ORIGINAL RESEARCH

Integrating Deep Learning and Radiomics in Differentiating Papillary Thyroid Microcarcinoma from Papillary Thyroid Carcinoma with Ultrasound Images

Bing Yu[1,2,*], Huijuan He[3,*], Qiao Zheng[2,*], Yao Ai[2], Xianwen Yu[2], Sunjun Li[4], Ji Zhang[2], Juebin Jin [5], Xiance Jin [2,6], Wenliang Yu[3]

[1]Purchasing Center Department, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, 310020, People's Republic of China; [2]Radiation and Medical Oncology Department, Wenzhou Medical University First Affiliated Hospital, Wenzhou, 325000, People's Republic of China; [3]Radiation Oncology Department, Quzhou People's Hospital, Quzhou, 324002, People's Republic of China; [4]Alberta College, Wenzhou Medical University, Wenzhou, 325000, People's Republic of China; [5]Medical Engineering Department, Wenzhou Medical University First Affiliated Hospital, Wenzhou, 325000, People's Republic of China; [6]School of Basic Medical Science, Wenzhou Medical University, Wenzhou, 325000, People's Republic of China

*These authors contributed equally to this work

Correspondence: Xiance Jin, Radiation and Medical Oncology Department, Wenzhou Medical University First Affiliated Hospital, Wenzhou, 325000, People's Republic of China, Email jinxc1979@hotmail.com; Wenliang Yu, Radiation Oncology Department, Quzhou People's Hospital, Quzhou, 324002, People's Republic of China, Email wlyu1982@aliyun.com

Purpose: The feasibility and accuracy of ultrasound-based radiomics, deep learning, and combined deep learning radiomics models were investigated in the differentiation of papillary thyroid carcinoma and papillary thyroid microcarcinoma to decrease the risk of overtreatment of papillary thyroid microcarcinoma.

Methods: A total of 549 patients with confirmed 180 papillary thyroid carcinoma and 436 papillary thyroid microcarcinoma nodules from Hospital One were enrolled and randomly divided into training and validation cohorts at a ratio of 8:2 with 56 patients left as independent testing set 1. Fifty patients from Hospital Two were enrolled as independent testing set 2. Radiomics signature and five deep learning networks, such as visual geometry group 13 (VGG13), VGG16, VGG19, AlexNet, and EfficientNet, were generated for papillary thyroid carcinoma and papillary thyroid microcarcinoma differentiation. Combined deep learning and radiomics models were constructed to further improve the differentiation ability.

Results: An area under curves of 0.826 and 0.822 was achieved with radiomics model for papillary thyroid carcinoma and papillary thyroid microcarcinoma differentiation in the independent testing set 1 and set 2, respectively. VGG19 achieved the best area under curves of 0.890 and EfficientNet achieved the best accuracy of 0.867. The best accuracy and area under curves of 0.904, 0.900, and 0.931, 0.946 were achieved with the combination of VGG + radiomics (R_V_Combined) and EffiecientNet + radiomics (R_E_Combined) in the independent testing set 1 and set 2, respectively.

Conclusion: Deep learning and radiomics combination models are promising in the noninvasively preoperative differentiation of papillary thyroid microcarcinoma and papillary thyroid carcinoma to decrease the overtreatment of patients with papillary thyroid microcarcinoma and to minimize the complications caused by overtreatment.

Keywords: papillary thyroid carcinoma, papillary thyroid microcarcinoma, ultrasound, deep learning, radiomics

Introduction

For the past few decades, the incidence of thyroid cancer is increasing rapidly worldwide with papillary thyroid carcinoma (PTC) being the most common type of thyroid cancer.[1] Except for the actual increase in tumor occurrence, the increased prevalence of PTC is due to the increasing use of high-resolution ultrasound (US) imaging and US-guided fine-needle aspiration biopsies (FNAB).[2] The increased accuracy of pathological thyroid examinations also contributes to the increased diagnosis of PTC, especially the overdiagnosis of papillary thyroid microcarcinoma (PTMC),[3] which is

defined as a PTC with a diameter of ≤10 mm.[4] Studies reported that most of PTMCs (over 70%) are diagnosed incidentally in autopsies and thyroidectomy specimens, and most PTMCs are benign with reported mortality from 0% to 1%.[5] Therefore, the risks of overdiagnosis and overtreatment are highly possible for PTMCs, which causes potential complications, distress, and economic burdens for patients with PTMCs.[6] Preoperative differentiation of PTMC from PTC is of great clinical significance to avoid overtreatment and to determine the appropriate treatment options for patients with PTMCs.[7]

US imaging of the thyroid and neck is usually the initial workup for a patient with a thyroid nodule, but the visual interpretation of US images in the diagnosis of PTC and PTMC is limited not only because the diagnosis highly depends on the radiologists' experience but also the interobserver variation.[8] Although the study of Ma et al demonstrated that combining conventional US, contrast-enhanced ultrasound (CEUS) and real-time elastography (RTE) is able to improve the diagnostic accuracy of PTMC, it is not a routine clinical practice for a patient with all three types of US exams.[9] With the emergence of radiomics, studies reported that US-based radiomics is able to differentiate benign and malignant thyroid nodules.[10,11] Deep learning models with convolutional neural network (CNN) have also been investigated for the differentiation of benign and malignant thyroid nodules with US images and demonstrated excellent performance compared with radiologists.[12–14] However, few studies have addressed the application of radiomics and deep learning models in the differentiation of PTC and PTMC.

In this study, the feasibility and accuracy of US-based radiomics, deep learning, and combined deep learning radiomics models were investigated in the differentiation of PTMC and PTC to decrease the risk of overtreatment with patients' data from two hospitals, one hospital for training and testing, and second hospital for external validation.

