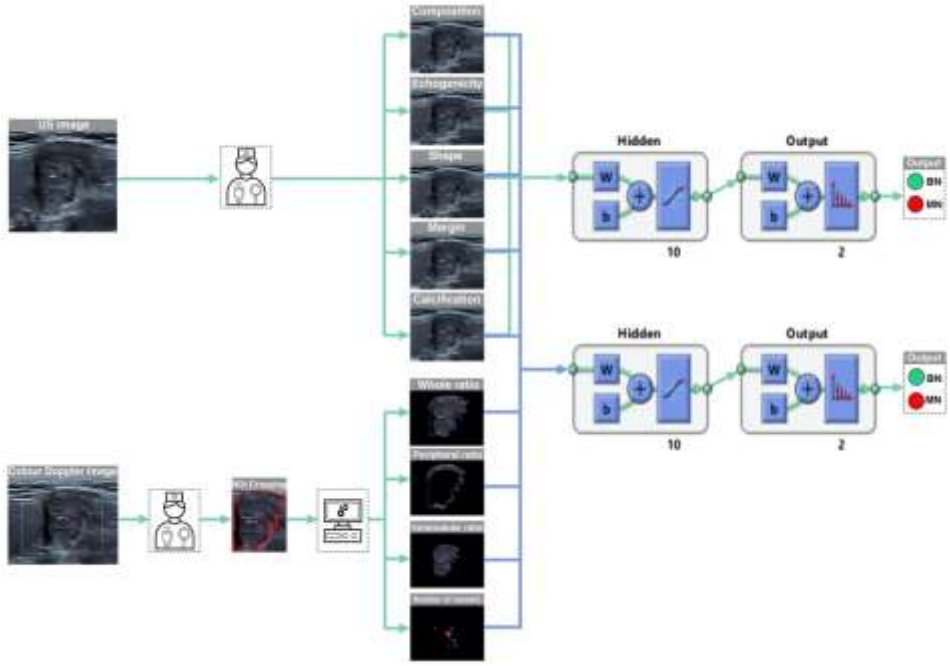


Figure 5. Number of vessels in malignant and benign thyroid nodules.

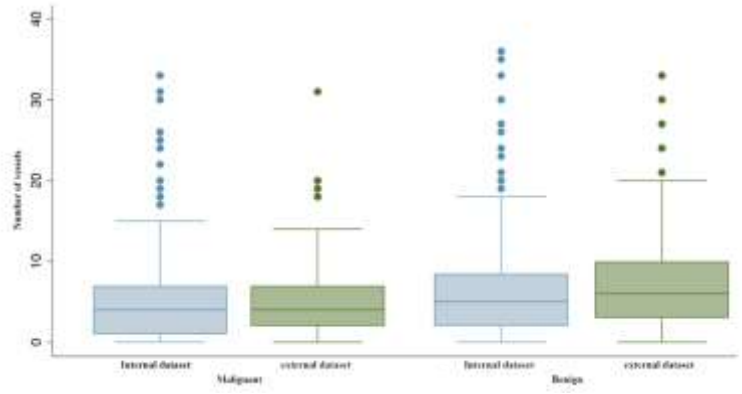
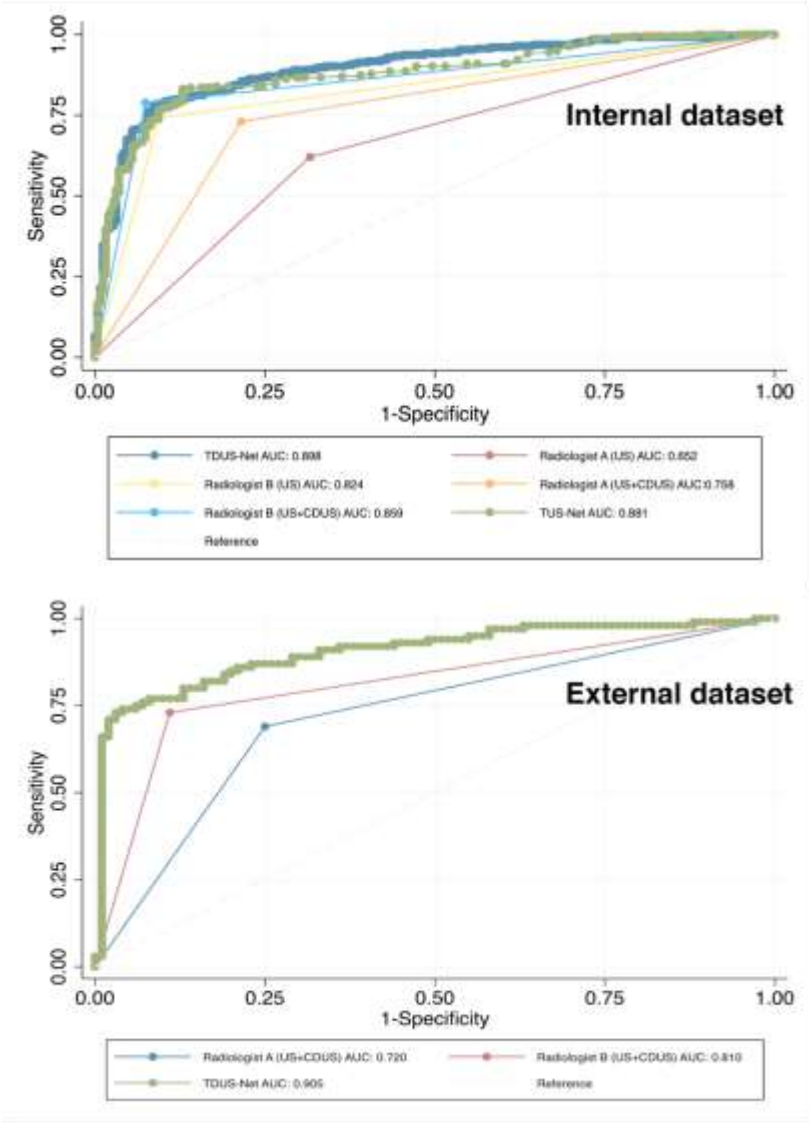


Figure 6. Diagnostic performance of radiologists, TUS-Net, and TDUS-Net.


Internal dataset
Sensitivity
1-Specificity
0.00
0.25
0.50
0.75
1.00
TDUS-Net AUC: 0.898
Radiologist A (US) AUC: 0.852
Radiologist B (US) AUC: 0.824
Radiologist A (US+CDUS) AUC:0.758
Radiologist B (US+CDUS) AUC: 0.859
TUS-Net AUC: 0.881
Reference
External dataset
Sensitivity
1-Specificity
0.00
0.25
0.50
0.75
1.00
Radiologist A (US+CDUS) AUC: 0.720
Radiologist B (US+CDUS) AUC: 0.810
TDUS-Net AUC: 0.905
Reference

497