Materials and Methods
Patients
Patients diagnosed with papillary thyroid neoplasm in Hospital One from January 2018 to September 2020 were retrospectively reviewed according to the electronic medical records. Enrolled patients were randomly divided into training, validation, and independent testing sets. Fifty PTC cases acquired from Hospital Two were used as an independent external testing dataset. The inclusion criteria were as follows: 1) pathologically confirmed PTMCs and PTCs; 2) diagnosed by US images with detailed sonographic features described. The exclusion criteria included the followings: 1) with preoperative therapy (resection biopsy, neoadjuvant radiotherapy, or chemotherapy); 2) benign thyroid lesions; 3) missing important histopathological results (immunohistochemical results or lymph-nodule results); 4) incomplete information or images. Routine clinical tests and patients' characteristics were also extracted from the records. Figure 1 shows the flowchart for patients' enrollment. This study was approved by the ethics committee in Clinical Research (ECCR) of the First Affiliated Hospital of Wenzhou Medical University and conducted following the Declaration of Helsinki (ECCR no. 2019059) with confirmed patient confidentiality. The requirement of informed consent was waived by the ECCR according to the retrospective nature of this study.

US Examinations and Clinical Factors
US examinations of thyroid nodules were performed with high-frequency linear probes (5 MHz to 14 MHz) with a variety of US systems: Philips EPIQ7C (Philips Medical Systems, the Netherlands), GE Volume E8 (GE Medical Systems, USA), Siemens ACUSON OXANA 2 (Siemens Medical Solutions, USA), Esaote MyLab Class C (Esaote, Italy), Hitachi HI VISION Preirus (Hitachi-Aloka Medical, Japan) and Mindray Resona 7T (Mindray Medical International, China). The US images included both transverse and longitudinal sections of nodules. Clinical characteristics included basic information (age and sex) and ultrasound findings, which consist of composition, echogenicity, shape, margin, and echogenic foci, as well as stages scored according to the thyroid imaging reporting and data system (TI-RADS) criteria of the American College of Radiology.[15]
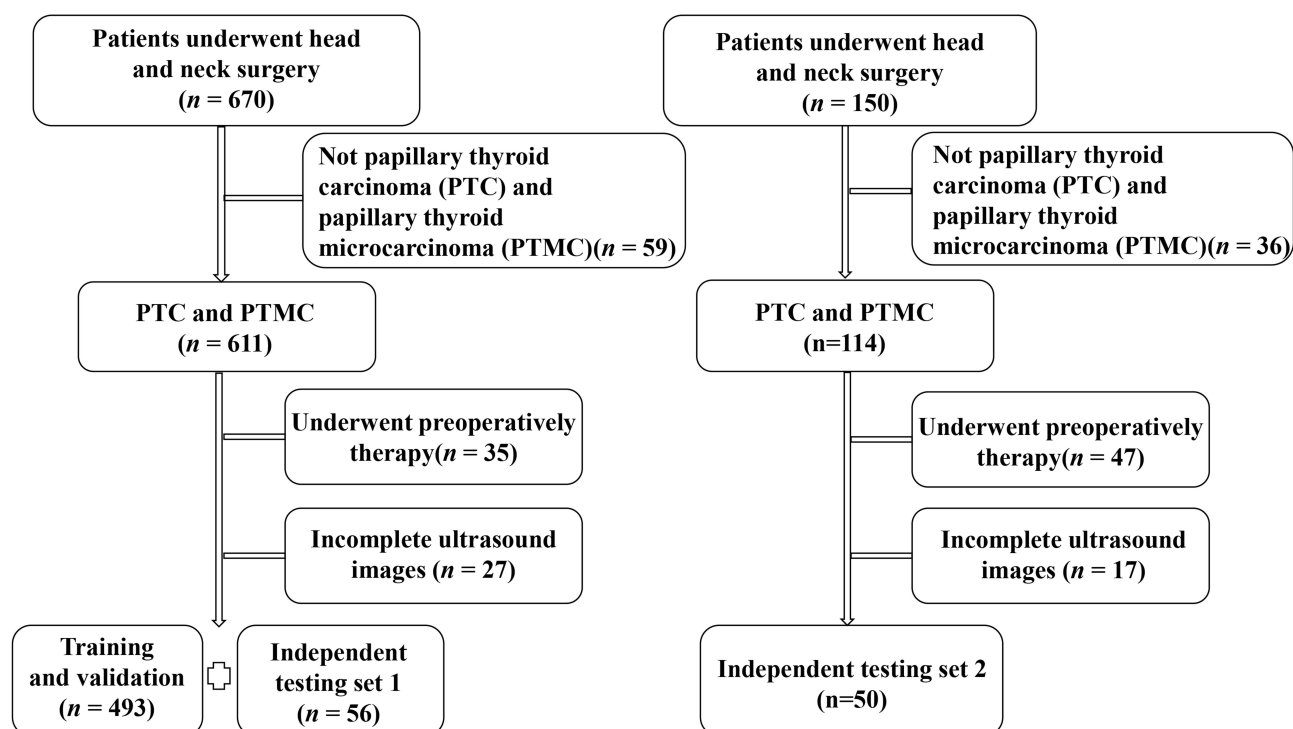
Figure 1 The patient enrollment process for the training set and the two independent testing sets. The training cohort and the independent testing set 1 patient enrollment process are shown on the left, and the independent testing set 2 is shown on the right. These two datasets are from two different hospital patients.

Radiomics Features Extraction and Modeling

Target volumes were contoured manually by one junior radiologist on the US images and confirmed by a senior radiologist with over 15 years of experience. Supplementary Figure 1 shows the typical US with contoured target volumes. Python (v. 3.7.0; https://www.python.org/) and package pyradiomics 2.2.0 (version 2.2) were used to extract radiomics features from the manually segmented target volumes. Based on different matrices that capture the spatial intensity distribution and wavelet filtering, a total of 1566 radiomics features were extracted, which includes 88 exponential features, 88 gradient features, 88 logarithm features, 70 square features, 352 wavelet features, and 880 log features, respectively. Radiomics features with a p <0.05 in Mann–Whitney *U*-tests were selected as potentially informative features, then the least absolute shrinkage selection operator (LASSO) was applied to identify optimal features for PTMC and PTC classification.[16] A ten-fold cross validation was applied to tune the elastic net parameters to reduce the redundant information and to avoid over-fitting. A minimum standard deviation and maximum area under curves (AUC) were achieved by tuning coefficient λ. The linear combination of selected radiomics features with respective weights makes the final radiomics signature.

Deep Learning Models

In the preprocessing, a rectangular region of interest (ROI) was cropped from raw US images according to the tumor segmentation mask and resized to 224×224 pixels for normalization. Five deep learning networks, visual geometry group 13 (VGG13),[17] VGG16, VGG19, AlexNet, and EfficientNet were pre-trained on Imagenet and then adopted as the deep learning networks in this study using python 3.7 programming language Tensorflow 2.4, and Keras 2.2.4 open-source programming packages.[18] As shown in Supplementary Figure 2 the structures of different deep learning networks, VGG uses a pooling layer as a demarcation and has six block structures with each having the same number of channels. Because both the convolutional and fully connected layers have weight coefficients, they are also referred to as weight layers. The pooling layer does not involve weights. For the VGG CNN, its convolutional layers and pooling layers are responsible for feature extraction, and the final 3 fully connected layers are responsible for the classification task (×2

means that the module in brackets is repeated twice). The AlexNet network consists of 8 layers with the first 5 layers convolutional and the last 3 layers of fully connected. Each backbone of the EfficientNet contains 7 blocks with each having different numbers of sub-blocks.

In the training stage, rectangular ROIs were fed into the networks to update model parameters with classification results as the output. The loss function was calculated based on the cross-entropy of the outputs and labels. A learning rate of 1e-4, dropout parameter of 0.6, training Epoch of 100, the activation function ReLu, the classification function softmax, and the Adam optimizer were applied to update the model parameters with a batch size of 32 and a maximum iteration step of 300. Compiled prediction codes were generated based on the data weights obtained from training to obtain the predicted value of each test picture. Identical parameters were applied for the training of these deep learning networks. The computer environment configuration for deep learning modeling was Linux 64-bit operating system, GPU RTX2080, and video memory 6GB with all other applications shut down while the program is running.

Fused Models and Evaluation

Firstly, radiomics and deep learning models were developed to differentiate PTC and PTMC independently. In order to improve the classification performance, the prediction scores of radiomics and deep learning models were fused by applying an information fusion method.[19] As shown in Figure 2, the prediction results were obtained independently by training each model in the early stage, then logistic regression was applied to fuse the outputs of each model to make a decision. The minimum loss value, positive predictive value (PPV), negative predictive value (NPV), recall rate, F1 score, and the AUCs of receiver operating characteristics (ROC) curves were calculated to evaluate and compare the performance of these models.[20]

Statistical Analysis

Detailed clinical differences between PTC and PTMC were compared by *t*-test, chi-square test, and Mann–Whitney *U*-test. LASSO regression model building was done using the "glmnet" package. Glmnet function in R language was applied for n cross validation (n = 10), which means that data was separated into 10 subsets. All statistics were two-sided and p-values less than 0.05 were considered to be statistically significant. Statistical analysis was performed using the R analysis platform (version 3.6.0), OriginPro2018, MedCalc (version 19.3.0), SPSS 19 software, and Python 3.7.
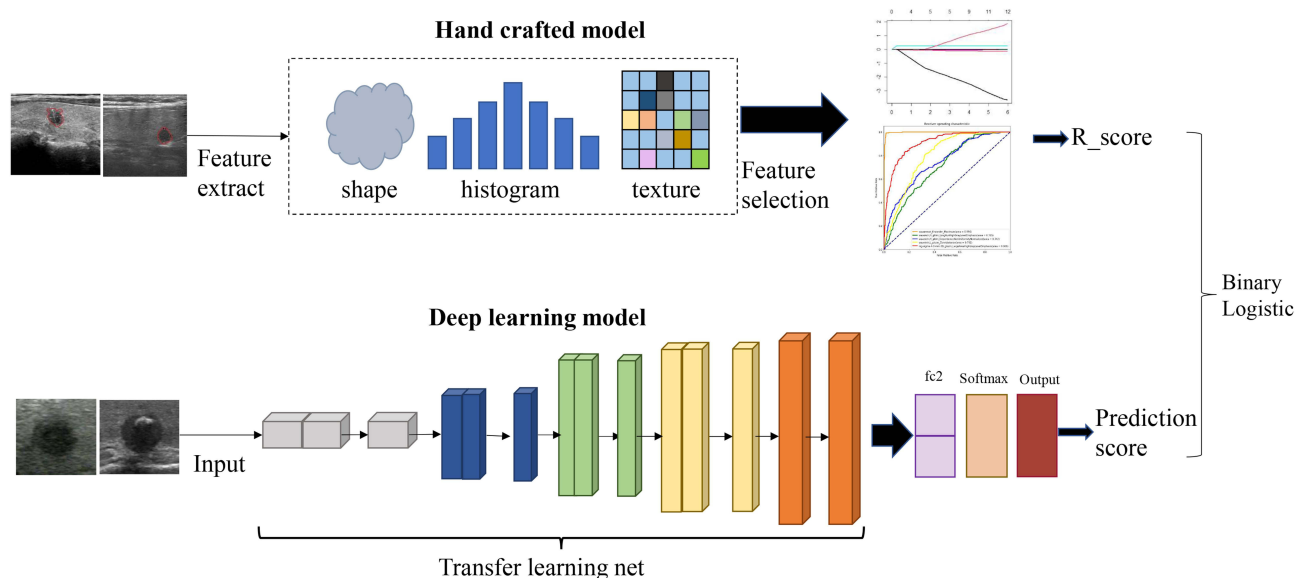

Figure 2 Schematic pipeline of the radiomics and deep learning modeling, as well as their combination for the differentiation of papillary thyroid tumor and papillary thyroid microcarcinoma.
Abbreviations: Fc2, 2 fully connected layer; R_score, radiomics score.

Results

A total of 549 patients with an average age of 46.55±11.33 years (ranges from 14 to 81 years) were enrolled in this study with confirmed 180 PTC and 436 PTMC nodules from Hospital One, as shown in Figure 1 the flowchart of patient selection. Patients were randomly divided into training cohort and validation cohort at a ratio of 8:2. There were 56 patients from Hospital One used as independent testing set 1. Fifty patients from Hospital Two with confirmed 25 PTC and 25 PTMC nodules were enrolled as independent testing set 2 for the external validation of these models. There were 205 PTC and 461 PTMC nodules with an average diameter of 15.80±7.15 mm and 6.89±3.25 mm, respectively. Detailed clinical characteristics of these patients were presented in Table 1. A total of 612 PTMC and 311 PTC US images were analyzed.

A total of 203 radiomics features were selected out of the extracted 1777 features according to the Mann–Whitney U-test with a p <0.05. A final 10 features were further screened out from the 203 features to build the radiomics signature using the LASSO logistic regression model, as shown in Supplementary Figure 3. These features included 5 first order features, 2 grey level run length matrix (GLRLM) features, 1 Gray Level Dependence Matrix (GLDM), and 2 Gray Level Size Zone Matrix (GLSZM). Detailed features and their corresponding non-zero coefficients were presented in Table S1. The ROC evaluation of radiomics signature in the differentiation of PTC and PTMC was shown in Figure 3 with an AUC of 0.908 (95% CI: 0.887–0.928), 0.826 (95% CI: 0.734–0.918), and 0.822 (95% CI: 0.710–0.936) in the validation cohort, independent testing set 1 and independent testing set 2, respectively.

Table 2 shows the performance of different deep learning models in the differentiation of PTC and PTMC with an accuracy of 0.800 (95% CI: 0.700–0.837), 0.850 (95% CI: 0.775–0.875), 0.850 (95% CI: 0.737–0.873), 0.850 (95% CI: 0.763–0.875) and 0.863 (95% CI: 0.787–0.932) in the validation cohort for AlexNet, VGG13, VGG16, VGG19, and EfficientNet, respectively. The ROC curves and the corresponding confusion matrices of different deep learning models were shown in Figure 4. As shown in Figure 4b the AUCs of AlexNet, VGG13, VGG16, VGG19, and EfficientNet were 0.800 (95% CI: 0.708–0.892), 0.850 (95% CI: 0.772–0.928), 0.846 (95% CI: 0.765–0.927), 0.890 (95% CI: 0.818–0.962), and 0.867 (95% CI: 0.789–0.945), respectively. Accordingly, VGG19 and EfficientNet were selected to combine with radiomics signature for further analysis. To further interpret the results of deep learning models, a heatmap of output features was generated by activating the visualization class. The generated heatmap was then overlaid onto the original image to generate the final deep learning visualization heatmap, as shown in Figure 4a the US images of 3 cases of PTMC and PTC, and their activated heatmap with VGG19. The heatmap could produce a coarse localization map highlighting the import regions for the classification targets.

Detailed comparison of the performance of radiomics signature, deep learning models, and combined deep learning radiomics models with the independent testing set 1 and set 2 were presented in Table 3. The accuracy of VGG19,

Table 1 The Clinical Characteristics of Enrolled Patients for Training and Testing

Characteristics	Training and Validation Sets			Independent Testing Set 1			Independent Testing Set 2		
	PTMC	PTC	P	PTMC	PTC	P	PTMC	PTC	P
Patient number	346	147		40	16		25	25	
Node number	396	164		40	16		25	25	
Sex			0.35			0.92			0.77
Male	94	34		8	3		10	9	
Female	252	113		32	13		15	16	
Age (years)	46.73±11.23	46.45±13.13	<0.0001	44.75±12.25	42.13±13.57	0.54	49.04±5.50	47.76±16.64	0.14
Diameter(mm)	6.89±3.26	16.01±7.30	<0.0001	13.61±5.15	7.14±3.17	<0.0001	8.01±2.09	15.21±5.94	<0.0001
TI-RADS			0.76			0.81			0.75
3	0	0		1	0		0	0	
4	27	10		5	2		19	18	
5	368	154		34	14		3	6	
6/7	1	0		0	0		3	1	

Notes: Categorical variables were compared by using the chi-square test; Continues variables were compared by using the Mann–Whitney U-test.
Abbreviations: PTMC, papillary thyroid microcarcinoma; PTC, papillary thyroid carcinoma; TI-RADS, thyroid imaging reporting and data system.
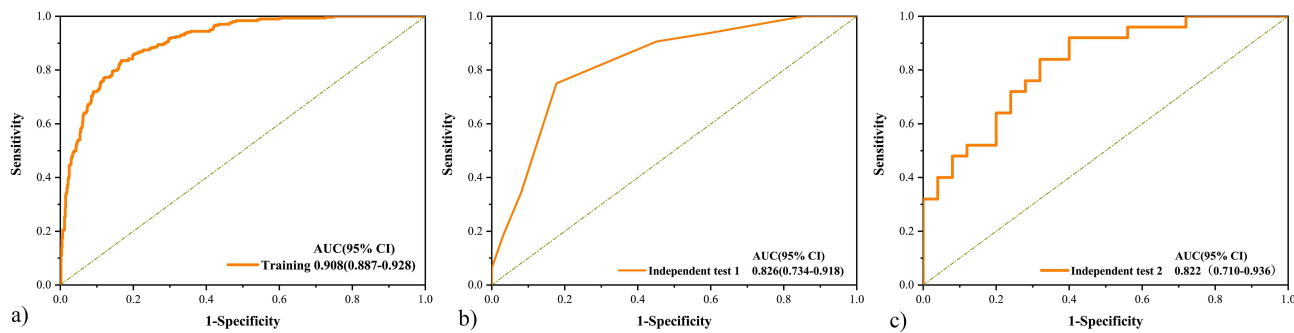
Figure 3 The performance of radiomics signature with ROC curves in (**a**) training, (**b**) independent testing set 1, and (**c**) independent testing set 2.

EfficientNet, Radiomics, combined radiomics VGG19 (R_V_combined), and radiomics EffiecientNet combination (R_E_combined) were 0.829, 0.798, 0.766, 0.904, 0.851 and 0.680, 0.680, 0.740, 0.780, 0.900 with the independent testing set 1 and set 2, respectively. Figure 5 shows the ROC curves of these models with an AUC of 0.826 (95% CI: 0.734–0.918), 0.890 (95% CI: 0.818–0.962), 0.867 (95% CI: 0.790–0.945), 0.931 (95% CI: 0.870–0.993), 0.908 (95% CI: 0.849–0.966) and 0.822 (95% CI: 0.709–0.936), 0.698 (95% CI: 0.546–0.849), 0.899 (95% CI: 0.806–0.993), 0.874 (95% CI: 0.778–0.969), 0.946 (95% CI: 0.885–1.000) for Radiomics, VGG19, EfficientNet, R_V_combined, and R_E_combined with the independent testing set 1 and set 2, respectively.

Discussion

In this study, the feasibility and accuracy of radiomics, deep learning models, and combined deep learning radiomics models were investigated in the differentiation of PTMC from PTC using US images. The models were further verified with external validation cohorts from a second hospital. An AUC of 0.826, 0.890, 0.867, 0.931, 0.908 and 0.822, 0.698, 0.899, 0.874, 0.946 for Radiomics, VGG19, EfficientNet, R_V_combined, and R_E_combined models was achieved with the independent testing set 1 and set 2, respectively.

With the increasing prevalence of high-resolution US and other imaging modalities, the diagnosis of thyroid cancers has increased remarkably and continuously.[21] Most of the newly diagnosed thyroid cancers are small PTCs including PTMCs.[22] Studies reported that the average size of thyroid tumors decreased from 1.51 cm in the year 2000 to 1.02 cm in 2005 with 36.9% of thyroid cancers being small than 1 cm in 2000, but PTMCs account for 61.48% of all thyroid cancers in 2005.[23] Similarly, PTMCs account for about 70% of all cases enrolled in this study. However, more evidence demonstrated that most PTMCs have a very indolent nature and excellent outcomes.[24] The increasing awareness of the impact of overtreatment on PTMCs also changed dramatically the international guidelines.[6] Therefore, non-invasive

Table 2 The Performance of Different Deep Learning Models in the Training and Validation Cohorts

		ACC	LOSS	Depth	Number of Layers	PARAM
AlexNet	Training	0.837	0.340	8	25	64,130,754
	Validation	0.800	0.444			
VGG13	Training	0.873	0.275	13	35	64,987,720
	Validation	0.850	0.358			
VGG16	Training	0.854	0.313	16	41	70,297,416
	Validation	0.850	0.304			
VGG19	Training	0.847	0.363	19	47	75,607,112
	Validation	0.850	0.325			
EfficientNet	Training	0.908	0.270	82	290	4,052,133
	Validation	0.863	0.610			

Abbreviations: ACC, accuracy; LOSS, training loss value; PARAM, model parameters amount; VGG, visual geometry group.

a)

Papillary thyroid microcarcinoma
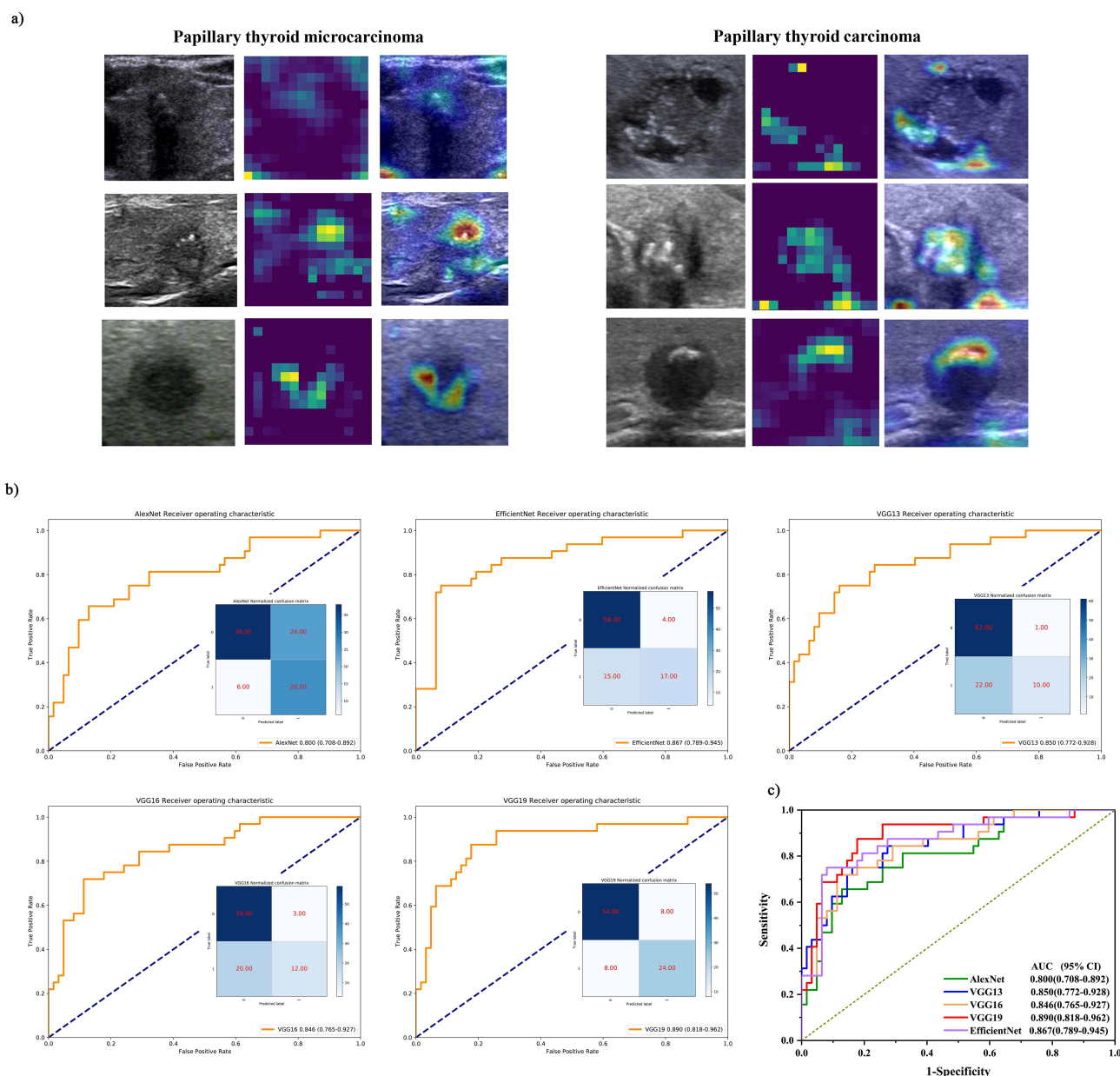
Papillary thyroid carcinoma

b)

Figure 4 (**a**) The network features and heatmaps of 3 cases of papillary microcarcinoma of the thyroid and papillary thyroid carcinoma. The ROC curves and the corresponding confusion matrices of different deep learning models, (**b**) AlexNet, VGG13, VGG16, VGG19, EfficientNet; (**c**) ROCs comparison.

methods are urgently needed to differentiate PTMC from PTC preoperative to avoid overtreatment for patients with PTMC regardless of the increased diagnosis of thyroid cancers.

US is widely applied for preoperative imaging of PTC for diagnosis and staging. US features, such as irregular border, halo sign microcalcifications, macrocalcifications, isoechoic, and hypoechoic appearance, had been investigated to differentiate PTC from PTMC clinically.[25] An AUC of 0.97, a sensitivity of 88.6%, and a specificity of 94.6% were reported in the prediction of PTMC with combined conventional US, CEUS, and RTE.[9] However, the accuracy of US diagnosis is easily affected by image quality and the experience of US technician and radiologists, who handles the probe and interprets the US images.[24] With the emergence of radiomics, the radiomics model had been proposed as a promising method to assess the risk of PTC metastasis using US images and achieved an AUC of 0.782 and an accuracy of 0.710, respectively.[26] The presence of extrathyroidal extension (ETE) of PTC was preoperatively predicted with a radiomics model using CT images and achieved an AUC of 0.812 in the validation cohort.[27] Radiomics models were also

Table 3 The Performance of Radiomics Signature, Deep Learning Models and Combined Deep Learning Radiomics Models with Independent Test 1 and 2

Models	Accuracy	Precision	Sensitivity	Specificity	PPV	NPV	F1 score	AUC
Independent test 1								
VGG19	0.829	0.750	0.750	0.871	0.750	0.871	0.829	0.890
EfficientNet	0.798	0.810	0.530	0.935	0.810	0.795	0.785	0.867
Radiomics	0.766	0.604	0.906	0.694	0.604	0.935	0.772	0.826
R_V_combined	0.904	0.848	0.934	0.848	0.919	0.875	0.905	0.931
R_E_combined	0.851	0.781	0.781	0.887	0.781	0.887	0.851	0.908
Independent test 2								
VGG19	0.680	0.760	0.520	0.840	0.765	0.636	0.67	0.698
EfficientNet	0.680	0.890	0.400	0.960	0.909	0.615	0.679	0.899
Radiomics	0.740	0.730	0.760	0.720	0.731	0.75	0.739	0.822
R_V_combined	0.780	0.818	0.750	0.818	0.840	0.720	0.779	0.874
R_E_combined	0.900	0.884	0.920	0.880	0.885	0.917	0.899	0.946

Abbreviations: PPV, positive predictive value; NPV, negative predictive value; AUC, area under curve; R_V_combined, radiomics and VGG19 combination; R_E_combined, radiomics and EfficientNet combination.

investigated to predict the presence of B-Raf proto-oncogene, serine/threonine kinase (BRAF) mutation in PTC with an average AUC of 0.651, an accuracy of 64.3%, a sensitivity of 66.8%, and a specificity of 61.8% using US images, respectively.[28] However, to the best of our knowledge, this is the first study to investigate the feasibility of US-based radiomics and deep learning models in the differentiation of PTC and PTMC. After manual target segmentation, feature extraction, and selection, an AUC of 0.826 and 0.822 was achieved with radiomics signature in the independent testing set 1 and set 2, respectively, in this study. The performance of our radiomics model was promising in comparison with other radiomics studies for PTC as mentioned previously.[28–30] However, it was inferior to the direct prediction of PTC and PTMC with combined US, CEUS, and RTE.[9]

Five deep learning networks were adapted and trained in this study to further investigate the differentiation of PTMC from PTC in this study. As shown in Table 2 and Figure 4, VGG19 achieved a best AUC of 0.890 and EfficientNet achieved a best accuracy of 0.867, respectively. As shown in Table 3, the performance of radiomics was inferior to two deep learning models in accuracy and AUC with independent testing set 1; however, the accuracy of the radiomics model was better than those of two deep learning models in accuracy with an external validation set (independent testing set 2). However, advanced deep machine learning approaches have been developed to handle the challenging health problems
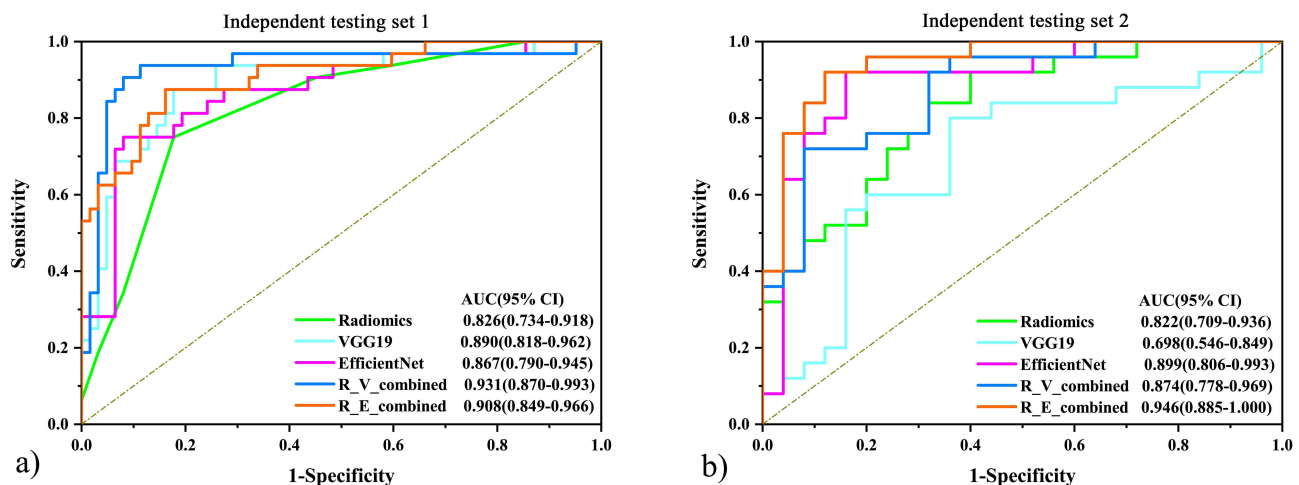

Figure 5 The performance of radiomics, deep learning, and combined deep learning radiomics models in differentiating PTMC from PTC with (**a**) independent testing set 1; (**b**) independent testing set 2.

and to diagnose thyroid cancer.[31,32] In this study, the application of deep learning networks did not guarantee superior differentiation ability for PTC and PTMC. On the other hand, the interpretability of deep learning models was weaker in comparison with radiomics features.[33] In this study, visualization class was activated to generate a heatmap after the differentiation with deep learning networks to increase the interpretability of our deep learning models. As shown in Figure 4a the heatmap corresponding to the boundary of the tumor and the low echo area inside the tumor were highlighted, which are the features corresponding to deep learning network features.

In this study, the differentiation ability and accuracy of these models were further improved by combining radiomics and deep learning networks, as shown in Table 3 and Figure 5. The best accuracy and AUC of 0.904, 0.900, and 0.931, 0.946 were achieved with the combination of VGG + radiomics (R_V_Combined) and EffiecientNet + radiomics (R_E_Combined) in the independent testing set 1 and set 2, respectively. Consistently, deep learning radiomics had been reported to improve the prediction ability using US images for breast cancers.[34,35] The combination of radiomics scores and deep learning prediction scores was usually fused by information fusion, which includes three strategies of early fusion, mid-term fusion, and late-stage fusion.[19] However, only the late-stage fusion strategy was applied in this study. Although automatic segmentation on US images for cervical cancer and ovarian cancer had been intensively investigated, further study is still necessary to transfer the automatic segmentation methods to PTC and PTMC on US images.[29,30] On the other hand, the stability and reproducibility of radiomics features may be easily affected by the type of scan machines and the different automatic segmentation algorithms.[36,37] Another limitation of this study is that the influence of clinical factors in the differentiation of PTMC and PTC was not fully investigated.

Conclusions

Deep learning and radiomics combination models are promising in the noninvasively preoperative differentiation of PTMC and PTC to decrease the overtreatment of patients with PTMC and to minimize the complications caused by overtreatment.

Ethics Approval and Informed Consent

This study was approved by the ethics committee in Clinical Research (ECCR) of authors' hospital and conducted following the Declaration of Helsinki (ECCR no. 2019059). The requirement of informed consent was waived by the ECCR.

Acknowledgments

We would like to thank Jianping Wu, who made a significant contribution in the revision of this manuscript.

Disclosure

No potential conflict of interest relevant to this article was reported. This study had been presented in the 2023 conference of the American Association of Physicists in Medicine.

References

1. Seib CD, Sosa JA. Evolving understanding of the epidemiology of thyroid cancer. *Endocrinol Metab Clin North Am*. 2019;48(1):23–35. doi:10.1016/j.ecl.2018.10.002

2. Vaccarella S, Franceschi S, Bray F, Wild CP, Plummer M, Dal Maso L. Worldwide thyroid-cancer epidemic? The increasing impact of overdiagnosis. *N Engl J Med*. 2016;375(7):614–617. doi:10.1056/NEJMp1604412

3. Vigneri R, Malandrino P, Vigneri P. The changing epidemiology of thyroid cancer: why is incidence increasing? *Curr Opin Oncol*. 2015;27(1):1–7. doi:10.1097/CCO.0000000000000148

4. Lloyd RV, Buehler D, Khanafshar E. Papillary thyroid carcinoma variants. *Head Neck Pathol*. 2011;5(1):51–56. doi:10.1007/s12105-010-0236-9

5. Wang W, Kong L, Guo H, Chen X. Prevalence and predictor for malignancy of contralateral thyroid nodules in patients with unilateral PTMC: a systematic review and meta-analysis. *Endocr Connect*. 2021;10(6):656–666. doi:10.1530/EC-21-0164

6. Sugitani I, Ito Y, Takeuchi D, et al. Indications and strategy for active surveillance of adult low-risk papillary thyroid microcarcinoma: consensus statements from the Japan association of endocrine surgery task force on management for papillary thyroid microcarcinoma. *Thyroid*. 2021;31 (2):183–192. doi:10.1089/thy.2020.0330

7. Li M, Dal Maso L, Vaccarella S. Global trends in thyroid cancer incidence and the impact of overdiagnosis. *Lancet Diabetes Endocrinol*. 2020;8 (6):468–470. doi:10.1016/S2213-8587(20)30115-7

8. Chen HY, Liu WY, Zhu H, et al. Diagnostic value of contrast-enhanced ultrasound in papillary thyroid microcarcinoma. *Exp Ther Med*. 2016;11 (5):1555–1562. doi:10.3892/etm.2016.3094

9. Ma HJ, Yang JC, Leng ZP, Chang Y, Kang H, Teng LH. Preoperative prediction of papillary thyroid microcarcinoma via multiparameter ultrasound. *Acta Radiol*. 2017;58(11):1303–1311. doi:10.1177/0284185117692167

10. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278(2):563–577. doi:10.1148/radiol.2015151169

11. Kwon MR, Shin JH, Hahn SY, et al. Histogram analysis of greyscale sonograms to differentiate between the subtypes of follicular variant of papillary thyroid cancer. *Clin Radiol*. 2018;73(6):591.e1–591.e7. doi:10.1016/j.crad.2017.12.008

12. Akkus Z, Cai J, Boonrod A, et al. A survey of deep-learning applications in ultrasound: artificial intelligence-powered ultrasound for improving clinical workflow. *J Am Coll Radiol*. 2019;16(9 Pt B):1318–1328. doi:10.1016/j.jacr.2019.06.004

13. Li X, Zhang S, Zhang Q, et al. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study [published correction appears in Lancet Oncol. 2020 Oct;21(10):e462. doi: 10.1016/S1470-2045(20) 30546-5]. *Lancet Oncol*. 2019;20(2):193–201. doi:10.1016/S1470-2045(18)30762-9

14. Peng S, Liu Y, Lv W, et al. Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study [published correction appears in Lancet Digit Health. 2021 Jul;3(7):e413]. *Lancet Digit Health*. 2021;3(4):e250–e259. doi:10.1016/S2589-7500(21)00041-8

15. Tessler FN, Middleton WD, Grant EG. Thyroid imaging reporting and data system (TI-RADS): a user's guide [published correction appears in Radiology. 2018 Jun;287(3):1082]. *Radiology*. 2018;287(1):29–36. doi:10.1148/radiol.2017171240

16. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1–22. doi:10.18637/jss.v033.i01

17. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arxiv:14091556[csCV].

18. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. arxiv:151203385[csCV].

19. Meurer WJ, Tolles J. Logistic regression diagnostics: understanding how well a model predicts outcomes. *JAMA*. 2017;317(10):1068–1069. doi:10.1001/jama.2016.20441

20. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983;148(3):839–843. doi:10.1148/radiology.148.3.6878708

21. Jung KW, Won YJ, Kong HJ, Lee ES; Community of Population-Based Regional Cancer Registries. Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2015. *Cancer Res Treat*. 2018;50(2):303–316. doi:10.4143/crt.2018.143

22. Ahn HS, Kim HJ, Welch HG. Korea's thyroid-cancer "Epidemic" — screening and overdiagnosis. *N Engl J Med*. 2014;371(19):1765–1767. doi:10.1056/NEJMp1409841

23. Cordioli MI, Canalli MH, Coral MH. Increase incidence of thyroid cancer in Florianopolis, Brazil: comparative study of diagnosed cases in 2000 and 2005. *Arq Bras Endocrinol Metabol*. 2009;53(4):453–460. doi:10.1590/s0004-27302009000400011

24. Haugen BR, Alexander EK, Bible KC, et al. 2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2016;26(1):1–133. doi:10.1089/thy.2015.0020

25. Zhang XL, Qian LX. Ultrasonic features of papillary thyroid microcarcinoma and non-microcarcinoma. *Exp Ther Med*. 2014;8(4):1335–1339. doi:10.3892/etm.2014.1910

26. Liu T, Zhou S, Yu J, et al. Prediction of lymph node metastasis in patients with papillary thyroid carcinoma: a radiomics method based on preoperative ultrasound images. *Technol Cancer Res Treat*. 2019;18:1533033819831713. doi:10.1177/1533033819831713

27. Chen B, Zhong L, Dong D, et al. Computed tomography radiomic nomogram for preoperative prediction of extrathyroidal extension in papillary thyroid carcinoma. *Front Oncol*. 2019;9:829. doi:10.3389/fonc.2019.00829

28. Kwon MR, Shin JH, Park H, Cho H, Hahn SY, Park KW. Radiomics study of thyroid ultrasound for predicting braf mutation in papillary thyroid carcinoma: preliminary results. *AJNR Am J Neuroradiol*. 2020;41(4):700–705. doi:10.3174/ajnr.A6505

29. Jin J, Zhu H, Zhang J, et al. Multiple U-net-based automatic segmentations and radiomics feature stability on ultrasound images for patients with ovarian cancer. *Front Oncol*. 2021;10:614201. doi:10.3389/fonc.2020.614201

30. Jin J, Zhu H, Teng Y, Ai Y, Xie C, Jin X. The accuracy and radiomics feature effects of multiple U-net-based automatic segmentation models for transvaginal ultrasound images of cervical cancer. *J Digit Imaging*. 2022;35(4):983–992. doi:10.1007/s10278-022-00620-z

31. Tutsoy O, Koç GG. Deep self-supervised machine learning algorithms with a novel feature elimination and selection approaches for blood test-based multi-dimensional health risks classification. *BMC Bioinf*. 2024;25(1):103. doi:10.1186/s12859-024-05729-2

32. Tutsoy O, Sumbul HE. A novel deep machine learning algorithm with dimensionality and size reduction approaches for feature elimination: thyroid cancer diagnoses with randomly missing data. *Brief Bioinform*. 2024;25(4):bbae344. doi:10.1093/bib/bbae344

33. Gandin I, Scagnetto A, Romani S, Barbati G. Interpretability of time-series deep learning models: a study in cardiovascular patients admitted to Intensive care unit. *J Biomed Inform*. 2021;121:103876. doi:10.1016/j.jbi.2021.103876

34. Zheng X, Yao Z, Huang Y, et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer [published correction appears in Nat Commun. 2021 Jul 12;12(1):4370]. *Nat Commun.* 2020;11(1):1236. doi:10.1038/s41467-020-15027-z

35. Jiang M, Li CL, Luo XM, et al. Ultrasound-based deep learning radiomics in the assessment of pathological complete response to neoadjuvant chemotherapy in locally advanced breast cancer. *Eur J Cancer.* 2021;147:95–105. doi:10.1016/j.ejca.2021.01.028

36. Yi J, Lei X, Zhang L, et al. The influence of different ultrasonic machines on radiomics models in prediction lymph node metastasis for patients with cervical cancer. *Technol Cancer Res Treat.* 2022;21:15330338221118412. doi:10.1177/15330338221118412

37. Teng Y, Ai Y, Liang T, et al. The effects of automatic segmentations on preoperative lymph node status prediction models with ultrasound radiomics for patients with early stage cervical cancer. *Technol Cancer Res Treat.* 2022;21:15330338221099396. doi:10.1177/15330338221099396